
WOJSKOWA AKADEMIA TECHNICZNA

im. Jarosława Dąbrowskiego

WYDZIAŁ CYBERNETYKI



Pakiet testów losowości do badania prymitywów kryptograficznych

ROZPRAWA DOKTORSKA

Autor:

mgr inż. Krzysztof MAŃK

Promotor pracy:

dr hab. Marek KOJDECKI, prof. WAT

Promotor pomocniczy:

dr inż. Michał WROŃSKI

Warszawa, 2025

*Any one who considers
arithmetical methods of producing random digits
is, of course, in a state of sin.*

John von Neuman

Various Techniques Used in Connection with Random Digits

Oznaczenia

$s^n = s_1, s_2, \dots, s_n$ – ciąg binarny o długości n , gdzie $s_i \in \{0;1\}$

$B^{d,n} = b_1, b_2, \dots, b_n$ – ciąg d bitowych bloków binarnych o długości n , gdzie $b_i \in (s_{i,1}, s_{i,2}, \dots, s_{i,d})$

$U^n = u_1, u_2, \dots, u_n$ – ciąg liczb rzeczywistych długości n , gdzie $u_i \in [0;1)$

$\varphi(x)$ – gęstość rozkładu normalnego $N(0;1)$ w punkcie x

$\Phi(x)$ – dystrybuanta rozkładu normalnego $N(0;1)$ w punkcie x

$\Phi^{-1}(\alpha)$ – kwantyl rozkładu normalnego $N(0;1)$ rzędu α

$T_k(x)$ – dystrybuanta rozkładu T Studenta z k stopniami swobody w punkcie x

$T_k^{-1}(\alpha)$ – kwantyl rozkładu T Studenta z k stopniami swobody rzędu α

$\chi_k^2(x)$ – dystrybuanta rozkładu chi-kwadrat z k stopniami swobody w punkcie x

$E[X]$ – wartość oczekiwana zmiennej losowej X

$D^2[X]$ – wariancja zmiennej losowej X

$\lg x$ – logarytm przy podstawie 2 z x

$\lfloor x \rfloor$ – największa liczba całkowita, nie większa od x

$a.b$ – kolejne liczby całkowite od a do b

$a | b$ – a dzieli b , $a, b \in \mathbb{Z}$

$a \nmid b$ – a nie dzieli b , $a, b \in \mathbb{Z}$

$Hwt(s)$ – waga Hamminga argumentu s , oznacza liczbę 1 w ciągu binarnym s

$I[\omega]$ – funkcja prawdy, $I[\omega] = 1$, gdy zdanie ω jest prawdziwe, zaś 0 w przeciwnym wypadku

$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$ – funkcja Gamma

$\binom{a}{b} = \frac{\Gamma(a+1)}{\Gamma(b+1)\Gamma(a-b+1)}$, $a, b \in \mathbb{R}$ – uogólniony symbol Newtona

$\left\{ \begin{matrix} k \\ r \end{matrix} \right\}$ – liczba Stirlinga II rodzaju, określająca liczbę sposobów podziału zbioru k -elementowego na r niepustych podzbiorów

$\Pr(\Xi)$ – prawdopodobieństwo zdarzenia Ξ

Wstęp	13
I. Losowość, generatory i testy statystyczne	16
1. Pojęcie losowości	16
2. Generatory	18
3. Testy statystyczne	19
3.1. Badanie losowości	19
3.2. Test statystyczny	20
3.3. Testy drugiego poziomu	21
3.4. Test trzeciego poziomu	22
4. Cel pracy	23
II. Testy losowości	25
1. Procedury ogólne	25
1.1. Test Pearsona – zgodności chi-kwadrat	25
1.2. Test Kołmogorowa-Smirnowa	26
1.3. Test Kołmogorowa	27
1.4. Test Kuipera	28
1.5. Rodzina testów Andersona-Darlinga	29
1.6. Test Cramera-von Misesa	29
1.7. Test Andersona-Darlinga	30
1.8. Test dla dwóch liczb	30
1.9. Test dla trzech liczb	31
2. Testy dla ciągów bloków bitów	31
2.1. Testy bazujące na procedurach ogólnych	32
2.1.1. Test częstości	32
2.1.2. Test zgodności rozkładu	32
2.2. Testy momentów	32
2.2.1. Test momentu pierwszego rzędu dla ciągu liczb	33
2.2.2. Test momentu pierwszego rzędu dla ciągu bitów	33
2.2.3. Test momentu drugiego rzędu dla ciągu liczb	33
2.2.4. Test momentu drugiego rzędu dla ciągu bitów	34
2.3. Test średniej	34
2.4. Test autokorelacji	35
2.5. Test korelacji	35
2.6. Test iloczynów	35
2.7. Test sumy logarytmów	36
2.8. Test częstości dla podprzedziału	36
2.9. Test kolizji indeksów maksymalnych elementów	36
2.10. Test sum do g	36
2.11. Test odstępów	37
2.12. Test sumy logarytmów odstępów	37
2.13. Test sumy kwadratów odstępów	38
2.14. Test odstępów dni urodzin	39
2.15. Test największej liczby elementów w przedziale	39
2.16. Uniwersalny test Maurera	40
2.17. Test Corona	41
2.18. Stos książek	41
2.18.1. Test stosu książek	41

2.18.2. Test porządku	42
2.19. Test zgodności liczby komórek z dokładnie b punktami (Quadratic Box- Test)	43
2.20. Test pochodnej ciągu liczb	46
2.21. Test kolekcjonera	47
2.22. Test serii rosnących / malejących	47
2.22.1. Test serii z macierzą kowariancji	47
2.22.2. Test serii z pomijaniem	49
2.23. Test kolizji	49
2.24. Test ściskania	49
2.25. Test oparty na grze w kości	50
2.26. Test największego wspólnego dzielnika	50
2.27. Test wypełniania drzewa binarnego na podstawie ciągu liczb	51
2.28. Test autokorelacyjny dla ciągu bloków bitów	52
3. Testy dla ciągów nienakładających się wektorów	54
3.1. Testy dla zliczeń liczby punktów w komórkach	54
3.2. Test m najbliższych par	57
3.3. Test zgodności bitowej	59
3.4. Test Bickela-Breimana	59
3.5. Test największy- z - t	60
3.6. Test permutacji	60
3.7. Test parkingowy	60
4. Testy dla ciągów nakładających się wektorów	61
4.1. Testy dla zliczeń liczby punktów w komórkach	61
4.1.1. Test potęgowo-dywergencyjny dla przypadku gęstego	61
4.1.2. Test potęgowo-dywergencyjny dla przypadku rzadkiego	61
4.1.3. Test liczby wolnych komórek	62
4.1.4. Test częstości dla nakładających się wektorów	62
4.1.5. Test entropii	64
4.2. Test nakładających się permutacji	65
4.3. Testy wolnych cel	66
4.4. Test „policz jedynek w ciągu bloków”	67
5. Testy dla ciągów bitów	67
5.1. Test autokorelacyjny	67
5.2. Test liczby serii	68
5.3. Test serii z macierzą kowariancji	68
5.4. Test do liczby serii N	70
5.5. Testy losowych wycieczek	71
5.6. Testy losowych wycieczek z liniowym wyznaczaniem kierunku	72
5.7. Testy złożoności liniowej ciągu	72
5.7.1. Test liczby skoków w profilu liniowym	72
5.7.2. Test rozkładu wysokości skoków w profilu liniowym	72
5.7.3. Test rozkładu złożoności liniowej podciągów	73
5.8. Test złożoności Lempela-Ziva	73
5.9. Testy spektralne	74
5.9.1. Test spektralny	74
5.9.2. Test spektralny sumy kwadratów modułów – wariant 1	74
5.9.3. Test spektralny sumy kwadratów modułów – wariant 2	75
5.10. Test okresowości w ciągu	75
5.11. Test najdłuższej serii jedynek w bloku	76
5.12. Testy wag Hamminga	76
5.12.1. Test częstości wag Hamminga	76
5.12.2. Test wag Hamminga	77

5.12.3. Test korelacji wag Hamminga	77
5.12.4. Testy niezależności wag Hamminga	77
5.13. Test rzędu macierzy binarnej	78
5.14. Test nienakładających się wzorców	78
5.15. Test nakładających się wzorców	78
5.16. Test wypełniania drzewa binarnego na podstawie ciągu bitów	79
III. Wybrane pakiety testów losowości	85
1. Knuth II.3.3.2	85
1.1. Test równomierności (Equidistribution Test)	85
1.2. Test częstości par (Serial Test)	85
1.3. Test odstępów (Gap Test)	85
1.4. Test podziałów (Partition Test)	85
1.5. Test kolekcjonera (Coupon Collector's Test)	86
1.6. Test permutacji (Permutation Test)	86
1.7. Test serii rosnących / malejących (Run Test)	86
1.7.1. Test z macierzą kowariancji	86
1.7.2. Test z pomijaniem	86
1.8. Test największy- z - t (Maximum-of- t Test)	86
1.9. Test kolizji (Collision Test)	86
1.10. Test autokorelacji (Serial Correlation Test)	86
2. DIEHARD	86
2.1. Test odstępów dni urodzin (Birthday Spacings Test)	87
2.2. Test nakładających się permutacji piątek (Overlapping 5-Permutation Test)	87
2.3. Test rzędu macierzy binarnej (Binary Rank Test)	87
2.4. Testy wolnych cel	88
2.4.1. Test nakładających się słów 20-bitowych (Bitstream Test)	88
2.4.2. Test nakładających się par (Overlapping-Pairs-Sparse-Occupancy Test)	88
2.4.3. Test nakładających się czwórek (Overlapping-Quadruples-Sparse-Occupancy Test)	89
2.4.4. Test DNA (DNA Test)	89
2.5. Test „policz jedyne w ciągu bajtów” (Count-the-1 Test on a Stream of Bytes)	89
2.6. Test „policz jedyne w bajtach” (Count-the-1's Test for Specific Bytes)	90
2.7. Test parkingowy (Parking Lot Test)	90
2.8. Test minimalnej odległości (Minimum Distance Test)	90
2.9. Test kul (3D Spheres Test)	90
2.10. Test ściskania (Squeeze Test)	91
2.11. Test nakładających się sum (Overlapping Sums Test)	91
2.12. Test serii (Runs Test)	94
2.13. Test oparty na grze w kości (Craps Test)	94
3. FIPS PUB 140	94
3.1. Test monobitowy	95
3.2. Test pokerowy	95
3.3. Testy liczby serii	95
3.4. Test długich serii	95
4. AIS-31	95
4.1. Test rozłączności (disjointness test)	95
4.2. Test monobitowy	96
4.3. Test pokerowy	96
4.4. Testy liczby serii	96
4.5. Test długich serii	96
4.6. Test autokorelacyjny	96
4.7. Test równomierności rozkładu	96

4.8. Test porównawczy dla rozkładów wielowymiarowych	97
4.9. Test entropii	97
5. NIST SP 800-22 Statistical Test Suite	97
5.1. Test częstości bitów (Frequency Monobit Test)	98
5.2. Test częstości w bloku (Frequency Test within a Block)	98
5.3. Test serii (Runs Test)	98
5.4. Test najdłuższej serii jedynek w bloku (Test for the Longest Run of Ones in a Block)	98
5.5. Test rzędu macierzy binarnej (Binary Matrix Rank Test)	100
5.6. Test spektralny – dyskretna transformata Fouriera (Discrete Fourier Transform Test)	101
5.7. Test nienakładających się wzorców (Non-overlapping Template Matching Test)	101
5.8. Test nakładających się wzorców (Overlapping Template Matching Test)	101
5.9. Uniwersalny test Maurera (Maurer's Universal Statistical Test)	101
5.10. Test złożoności liniowej (Linear Complexity Test)	102
5.11. Test częstości nakładających się bloków (Serial Test)	103
5.12. Test entropii (Approximate Entropy Test)	103
5.13. Test sum (Cumulative Sums Test)	103
5.14. Test losowych wycieczek (Random Excursions Test)	104
5.15. Alternatywny test losowych wycieczek (Random Excursions Variant Test)	106
5.16. Test drugiego poziomu	106
6. TestU01	107
6.1. Klasa scomp	107
6.1.1. Testy złożoności liniowej	107
6.1.2. Test złożoności Lempela-Ziva	107
6.2. Klasa smultin	108
6.2.1. Wektory nienakładające się	108
6.2.2. Wektory nakładające się	108
6.3. Klasa snpair	108
6.3.1. Test m najbliższych par (m Nearest Pairs Test)	108
6.3.2. Test zgodności bitowej (Close Pairs Bit Match)	108
6.3.3. Test Bickela-Breimana	108
6.4. Klasa sspacings	108
6.4.1. Test sumy logarytmów odstępów	108
6.4.2. Test sumy kwadratów odstępów	109
6.4.3. Test największej liczby elementów w przedziale	109
6.5. Klasa sspectral	109
6.5.1. Test spektralny	109
6.5.2. Test spektralny sumy kwadratów modułów – wariant 1	109
6.5.3. Test spektralny sumy kwadratów modułów – wariant 2	109
6.6. Klasa sstring	109
6.6.1. Test okresowości w ciągu (Periods in Strings Test)	109
6.6.2. Test najdłuższej serii jedynek w bloku	109
6.6.3. Test częstości wag Hamminga	109
6.6.4. Test wag Hamminga	109
6.6.5. Test korelacji wag Hamminga	109
6.6.6. Testy niezależności wag Hamminga	109
6.6.7. Testy do liczby serii N	109
6.6.8. Test autokorelacyjny	109
6.7. Klasa svaria	109
6.7.1. Test średniej (Sample Mean)	109
6.7.2. Test korelacji (Sample Correlation)	110

6.7.3. Test iloczynów (Sample Product)	110
6.7.4. Test sumy logarytmów (Sum of Logarithms)	110
6.7.5. Test częstości dla podprzedziału (Weight Distribution)	110
6.7.6. Test kolizji indeksów maksymalnych elementów (Collision Argument Maximal)	110
6.7.7. Test sum do g (Sum Collector)	110
6.7.8. Test odległości (Appearance Spacings)	110
6.8. Klasa swalk	110
6.8.1. Testy losowych wycieczek (Random Walk)	110
6.8.2. Testy losowych wycieczek z liniowym wyznaczaniem kierunku	110
6.8.3. Test drogi do (Random Walk)	110
7. Dieharder	111
7.1. Testy zaczerpnięte z pakietu DIEHARD	111
7.1.1. Test odstępów dni urodzin (diehard_birthdays)	111
7.1.2. Test nakładających się permutacji piątek (diehard_operm5)	111
7.1.3. Test rzędu macierzy binarnej (diehard_rank_32x32 i diehard_rank_6x8)	111
7.1.4. Test nakładających się słów 20-bitowych (diehard_bitstream)	111
7.1.5. Test nakładających się par (diehard_opso)	111
7.1.6. Test nakładających się czwórek (diehard_oqso)	112
7.1.7. Test DNA (diehard_dna)	112
7.1.8. Test „policz jedynki w ciągu bajtów” (diehard_count_1s_str)	112
7.1.9. Test „policz jedynki w bajtach” (diehard_count_1s_byt)	112
7.1.10. Test parkingowy (diehard_parking_lot)	112
7.1.11. Test minimalnej odległości (diehard_2dsphere)	112
7.1.12. Test kul (diehard_3dsphere)	112
7.1.13. Test ściskania (diehard_squeeze)	112
7.1.14. Test nakładających się sum (diehard_sums)	112
7.1.15. Test serii wg DIEHARD (diehard_runs)	112
7.1.16. Test oparty na grze w kości (diehard_craps)	112
7.2. Testy zaczerpnięte z pakietu STS	112
7.2.1. Test częstości bitów (sts_monobit)	112
7.2.2. Test serii wg NIST (sts_runs)	112
7.2.3. Test częstości nakładających się bloków (sts_serial)	112
7.3. Test największego wspólnego dzielnika (marsaglia_tsang_gcd)	113
7.4. Test częstości bloków bitów (rgb_bitdist)	113
7.5. Uogólniony test minimalnej odległości (rgb_minimum_distance)	114
7.6. Test opóźnionych sum (rgb_lagged_sums)	114
7.7. Test lmn (rgb_lmn)	115
7.8. Test nakładających się permutacji (rgb_operm)	115
7.9. Test permutacji (rgb_permutations)	116
7.10. Test niezmienności bitu (rgb_persist)	117
7.11. Test częstości bajtów (dab_bytedistrib)	118
7.12. Test dyskretnej transformaty kosinusowej (dab_dct)	119
7.13. Test wypełniania drzewa binarnego na podstawie ciągu liczb (dab_filltree)	119
7.14. Test wypełniania drzewa binarnego na podstawie ciągu bitów (dab_filltree2)	119
7.15. Test wag Hamminga bloków (dab_monobit2)	119
8. Crypt-X	119
8.1. Test częstości	120
8.2. Test autokorelacji	120
8.3. Test częstości dla różnych długości bloków nienakładających się	120
8.4. Test serii	120
8.5. Test złożoności sekwencji	120

8.6. Test złożoności liniowej	120
8.7. Test entropii w bloku	120
IV. Trójpoziomowa procedura testowa	121
1. Generator	121
2. Test	122
3. Weryfikacja generatorów	122
4. Weryfikacja testów	127
V. Określenie przydatności procedur testowych do badania prymitywów kryptograficznych	132
1. Procedura	132
1.1. Weryfikacja rozróżnialności generatorów	132
1.2. Weryfikacja przydatności procedury testowej do badania pojedynczych ciągów	133
1.3. Weryfikacja przydatności procedury testowej do badania generatorów przy małej liczbie powtórzeń	133
1.4. Weryfikacja przydatności procedury testowej do badania generatorów przy dużej liczbie powtórzeń	134
2. Procedury dla ciągów długości 1 kiB	134
2.1. Zestaw początkowy	134
2.2. Weryfikacja rozróżnialności generatorów	134
2.3. Weryfikacja przydatności do badania pojedynczych ciągów długości 1 kiB	135
2.4. Weryfikacja przydatności do badania generatorów przy małej liczbie powtórzeń	136
2.5. Weryfikacja przydatności do badania generatorów przy dużej liczbie powtórzeń	136
2.6. Podsumowanie eliminacji procedur dla ciągów 1 kiB	137
3. Procedury dla ciągów długości 1 Mb	137
3.1. Zestaw początkowy	137
3.2. Weryfikacja rozróżnialności generatorów	137
3.3. Weryfikacja przydatności do badania pojedynczych ciągów długości 1 Mb	138
3.4. Weryfikacja przydatności do badania generatorów przy małej liczbie powtórzeń	139
3.5. Weryfikacja przydatności do badania generatorów przy dużej liczbie powtórzeń	139
3.6. Podsumowanie eliminacji procedur dla ciągów 1 Mb	140
3.7. Weryfikacja procedur testowych pakietu Statistical Test Suite	140
4. Podsumowanie wyników eliminacji	141
VI. Skorelowanie procedur testowych	142
1. Weryfikacja istotności korelacji	142
1.1. Procedura weryfikacji istotności korelacji	142
1.2. Wyniki weryfikacji istotności korelacji dla procedur dla ciągów o długości 1 kiB	143
1.3. Wyniki weryfikacji istotności korelacji dla procedur dla ciągów o długości 1 Mb	145
2. Dwuwymiarowa regresja liniowa i nieliniowa dla prawdopodobieństw testowych	147
2.1. Regresja dwuwymiarowa	147
2.2. Korelacja wystarczająco bliska 1	151
2.3. Wyniki dla regresji dwuwymiarowej uzyskane dla procedur dla ciągów o długości 1 kiB	155
2.3.1. Korelacja równa 1	155
2.3.2. Korelacja bliska 1	157
2.3.3. Korelacja nieliniowa	163
2.3.4. Wyniki dla prawdopodobieństw testowych drugiego poziomu dla 128 powtórzeń	165

2.4. Wyniki dla regresji dwuwymiarowej uzyskane dla procedur dla ciągów o długości 1 Mb	166
2.4.1. Korelacja równa 1	166
2.4.2. Korelacja bliska 1	167
2.4.3. Korelacja nieliniowa	170
2.5. Wpływ długości ciągu na siłę korelacji	170
3. Grupy skorelowanych procedur	171
3.1. Grupy skorelowanych procedur dla ciągów o długości 1 kiB	171
3.2. Grupy skorelowanych procedur dla ciągów o długości 1 Mb	175
4. Wielowymiarowa regresja liniowa	179
4.1. Regresja wielowymiarowa	179
4.2. Dobór procedur do kombinacji	179
4.3. Wyniki regresji wielowymiarowej dla procedur dla ciągów o długości 1 kiB	180
4.4. Wyniki regresji wielowymiarowej dla procedur dla ciągów o długości 1 Mb	183
5. Podsumowanie analizy skorelowania procedur	192
Podsumowanie	193
Bibliografia	196
 Załącznik A – Zestawy procedur dla ciągów długości 1 kiB	204
1. Zestaw początkowy	204
1.1. Procedury dla ciągów bloków bitów	204
1.2. Procedury dla ciągów wektorów nienakładających się	207
1.3. Procedury dla ciągów wektorów nakładających się	209
1.4. Procedury dla ciągów bitów	210
2. Procedury testowe przydatne do badania pojedynczych ciągów	212
2.1. Procedury dla ciągów bloków bitów	212
2.2. Procedury dla ciągów wektorów nienakładających się	215
2.3. Procedury dla ciągów wektorów nakładających się	217
2.4. Procedury dla ciągów bitów	218
3. Procedury testowe przydatne do badania generatorów przy małej liczbie powtórzeń	219
3.1. Procedury dla ciągów bloków bitów	219
3.2. Procedury dla ciągów wektorów nienakładających się	221
3.3. Procedury dla ciągów wektorów nakładających się	223
3.4. Procedury dla ciągów bitów	224
4. Procedury testowe przydatne do badania generatorów przy dużej liczbie powtórzeń	225
4.1. Procedury dla ciągów bloków bitów	225
4.2. Procedury dla ciągów wektorów nienakładających się	226
4.3. Procedury dla ciągów wektorów nakładających się	228
4.4. Procedury dla ciągów bitów	228
5. Rekomendowany pakiet procedur dla ciągów długości 1 kiB	228
5.1. Procedury dla ciągów bloków bitów	228
5.2. Procedury dla ciągów wektorów nienakładających się	230
5.3. Procedury dla ciągów wektorów nakładających się	231
5.4. Procedury dla ciągów bitów	231
 Załącznik B – Zestawy procedur dla ciągów długości 1 Mb	232
1. Zestaw początkowy	232
1.1. Procedury dla ciągów bloków bitów	232
1.2. Procedury dla ciągów wektorów nienakładających się	236
1.3. Procedury dla ciągów wektorów nakładających się	238
1.4. Procedury dla ciągów bitów	238
2. Procedury testowe przydatne do badania pojedynczych ciągów	251
2.1. Procedury dla ciągów bloków bitów	251

2.2. Procedury dla ciągów wektorów nienakładających się	256
2.3. Procedury dla ciągów wektorów nakładających się	260
2.4. Procedury dla ciągów bitów	261
3. Procedury testowe przydatne do badania generatorów przy małej liczbie powtórzeń	270
3.1. Procedury dla ciągów bloków bitów	270
3.2. Procedury dla ciągów wektorów nienakładających się	274
3.3. Procedury dla ciągów wektorów nakładających się	278
3.4. Procedury dla ciągów bitów	279
4. Procedury testowe przydatne do badania generatorów przy dużej liczbie powtórzeń	283
4.1. Procedury dla ciągów bloków bitów	283
4.2. Procedury dla ciągów wektorów nienakładających się	287
4.3. Procedury dla ciągów wektorów nakładających się	290
4.4. Procedury dla ciągów bitów	290
5. Rekomendowany pakiet procedur dla ciągów długości 1 Mb	293
5.1. Procedury dla ciągów bloków bitów	293
5.2. Procedury dla ciągów wektorów nienakładających się	296
5.3. Procedury dla ciągów wektorów nakładających się	298
5.4. Procedury dla ciągów bitów	298
5.5. Procedury realizowane poprzez kombinację liniową p_{val} innych procedur	301
5.5.1. Test potęgowo-dywergencyjny	301
5.5.2. Test sumy kwadratów odstępów bez zapętlania	302
5.5.3. Test sumy kwadratów odstępów z zapętlaniem	303

Wstęp

Analiza statystyczna jest jednym z podstawowych i najpotężniejszych narzędzi w nauce. W szczególności, jest ona jednym z narzędzi wykorzystywanych w pracy kryptologa, gdzie pod postacią tzw. testów losowości, służy do wykrywania wszelkiego rodzaju nieprawidłowości mogących pojawiać się w badanych sekwencjach bitów, wskazując niedoskonałości badanych prymitywów kryptograficznych.

Ogromna moc obliczeniowa i pojemność pamięci współczesnych komputerów jest jednak ograniczona i stawia skuteczną zaporę na drodze do sprawdzenia wszystkiego, stąd za ważne uznajemy, aby zużywać ją na obliczenia dające wiarygodne wyniki i nieposiadające dających się usunąć nadmiarowości, a jednocześnie pokryć nimi maksymalny obszar własności, które posiadać powinien ciąg losowy.

Poddanie zbioru ciągów wytwarzanych przy użyciu badanego prymitywu kryptograficznego (np. generator losowy lub pseudolosowy, szyfr blokowy, czy funkcja skrótu) serii testów losowości, jest jednym z ważnych elementów oceny jego jakości, decydującym niekiedy o jego przyjęciu albo odrzuceniu. Procedurę taką zastosowano na przykład w trakcie konkursu na nowy standard algorytmu szyfrowania/deszyfrowania danych – AES.¹ Przykład ten pokazuje, że konieczna jest duża wiarygodność i jednoznaczność otrzymywanych wyników. Wiarygodność rozumiemy jako matematyczną poprawność oceny wydawanej przez każdy z testów z osobna, jak również oceny końcowej. Jednoznaczność oznaczać będzie niezmienność oceny w warunkach niezależnego powtórzenia badania lub modyfikacji aparatu badawczego.

W chwili obecnej, w światowej literaturze, znaleźć można odniesienie do co najmniej siedmiu zbiorów procedur testowych, takich jak: Rozdział 3.2 Sztuki programowania Donalda Knutha², DIEHARD³, NIST Statistical Test Suite⁴, Dieharder⁵, Test U01⁶, ENT, Crypt-X. Część z nich jest pozycjami literaturowymi lub też dostępna jest poprzez Internet, wszystkie jednak, w naszym odczuciu, cechuje brak spełnienia powyższych wymogów. Są to zbiory bardzo różnorodne, częstokroć zawierające się wzajemnie, bądź też całkowicie różniące się zestawem podstawowych testów. W dalszej części pośrednio lub bezpośrednio wykazujemy, że zawarte w nich procedury testowe nie zostały poprawnie zweryfikowane ani pod względem zgodności rozkładu, ani skorelowania staty-

¹ J. Soto, L. Bassham, *Randomness Testing of the Advanced Encryption Standard Finalist Candidates*, NIST IR 6483, 2000

² D.E. Knuth, *The Art of Computer Programming*, Addison-Wesley, 1981

³ G. Marsaglia, *DIEHARD Battery of Tests of Randomness*, 1995

⁴ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001

⁵ R.G. Brown, *DIEHARDER: A GNU Public Licensed Random Number Tester*, Duke University Physics Department, 2006

⁶ P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, Université de Montreal, 2007

styk. W takich warunkach bardzo trudne jest podjęcie decyzji o korzystaniu z konkretnego pakietu, bądź ustalenie kryteriów oceny na podstawie wielu wyników częściowych. Dodatkowym czynnikiem jest również kwestia praw autorskich i możliwości wykorzystywania dostępnych implementacji w działalności komercyjnej.

Tak zrodziła się konieczność stworzenia własnego, poprawnie funkcjonującego, zestawu testów, który będzie nietrywialną kompilacją dostępnej w tej dziedzinie wiedzy. Przedstawiona praca, a przede wszystkim, rezultaty pracy badawczej, mają dostarczyć przesłanek i doprowadzić do sformułowania oceny dotyczącej wszystkich rozpoznanych testów losowości w dwóch płaszczyznach. Będzie to jednostkowa poprawność wyników w każdym z testów, w tym określenie zakresu jego stosowalności. Drugim aspektem będzie przeanalizowanie zależności pomiędzy różnymi testami, mające na celu możliwie duże zawężenie zbioru koniecznych do przeprowadzenia testów. Ma to podwójne znaczenie, pierwsze czysto ekonomiczne, pozwalające w krótszym czasie realizować badanie, drugie bardziej naukowe – uzmysławia nieoczywiste na pierwszy rzut oka relacje pomiędzy różnymi testami.

Sam proces powstawania rozprawy rozłożony został na kilka etapów, które znalazły odzwierciedlenie w jej strukturze. Praca została podzielona na sześć zasadniczych, numerowanych rozdziałów, poprzedzonych wstępem i zwieńczonych podsumowaniem. Rozdział I ma za zadanie wprowadzić czytelnika w ogólne pojęcia związane z badaniem losowości i na ich tle przedstawić cele pracy. Rozdziały II i III stanowią swoiste kompendium testów i pakietów losowości. Opisane tam testy zostały pogrupowane ze względu na przeznaczenie oraz sposób, w jaki interpretują badany ciąg. Rozdziały te, prócz przytoczenia wiedzy literaturowej, rozbudowaliśmy o omówienia własnych wyników uzyskanych dla niektórych testów losowości oraz zauważonych w poszczególnych pakietach mankamentów analitycznych i implementacyjnych. Elementy te wyróżniliśmy w sposób widoczny po lewej.

W rozdziale IV przedstawiliśmy i uzasadniliśmy podstawowe narzędzie wykorzystane w tej pracy – trójpoziomą procedurę testową. Przygotowało to grunt pod zasadniczą część pracy zawartą w rozdziałach V i VI. Pierwszy z nich nosił oryginalnie tytuł „Eliminacja procedur”, który doskonale oddawał czemu poświęciliśmy tę część – usunęliśmy z budowanego pakietu wszystkie procedury, które nie przechodziły trójpoziomej procedury testowej. Wykonaliśmy to w trzech krokach, stopniowo zawężając zbiór rozważanych procedur, co pozwoliło na określenie trzech zbiorów procedur o różnych zakresach stosowalności, a przy okazji ograniczyło nakład obliczeniowy.

Zawartość rozdziału VI można określić jako próbę dalszego ograniczenia liczby procedur przy użyciu korelacji liniowej, nieliniowej, aż wreszcie wielowymiarowej. Ponownie zastosowaliśmy stopniowanie wykorzystywanych narzędzi. Po uzyskaniu dla przeważającej części procedur potwierdzenia istnienia skorelowania, wykorzystaliśmy obliczeniowo najprostszą korelację liniową jako narzędzie eliminacji nadmiarowych procedur. Następnie uogólniliśmy to narzędzie do testo-

wania 11 modeli nieliniowych. Analiza klas, na które zbiór procedur dzieliła relacja skorelowania, doprowadziła nas do zastosowania kolejnego narzędzia w postaci wielowymiarowej regresji liniowej.

Zawartość dwóch dodanych do pracy załączników skierowana jest do tych czytelników, którzy chcieliby użyć naszych wyników do konstrukcji zestawów testów o różnorodnym przeznaczeniu.

Rozdział I

Losowość, generatory i testy statystyczne

1. Pojęcie losowości

W klasycznej już pracy Donalda Knutha *Sztuka programowania*¹ znaleźć można cały podrozdział poświęcony rozważaniom dotyczącym natury i próby sformalizowania pojęcia losowości. Z punktu widzenia tej pracy, bardziej interesujące jest ujęcie losowości zaproponowane przez Yao oraz Bluma i Micaliiego.^{2,3}

Dla komputerów cyfrowych, obliczalnym jest algorytm, którego czas wykonania, mierzony liczbą operacji odpowiedniej maszyny Turinga, nie wykracza poza ramy wielomianowe. Cała współczesna kryptografia, a w szczególności testowania losowości, musi ograniczyć się jedynie do analizy rozwiązań mieszczących się w tych granicach. Indukuje to również naturalną tolerancję stanowiącą kompromis między rozwiązaniem kryptograficznie doskonałym, a jego przybliżeniem realizowanym przez szeroko pojmowane, praktyczne algorytmy kryptograficzne. Jeśli więc realnie dostępne są tylko rozwiązania działające w czasie wielomianowym oraz rzeczona tolerancja jest asymptotycznie ograniczona przez odwrotność dowolnego wielomianu, to szansa na odróżnienie dwóch obiektów mieszczących się w zakresie rzeczonej tolerancji, jest zanedbywalnie mała. Prowadzi to w naturalny sposób do formalnej definicji *funkcji zanedbywalnej*.

Funkcję $\text{negl}: \mathbb{Z}_{\geq 1} \rightarrow \mathbb{R}_{\geq 0}$ nazywa się zanedbywalną, jeśli dla dowolnego $c > 0$ istnieje $n_0 \in \mathbb{Z}_{\geq 0}$ takie, że dla wszystkich $n \geq n_0$, zachodzi $\text{negl}(n) < 1/n^c$.

Można łatwo pokazać, że funkcja $\text{negl}: \mathbb{Z}_{\geq 1} \rightarrow \mathbb{R}_{\geq 0}$ jest zanedbywalna wtedy i tylko wtedy, gdy dla każdego $c > 0$, mamy $\limsup_{n \rightarrow \infty} n^c \cdot \text{negl}(n) = 0$. Dodatkowo, warto zauważyć, że każda funkcja $f: \mathbb{Z}_{\geq 1} \rightarrow \mathbb{R}_{\geq 0}$, która dla prawie wszystkich n spełnia warunek $f(n) \leq \text{negl}(n)$, również jest funkcją zanedbywalną.

Funkcję $f: \mathbb{Z}_{\geq 1} \rightarrow \mathbb{R}$ nazywa się funkcją typu wielomianowego lub ograniczoną wielomianowo, jeżeli istnieją $a, c > 0$ takie, że $|f(n)| \leq n^c + a$, dla wszystkich $n \geq 0$. W takim przypadku stosuje się oznaczenie $f(n) = \text{poly}(n)$.

Łatwo pokazać, że jeśli $f(n) = \text{poly}(n)$, to $1/f(n)$ nie jest funkcją zanedbywalną.

Niech Y będzie algorytmem (również probabilistycznym), który przyjmuje na wejściu parametr bezpieczeństwa $n \in \mathbb{Z}_{\geq 1}$ oraz pozostałe elementy wejścia zakodowane w postaci ciągu bitowe-

¹ D.E. Knuth, *Sztuka programowania*, t.2. *Algorytmy seminumeryczne*, WNT, 2002, s. 159-191

² A.C. Yao, *Theory and Application of Trapdoor Functions*, 23rd IEEE Symposium on Foundations of Computer Science, Chicago, 1982, s. 80-91

³ M. Blum, S. Micali, *How to generate cryptographically strong sequences of pseudo random bits*, 23rd Annual Symposium on Foundations of Computer Science (sfcs 1982), Chicago, IL, USA, 1982, s. 112-117

go $x \in \{0,1\}^{\leq p(n)}$, gdzie p jest pewnym ustalonym wielomianem. Algorytm Υ nazywa się obliczalnym jeżeli istnieje funkcja $t(n) = \text{poly}(n)$ oraz funkcja zaniedbywalna $\text{negl}(n)$ taka, że dla wszystkich $n \in \mathbb{Z}_{\geq 1}$ oraz dla wszystkich $x \in \{0,1\}^{\leq p(n)}$, prawdopodobieństwo zdarzenia takiego, że czas działania Υ przekroczy $t(n)$ wynosi, co najwyżej $\text{negl}(n)$.

Przypuśćmy, że $n, l \in \mathbb{Z}_{\geq 1}$, gdzie n jest parametrem bezpieczeństwa takim, że $n \leq l$ (zwykle $n \ll l$). Niech dodatkowo $G: \{0,1\}^n \rightarrow \{0,1\}^l$ będzie obliczalnym algorytmem deterministycznym (generatorem pseudolosowym), zaś Υ adwersarzem, będącym obliczalnym algorytmem probabilistycznym. Jak każdy prymityw kryptograficzny, również generator pseudolosowy musi spełniać odpowiednie wymogi bezpieczeństwa, które w tym przypadku są rozumiane w taki sposób, że losowe n -bitowe ziarna są niemal doskonałe, w sensie akceptowalnej tolerancji, „rozciągane” do wymaganej długości l .

Algorytmy obliczalne reprezentujące stronę atakującą (w praktyce zaimplementowane na maszynach znacząco przewyższających moce obliczeniowe typowych użytkowników) posiadają potencjał do odróżnienia wyjść generatora od ciągów losowych, z dokładnością do co najwyżej zaniedbywalnego prawdopodobieństwa. W celu sformalizowania tych intuicji rozważmy dwa eksperymenty, Exp_b dla $b \in \{0,1\}$, zdefiniowane jako

$\text{Exp}_{b,\Upsilon}(1^n)$:

- l -bitowy ciąg z^l wyznaczany jest w następujący sposób:
 - jeżeli $b = 0$, wtedy $x \xleftarrow{\$} \{0,1\}^n$ oraz $z \leftarrow G(x)$;
 - jeżeli $b = 1$, wtedy $z \xleftarrow{\$} \{0,1\}^l$, gdzie $\xleftarrow{\$}$ oznacza losowanie z rozkładem równomiernym;
- ciąg z^l zostaje wysłany do adwersarza Υ , który zwraca wartość bitu $\hat{b} \in \{0,1\}$ (po dokonaniu wielomianowych obliczeń).

Dla $b \in \{0,1\}$, niech E_b oznacza zdarzenie polegające na tym, że Υ zwrócił $\hat{b} = 1$ w $\text{Exp}_{b,\Upsilon}$. Potencjałem adwersarza Υ względem generatora pseudolosowego G (potencjałem do odróżnienia wyjść G od ciągów losowych), nazywa się

$$\text{Adv}_{\Upsilon}(n, G) = |\Pr(E_0) - \Pr(E_1)|.$$

Generator pseudolosowy G nazywa się bezpiecznym, jeżeli dla każdego adwersarza Υ , będącego obliczalnym algorytmem probabilistycznym, zachodzi warunek

$$\text{Adv}_{\Upsilon}(n, G) \leq \text{negl}(n).$$

Mówimy, że adwersarz Υ odróżnia generator G od źródła losowego, jeżeli

$$Adv_{\Upsilon}(n, G) \neq \text{negl}(n).$$

Yao udowodnił, że dowolny algorytm odróżniający generator G od źródła losowego daje się trywialnie przekształcić w algorytm przewidujący kolejny bit ciągu wyjściowego generatora G . Istnienie takowego jest zaś niedopuszczalne w przypadku prymitywów kryptograficznych.

Skończony ciąg nazywać będziemy **ciągiem wyglądającym losowo**, jeśli przechodzi pomyślnie wszystkie testy statystyczne. Generator wytwarzający takie ciągi określać będziemy generatorem ciągów losowych.

Zwykle przyjmuje się, że ciąg wyglądający losowo jest ciągiem losowym, przy czym taka definicja ciągu losowego nie będzie zgodna z podaną przez Knutha⁴, choć podobnie jak tamta będzie zgodna z intuicyjnym pojmowaniem losowości.

Dalej przyjmować będziemy, że ciąg wygląda losowo, jeśli przechodzi pomyślnie wszystkie znane testy statystyczne. Należy od razu zauważyć, że dodanie słowa „znane” do wcześniejszej definicji radykalnie ją osłabia, pozwala jednak na skuteczne badanie losowości ciągów.

W tym miejscu chcemy poczynić pewne definicyjne rozróżnienie – w dalszej części pracy przez **test losowości** rozumieć będziemy pewien ogólny pomysł na badanie własności statystycznych ciągu lub generatora, natomiast jego konkretyzację, realizowaną dla ściśle określonego zestawu parametrów nazywać będziemy **procedurą testową**. Tak więc, na bazie jednego testu budować będziemy niekiedy tysiące procedur testowych. Wprowadzoną powyżej definicję testu statystycznego należy utożsamiać z procedurą testową.

2. Generatory

Generator definiowany jest w literaturze jako ciąg probabilistycznych algorytmów. Jest to bardzo ogólna definicja, która pozwala nie rozpatrywać w jaki sposób ciąg jest wytwarzany. W rzeczywistości mamy jednak do czynienia z dwoma jakościowo różnymi klasami generatorów.

Pierwsza z nich to wszelkiego rodzaju źródła fizyczne, wykorzystujące zjawiska przebiegające w sposób chaotyczny lub losowy, również w rozumieniu mechaniki kwantowej. Pod warunkiem prawidłowego skonstruowania i eksploatacji można by je uznać za **źródła** rzeczywiście **losowe**, gdyby tylko można było stwierdzić, iż oba te warunki zachodzą.

Druga kategoria, to tzw. **generatory pseudolosowe**, czyli wszelkiego rodzaju algorytmy, które, tak jak to było w definicji, przekształcają stosunkowo krótki ciąg na długą sekwencję wartości. Łatwo zauważyć, że na gruncie teorii informacji, generatory takie nie wytwarzają ciągów losowych, jeśli tylko ich długość przewyższa długość losowego klucza, a sam algorytm jest znany, a dokładniej – wystarczające jest jego dwukrotne użycie. Przy jednorazowym wykorzystaniu algorytmu

⁴ D.E. Knuth, *Sztuka programowania, t.2. Algorytmy seminumeryczne*, WNT, 2002, s. 179

można przyjąć, że pewna część generowanego ciągu jest jeszcze niezależna od klucza, ponieważ niesie informację o algorytmie jako takim.

Jak widać, w obu przypadkach, znajdujemy szerokie pole do stosowania testów statystycznych badających losowość wytwarzanych przez te generatory ciągów. W pierwszym przypadku celem stosowania testu będzie wykrycie nieprawidłowości powstałych bądź to na etapie produkcji, bądź eksploatacji konkretnego egzemplarza urządzenia – wykrycie usterek w losowości wytwarzanych ciągów równoznaczne jest ze stwierdzeniem awarii urządzenia. W przypadku generatorów pseudolosowych testy statystyczne są zwykle pierwszym narzędziem w rękach kryptoanalityka, wskazującym mu ewentualne błędy popełnione w trakcie projektowania algorytmu.

3. Testy statystyczne

W punkcie tym przedstawiamy podstawowe pojęcia dotyczące testów statystycznych.

3.1. Badanie losowości

Omawianie tego zagadnienia należy rozpocząć od tzw. postulatów losowości Golomba. W pracy *Shift Register Sequences*⁵, w której rozpatrywał użycie rejestrów przesuwających do wytwarzania ciągów wyglądających losowo, przedstawił trzy podstawowe warunki, które odnieść można nie tylko do rejestrów przesuwających, ale do każdej deterministycznej maszyny o skończonej liczbie stanów wytwarzającej okresowy ciąg binarny. Ciąg spełniający te warunki Golomb nazywał pseudolosowym (ang. *pseudo-noise sequence*).

Poniżej przedstawiamy postulaty losowości Golomba dla ciągu o okresie p :

- W każdym okresie liczba jedynek jest bliska liczbie zer:

$$\sum_{i=1}^p s_i \approx \frac{p}{2}.$$

- Liczba serii jedynek (sekwencji następujących po sobie bitów równych 1 poprzedzonych i wieńczonych przez 0) jest równa całkowitej liczbie serii zer (analogicznie jak dla jedynek). W każdym okresie, połowa serii ma długość 1, jedna czwarta długość 2, jedna ósma długość 3, itd., tak długo, aż liczba serii tak wyznaczona przekracza 1. Co więcej dla każdej z tych długości występuje tyle samo serii jedynek i zer.

- Funkcja autokorelacji $C(\tau)$ jest dwuwartościowa:

$$p \cdot C(\tau) = \sum_{i=1}^p (2 \cdot a_i - 1)(2 \cdot a_{i+\tau} - 1) = \begin{cases} p, & \text{gdy } \tau = 0, \\ k \ll p, & \text{gdy } 0 < \tau < p. \end{cases}$$

Te trzy warunki przybrały obecnie postać trzech bardzo podstawowych testów statystycznych, służących do badania losowości ciągów.

⁵ S.W. Golomb, *Shift Register Sequences*, San Francisco, Holden-Day, 1967

3.2. Test statystyczny

W dalszej części, nieco odmiennie niż w punkcie I.1, testem statystycznym nazywać będziemy algorytm, który dla zadanego ciągu określa prawdopodobieństwo, że losowo wybrany, z populacji o rozkładzie równomiernym, ciąg tej samej długości posiadać będzie właściwości, według określonej miary, gorsze od ciągu badanego. Miarę tę nazywać będziemy statystyką testową, a wyznaczane prawdopodobieństwo testowe oznaczać p_{val} (ang. *p-value*). Formalnie można je zdefiniować przy użyciu prawdopodobieństwa warunkowego:

$$p_{val} = \Pr(\|S_{obs}\| \geq x \mid H_0),$$

gdzie x jest zaobserwowaną wartością statystyki testowej S_{obs} , zaś H_0 hipotezą zerową zakładającą posiadanie przez badany ciąg badanej własności.

Rozumieć je będziemy jako prawdopodobieństwo przypadkowego zaobserwowania ciągu, dla którego badana własność jeszcze bardziej odstaje od idealnego punktu, niż w badanym przypadku. W tym miejscu należy wskazać na istotną różnicę pomiędzy tak zdefiniowanym prawdopodobieństwem p_{val} , a prawdopodobieństwem przyjęcia przez statystykę testową określonej wartości. W zależności od przyjętego w teście modelu (rozkładu statystyki testowej) relacja pomiędzy nimi przyjmować będzie bardzo różne postaci. Dwa najczęściej spotykane w niniejszej pracy przypadki to wykorzystanie rozkładu normalnego lub T-Studenta albo rozkładu chi-kwadrat.

W testach, których statystyka ma rozkład normalny lub T-Studenta, przypadkiem idealnym jest przyjęcie przez nią wartości oczekiwanej, a oddalanie się od niej świadczy o nielosowości badanego ciągu. Prawdopodobieństwo testowe wyznaczamy wówczas jako:

$$p_{val} = 1 - |1 - 2F(S_{obs})|,$$

gdzie F jest dystrybuantą rozkładu statystyki testowej S_{obs} .

W testach, których statystyka ma rozkład chi-kwadrat, o zajściu przypadku idealnego świadczy jej wartość równa zero, zaś jej powiększanie się jest następstwem mniejszej zgodności z hipotezą zerową. Prawdopodobieństwo testowe wyznaczamy jako:

$$p_{val} = 1 - \chi^2(S_{obs}).$$

W przypadku pozostałych testów wykorzystywanych w pracy, staraliśmy się w każdym przypadku określić, co najmniej, rozkład statystyki testowej. Szczególnie kłopotliwe, w omawianym kontekście, okazują się przypadki, w których należy wprost posłużyć się rozkładem dyskretnym, definiuje się wówczas tzw. lewo i prawostronne prawdopodobieństwo:

$$p_{val}^L = \Pr(S_{obs} \leq x \mid H_0) \text{ i } p_{val}^R = \Pr(S_{obs} \geq x \mid H_0),$$

gdzie x , S_{obs} i H_0 określone są jak wcześniej.

p_{val} wyznaczone jest następująco:

$$p_{val} = \begin{cases} p_{val}^R, & \text{gdy } p_{val}^R < p_{val}^L, \\ 1 - p_{val}^L, & \text{gdy } p_{val}^R \geq p_{val}^L < 0,5, \\ 0,5, & \text{w p.p.} \end{cases}$$

Najczęściej jednak dyskretny rozkład teoretyczny służył jedynie do konstrukcji testu zgodności, opartego na statystyce Pearsona (patrz punkt II.1.1.), co usuwało powyższy problem.

Wartości p_{val} bliskie 1 wskazują na dużą zgodność rozpatrywanej własności w badanym ciągu z przewidywaniem wynikającym z hipotezy zerowej, natomiast bliskie 0 – na duże odstępstwo, a niekiedy nawet jej brak.

W przypadku pojedynczego badania należy przyjąć pewną konkretną wartość progową dla p_{val} , poniżej której stwierdzać będziemy, że utraciliśmy zaufanie co zasadności przyjęcia hipotezy zerowej i orzekamy, że badany ciąg testu nie przeszedł. W ten sposób otrzymuje się, doskonale znany w statystyce, **test istotności**, a wartość progową dla p_{val} nazywa się **poziomem istotności**.

3.3. Testy drugiego poziomu

Postulaty losowości Golomba zostały określone w odniesieniu do ciągu okresowego, w praktyce jednak nie jest możliwe przebadanie własności ciągu w całym jego okresie. Co więcej od generatorów wymagamy, aby wytwarzały ciągi pseudolosowe nie tylko w sensie całego okresu, ale również w znacznie mniejszych wycinkach, co prowadzi do konieczności badania różnej długości fragmentów oraz podciągów badanego ciągu. W wyniku takiego badania otrzymujemy zestaw prawdopodobieństw p_{val} odpowiadających każdej kombinacji ciągu i wariantu przeprowadzonych testów. Aby na ich podstawie móc wydać końcowy werdykt co do pseudolosowości generatora musimy posiadać dodatkowe narzędzie, zwane testem drugiego poziomu.

W literaturze znaleźć można szereg takich procedur, np. metody:

- 1-z-n dla testów niezależnych,
- Bonferroniego – 1-z-n dla testów zależnych,
- k-z-n dla testów niezależnych,
- Rudera – k-z-n dla testów zależnych,
- Hommela,
- Simesa,
- Harta i Weissa,
- wykorzystujące rozkład Bernoulliego,
- badania rozkładu zaobserwowanych wartości p_{val} .

Podstawowym problemem, przy stosowaniu którejkolwiek z powyższych metod, jest statystyczna zależność, a przede wszystkim korelacja, wartości p_{val} uzyskiwanych dla konkretnego cią-

gu, która musi występować, nawet jeśli stosujemy bardzo zróżnicowane testy. Jednocześnie wręcz niewykonalne jest zbadanie tej zależności i posłużenie się którąś z metod przeznaczonych dla testów zależnych.

Obecnie, zdecydowanie najchętniej stosowaną, jest metoda badania rozkładu wartości p_{val} . Jej uzasadnienie jest nader proste: jeśli badane ciągi pochodzą z populacji o rozkładzie równomiernym na zbiorze wszystkich ciągów zadanej długości, to uzyskiwane dla konkretnego testu wartości prawdopodobieństw mają rozkład równomierny na przedziale $(0,1)$, przy założeniu ciągłości rozkładu statystyki testowej. Dla zbioru zaobserwowanych wartości p_{val} przeprowadza się test zgodności z rozkładem równomiernym na przedziale (najczęściej) $[0,1]$, który obala albo nie założenie o pseudolosowości generatora.⁶

Należy zdawać sobie sprawę, z poważnych mankamentów tej metody:

- metoda nie uwzględnia korelacji pomiędzy wynikami różnych testów przeprowadzonych dla tego samego ciągu,
- rozkłady wartości dystrybuant są dyskretne, jeśli obliczenia prowadzimy dla skończonych ciągów lub na komputerach cyfrowych,
- niedokładność obliczeń numerycznych zmusza do przejścia na przedział $[0,1]$.

W praktyce jednak, bardziej zwraca się uwagę na fakt, iż:

- metoda jest w zasadzie odporna na błędy numeryczne podczas wyznaczania p_{val} ,
- daje się stosować z powodzeniem zarówno dla niewielkich (od kilku), jak i bardzo licznych zbiorów wartości p_{val} ,
- pozwala wychwycić również generatory wytwarzające ciągi zbyt losowe, tzn. dające ponadoczekiwane dobre wyniki w testach,
- przy ostrożnym użyciu daje się odnieść również do testów z dystrybuantą nieciągłą.

3.4. Test trzeciego poziomu

Test losowości służy do oceny własności ciągu, wielokrotne powtórzenie testów pierwszego poziomu dla ciągów wytworzonych przez badany generator, pozwala ocenić własności generatora. Kolejna agregacja – trzeci poziom, z teoretycznego punktu widzenia będzie niczym innym niż rozłożeniem drugiego poziomu na dwa kroki, chyba że dodamy tu pewne założenia.

Założmy, iż dysponujemy:

- idealnym generatorem liczb losowych z przedziału $[0;1]$,

⁶ patrz rozdział II. Ogólne procedury testowe

- idealnym testem służącym do weryfikacji zgodności z rozkładem równomiernym na przedziale $[0;1]$.

Para jest idealna, jeżeli generator, którego wyjściem są p_{val} uzyskiwane przy użyciu testu, na wejście którego podawane są ciągi pochodzące z idealnego generatora, jest generatorem idealnym.

Idealność takiej pary sprowadza się do przechodzenia przez generator dowolnie wielostopniowej procedury testowania przy użyciu tego testu, gdzie na pierwszym poziomie badane są ciągi pochodzące z generatora, a kolejne poziomy polegają na agregacji częściowych wyników poprzedniego poziomu. Dla porządku należałoby jeszcze przyjąć, że możliwa jest podmiana generatora lub testu na inny idealny bez utraty własności idealności pary.

Chcąc ocenić własności testu losowości – jakość uzyskiwanych przy jego użyciu wyników, budujemy trójstopniową procedurę, w której badanym obiektem jest idealny generator, a testem użytym w roli testu drugiego i trzeciego poziomu jest nasz idealny test. Negatywny wynik takiego badania, wobec założenia o idealności generatora i testu drugiego poziomu, świadczy o niedostatecznej jakości wyników ocenianego testu losowości.

Naszkicowana procedura ma tylko dwa mankamenty – kandydat na idealny generator musiałby zostać uprzednio sprawdzony poprzez wszystkie testy statystyczne o złożoności wielomianowej lub musiałby być to jeden z generatorów opartych na problemie trudnym obliczeniowo, dla którego udowodniono analitycznie jego pseudolosowość. Idealny test musiałby z kolei być wykonywany na komputerze analogowym, gdyż skończoność arytmetyki komputera cyfrowego neguje możliwość jego implementacji. Użycie generatora takiego jak Blum-Blum-Shub, Shamira, czy RSA, oceniamy jako skrajnie niepraktyczne ze względu na gigantyczne, wymagane w przewidzianym badaniu, objętości danych w porównaniu do wydajności takich źródeł. Toteż w rozdziale IV przedstawiamy sposób obejścia tego problemu.

4. Cel pracy

W pracy tej chcemy zwrócić uwagę na dwa aspekty użytkowania pakietu testów losowości, którym naszym zdaniem nie poświęcono dotychczas należytej uwagi. O pierwszym wspomnieliśmy już wcześniej, jest to kwestia poprawności wyników otrzymywanych na podstawie pojedynczego testu. Praktycznie w opisie każdego z testów, które przedstawimy w kolejnym rozdziale, znajduje się sformułowanie: „statystyka testowa ma asymptotycznie rozkład X ” (lub „w przybliżeniu”), które należy odczytać „dla krótkich ciągów to coś zupełnie innego, ale dla wystarczająco długich prawie ma rozkład X ”. Dla wszystkich przedstawionych testów postaramy się pokazać, przy określonej liczbie powtórzeń, czy badany ciąg jest już wystarczająco długi, aby nie dało się tego wychwycić metodami statystycznymi. Drugą, jest skorelowanie statystyk procedur testowych, a dokładniej, dające się empirycznie mierzyć skorelowanie wartości p_{val} otrzymywanych dla tych samych ciągów.

Ponieważ nawet bardzo proste statystyki testowe wciąż skutecznie opierają się próbom analitycznego rozwiązania tych kwestii, postanowiliśmy uciec się do kolejnego przybliżenia w tym świecie przybliżeń – odwrócimy procedurę testową. Biorąc dużą liczbę ciągów pochodzących z „powszechnie uważanych za dobre”, a więc dających, bliskie równomiernemu, rozkłady wytwarzanych ciągów, zbadamy równomierność otrzymywanych w poszczególnych testach prawdopodobieństw. Na tej podstawie spodziewamy się ustalić warunki stosowalności – głównie minimalne długości badanych ciągów – dla każdego z testów. W drugim etapie zbadamy korelacje pomiędzy otrzymywanymi wartościami p_{val} , na ich podstawie spodziewamy się stwierdzić na ile znacząca jest nadmiarowość wyjściowego zbioru testów oraz które testy można z niego usunąć bez szkody dla kompletności oceny.

Informację na temat przeprowadzenia podobnej procedury można znaleźć w NIST Special Publication 800-22,⁷ jednak szczątkowość opisu, specyficzne kryteria doboru testów do pakietu, jak również wyniki analizy dołączonej dokumentacji zastosowanych testów, przemawiają za wykonaniem zamierzonej przez nas pracy.

Na podstawie wcześniejszych punktów można stwierdzić, że w praktyce nie jest możliwe stwierdzenie losowego, czy nawet pseudolosowego, charakteru badanego ciągu poprzez poddanie go serii testów statystycznych. Wynika to z niewykonalności zastosowania wszystkich testów, nawet tylko tych o złożoności wielomianowej. Zatem powszechna praktyka stosowania pewnego, w ten czy inny sposób ograniczonego, zbioru testów jest z założenia błędna – niestety wydaje się ona być jednocześnie najlepszą z możliwych do zastosowania. Poważnym argumentem przemawiającym za takim właśnie podejściem jest przeświadczenie, że ciąg przechodzący pomyślnie serię różnorodnych testów przechodzić będzie również kolejne. Jednocześnie jednak nie jest możliwe istnienie skończonego, tak zwanego kompletnego, zbioru testów, zdolnych łącznie do wykrycia wszystkich usterek losowości.

Warto zwrócić uwagę, że dokładnie ten sam ułomny mechanizm stosowany jest w konstrukcji algorytmów kryptograficznych, gdzie przez bezpieczeństwo rozumie się odporność na znane ataki i stąd czerpie się nadzieję na ogólną odporność algorytmu. Można powiedzieć, że historia kryptologii sprowadza się do ciągu negacji tych założeń, co nie powstrzymuje nas przed opracowywaniem i wdrażaniem kolejnych rozwiązań.

Podobnie tu nie będziemy zakładać, że pakiet testów, opracowany w ramach realizacji tej pracy jest kompletny i doskonały, mamy jednak nadzieję, że proste zestawienie go z którymkolwiek z powszechnie dostępnych przekona czytelnika o jego wyższości.

⁷ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001, s. 104-105

Rozdział II

Testy losowości

W rozdziale tym zestawiliśmy opisy analityczne, zebranych na potrzeby tej pracy, testów statystycznych wykorzystywanych do badania losowości ciągów liczbowych i binarnych.

Testy pogrupowaliśmy według sposobu, w jaki interpretują one badany ciąg, wyróżniając pięć grup:

- procedury ogólne, które wprost nie muszą stanowić testu losowości, ale stanowią często element składowy testów losowości,
- testy dla ciągów bloków bitów,
- testy dla ciągów nienakładających się wektorów t -wymiarowych, $t > 1$,
- testy dla ciągów nakładających się wektorów t -wymiarowych, $t > 1$,
- testy badające ciągi bitów.

Zdajemy sobie sprawę, że w części przypadków przyporządkowanie ma charakter arbitralny, a granica pomiędzy poszczególnymi grupami bywa umowna. W takich przypadkach kierowaliśmy się ideą (własnością, miarą) stojącą za danym testem.

1. Procedury ogólne

Poniżej przedstawimy kilka ogólnych procedur testowych, przy użyciu których skonstruowane jest wiele spośród testów statystycznych prezentowanych w kolejnych rozdziałach. W większości omówione tu testy przeznaczone są dla ciągów liczb rzeczywistych U^n .

1.1. Test Pearsona – zgodności chi-kwadrat

Test służy do weryfikowania zgodności empirycznego rozkładu obserwowanej zmiennej losowej z założonym dla niej rozkładem teoretycznym. Został sformułowany przez Karla Pearsona w 1900 roku.

Pearson pokazał, że statystyka testowa postaci:

$$S_{obs} = \sum_{i \in L} \left(\frac{O_i - E_i}{\sigma_i} \right)^2,$$

gdzie

L – zbiór pomiarów,

O_i – wartości mierzone,

E_i i σ_i – odpowiadające im wartości oczekiwane i odchylenia standardowe,

ma asymptotycznie rozkład chi-kwadrat, a liczba stopni swobody równa jest liczności zbioru L pomniejszonej o liczbę parametrów rozkładu wyznaczanych na podstawie wykonywanego pomiaru.

W szczególności, gdy mierzonymi wartościami są zliczenia, przyjmuje się, że ich odchylenia standardowe równe są $\sqrt{E_i}$ i wówczas powyższa formuła przyjmuje postać:

$$S_{obs} = \sum_{i \in L} \frac{(O_i - E_i)^2}{E_i}.$$

S_{obs} ma asymptotycznie rozkład chi-kwadrat z $\#L - 1$ stopniami swobody. Aby przybliżenie to było dostatecznie dobre, w zależności od źródła, podaje się, że E_i powinny być nie mniejsze od 5, 10 lub nawet 20, przy czym nie jest znane żadne ścisłe wyprowadzenie tej wielkości. Powszechną praktyką, pozwalającą spełnić ten warunek przy stosunkowo małych licznosciach próby jest grupowanie pomiarów w kategorii.

Źródła podają również, iż liczba klas powinna zależeć od licznosci próbki w funkcji $n^{2/5}$.⁸

1.2. Test Kołmogorowa-Smirnowa

Dokonujemy n niezależnych obserwacji losowej wielkości X , otrzymując wartości x_i dla $i = 1..n$. Funkcję

$$F_n(x) = \frac{\#\{i : x_i \leq x, i = 1..n\}}{n},$$

nazywać będziemy dystrybuantą empiryczną.

Test Kołmogorowa-Smirnowa opiera się na fakcie, że jeśli obserwowana zmienna losowa X ma założony rozkład, to wraz ze wzrostem n dystrybuanta empiryczna $F_n(x)$ będzie coraz lepiej aproksymować dystrybuantę teoretyczną $F(x)$.

Jedynym praktycznym ograniczeniem testu jest wymóg by $F(x)$ nie miała skoków.

W najbardziej rozpowszechnionej wersji test wykorzystuje następujące dwie statystyki:

$$K_n^+ = \sqrt{n} \max_{-\infty < x < +\infty} \{F_n(x) - F(x)\},$$

$$K_n^- = \sqrt{n} \max_{-\infty < x < +\infty} \{F(x) - F_n(x)\}.$$

K_n^+ jest miarą największego dodatniego, zaś K_n^- – największego ujemnego odchylenia dystrybuanty empirycznej od teoretycznej. Obie te zmienne losowe mają identyczny rozkład, określony formułami:

$$\begin{aligned} \Pr\left(K_n^{\pm} \leq \frac{t}{\sqrt{n}}\right) &= \frac{t}{n} \sum_{0 \leq k \leq t} \binom{n}{k} (k-t)^k (t+n-k)^{n-k-1} \\ &= 1 - \frac{t}{n} \sum_{t < k \leq n} \binom{n}{k} (k-t)^k (t+n-k)^{n-k-1}, \end{aligned}$$

⁸ M.G. Kendall, A. Stuart, *The Advanced Theory of Statistics* 2. Ch. Griffin & Company Limited, London, 1973

dla dużych n stosuje się również wzór asymptotyczny:

$$\Pr(K_n^\pm \leq \tau) = 1 - e^{-2\tau^2} \left(1 - \frac{2\tau}{9n} \left(3\sqrt{n} - \tau(3 - 2\tau^2) - \frac{2}{15n} (10\tau^4 - 19\tau^2 + 3) \right) + O(n^{-2}) \right), \tau \geq 0.^9$$

Knuth w *Sztuce programowania*¹⁰ podaje formuły pozwalające obliczać K_n^+ i K_n^- , na podstawie tego, że $F_n(x)$ rośnie w skończonej liczbie skoków. Jeśli dokonamy uporządkowania ciągu obserwacji x_i w kolejności niemalejącej, to statystyki dane są wówczas wzorami:

$$\begin{aligned} K_n^+ &= \sqrt{n} \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i) \right\}, \\ K_n^- &= \sqrt{n} \max_{1 \leq i \leq n} \left\{ F(x_i) - \frac{i-1}{n} \right\} \\ &= \sqrt{n} \left(\frac{1}{n} - \min_{1 \leq i \leq n} \left\{ \frac{i}{n} - F(x_i) \right\} \right). \end{aligned}$$

Możliwe jest, kosztem otrzymania przybliżonych wartości statystyk, wyeliminowanie czasochłonnego sortowania.¹¹

Spotykana jest również postać tego testu wykorzystująca statystyki testowe nie zawierające współczynnika $\sqrt{n}$, które są zwyczajowo oznaczane D_n^+ i D_n^- .

1.3. Test Kołmogorowa

Test ten można uznać za pochodną testu Kołmogorowa-Smirnowa, gdyż jego statystyka testowa może być określona następująco:

$$D_n = \max \{ D_n^+, D_n^- \}.$$

D_n ma bardziej złożony rozkład. W opisie, dołączonej do Test U01, biblioteki statystycznej¹² znaleźć można następującą recepturę wyznaczania prawdopodobieństwa w tym teście:

- jeśli $x \leq 1/2n$, to $p_{val} = 1$,
- jeśli $1/2n < x \leq 1/n$, to $p_{val} = 1 - n!(2x - 1/n)^n$,¹³
- jeśli $n \leq 10$ lub $x < 1/\sqrt{n}$ należy posłużyć się metodą opisaną w [MTW03],¹⁴
- w pozostałych przypadkach wykorzystywane jest przybliżenie:

$$p_{val} \approx (2 - 4x/3)e^{-2nx^2} + (8x/3 - 2)e^{-8nx^2},$$

⁹ P. L'Ecuyer, R. Simard, *ProbDist: A Software Library of Probability Distributions and Goodness-of-Fit Statistics in ANSI C*, Universite de Montreal, wersja 2007, s. 10

¹⁰ D.E. Knuth, *Sztuka programowania, t.2. Algorytmy seminumeryczne*, WNT, 2002, s. 53

¹¹ D.E. Knuth, *Sztuka programowania, t.2. Algorytmy seminumeryczne*, WNT, 2002, s. 605-606

¹² P. L'Ecuyer, R. Simard, *ProbDist: A Software Library of Probability Distributions and Goodness-of-Fit Statistics in ANSI C*, wyd. cyt., s. 10

¹³ J.H. Drew, A.G. Glen, L.M. Leemis, *Computing the Cumulative Distribution Function of the Kolmogorov-Smirnov Statistic*, Computational Statistics and Data Analysis, 34, 2000, s. 1-15

¹⁴ G. Marsaglia, W.W. Tsang, J. Wang, *Evaluating Kolmogorov's Distribution*, Journal of Statistical Software, 8, 18, 2004, s. 1-4

jeśli uzyskana wartość przekracza 0,85 lub $x \leq 0,441 / (\sqrt{n} + 0,275 - 0,04/\sqrt{n})$, to wykonywana jest liniowa interpolacja ogona.¹⁵

1.4. Test Kuipera

Test Kuipera jest kolejną pochodną testu Kołmogorowa-Smirnowa. Statystyka testowa ma postać:

$$K_n = D_n^+ + D_n^-.$$

Prawostronną dystrybuantę możemy obliczyć wykorzystując formułę:

$$\Pr(K_n > \lambda / (\sqrt{n} + 0,155 + 0,24/\sqrt{n})) = Q_{KD}(\lambda),^{16}$$

gdzie $Q_{KD}(\lambda) = 2 \sum_{j=1}^{\infty} (4j^2 \lambda^2 - 1) e^{-2j^2 \lambda^2}$.

Można znaleźć dodatkowe formuły dla niektórych przypadków $\Pr(K_n > \beta)$:

- dla $\beta < 2/n$ obliczane jest:

$$\Pr(K_n > \beta) = 1 - n! \left(\beta - \frac{1}{n} \right)^{n-1},$$

- dla $2/n \leq \beta < 3/n$ obliczane jest:

$$\Pr(K_n > \beta) = 1 - \frac{(a+b)^{n-1}(1-a+b) - (a-b)^{n-1}(1-a-b)}{2b \cdot n^{n-2}} (n-1)!,$$

gdzie $a = \frac{n\beta-1}{2}$ i $b = \sqrt{a^2 - \frac{n\beta-2}{2}}$,

- dla $\beta > 0.5$ i $2 \mid n$ lub $\beta > \frac{n-1}{2n}$ i $2 \nmid n$ obliczane jest:

$$\Pr(K_n > \beta) = \sum_{t=0}^{\lfloor n(1-\beta) \rfloor + 1} T(t) \binom{n}{t} \left(1 - \beta - \frac{t}{n} \right)^{n-t-1},$$

gdzie $T(t) = y^{t-3} \left(ny^3 - \frac{ty^2}{n} \left(3 - \frac{2}{n} \right) - \frac{t(t-1)(t-2)}{n^2} \right)$ i $y = \beta + \frac{t}{n}$,¹⁷

- w pozostałych przypadkach korzystamy z formuły:

$$Q_{KD}(\lambda) = 2 \sum_{j=1}^{\infty} (4j^2 \lambda^2 - 1) e^{-2j^2 \lambda^2} - \frac{8\lambda}{3\sqrt{n}} \sum_{j=1}^{\infty} j^2 (4j^2 \lambda^2 - 3) e^{-2j^2 \lambda^2}.^{18}$$

¹⁵ M.A. Stephens, *Use of the Kolmogorov-Smirnov, Cramer-von Mises and Related Statistics without Extensive Tables*, Journal of the Royal Statistical Society, Series B, 32, 1, 1970, s. 115-122

¹⁶ W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992, s. 627

¹⁷ <https://github.com/aarchiba/kuiper/blob/master/kuiper.py>

¹⁸ `kstest.c` z pakietu Dieharder

Obydwa powyższe warianty testu charakteryzują się większą czułością na odchylenia w ogonach rozkładu, co ma szczególnie duże znaczenie w przypadku badania generatorów losowych, gdy szczególnie interesujące jest zachowanie się otrzymywanych rozkładów wartości p_{val} w pobliżu 0 i 1. Niektóre prace wskazują nawet, że pod tym względem test Kuipera jest mocniejszy od, przedstawionego dalej, testu Andersona-Darlinga. Jest on używany, jako test drugiego poziomu w pakiecie Dieharder.

1.5. Rodzina testów Andersona-Darlinga

Jest to rodzina testów zaproponowana przez T.W. Anderson'a i D.A. Darling'a, będącej rozwinięciem wcześniejszych prac A. Kołmogorowa, W.I. Smirnowa, H. Cramera i R.E. von Misesa.¹⁹

Statystyka testowa ma ogólną postać:

$$W_n^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F(x))^2 \psi(F(x)) dF,$$

gdzie $\psi(t)$ jest funkcją wagi, zaś $F(x)$ i $F_n(x)$ są, jak w poprzednim punkcie, dystrybuantami, odpowiednio, teoretyczną i empiryczną.

W niniejszej pracy przedstawimy dwa przypadki powyższej statystyki, określone przez $\psi(t) = 1$ oraz $\psi(t) = (t(t-1))^{-1}$.

1.6. Test Cramera-von Misesa

Pod tą nazwą znany jest jeden z najstarszych wariantów testu opartego na powyższej statystyce, w którym należy przyjąć $\psi(t) = 1$.

Jeśli badanym ciągiem jest U^n , zaś testowanym rozkładem rozkład równomierny na przedziale $[0;1)$, to statystyka testowa daje się sprowadzić do postaci:

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^n \left(u_{(i)} - \frac{i-0,5}{n} \right)^2,$$

gdzie $u_{(1)}, u_{(2)}, \dots, u_{(n)}$ są elementami wektora U^n uporządkowanymi rosnąco.

Graniczny, przy $n \rightarrow +\infty$, rozkład W_n^2 dany jest następująco:

$$\Pr(W^2 \leq x) = \frac{1}{\pi\sqrt{2}} \sum_{i=0}^{\infty} (-1)^i \sqrt{4i+1} \binom{-0,5}{i} e^{-\frac{(4i+1)^2}{16x}} K_{1/4} \left(\frac{(4i+1)^2}{16x} \right),^{20}$$

gdzie

$$\binom{-0,5}{i} = \frac{(-1)^i \Gamma(i+0,5)}{i! \sqrt{\pi}} \text{ (na podstawie uogólnionego symbolu Newtona),}$$

¹⁹ T.W. Anderson, D.A. Darling, *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes*, Annals of Mathematical Statistics, 23, 2, 1952, s. 193-212

²⁰ T.W. Anderson, D.A. Darling, *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes*, Annals of Mathematical Statistics, 23, 2, 1952, s. 202

$K_{\alpha}(x)$ jest zmodyfikowaną funkcją Bessela drugiego rodzaju.

Dla małych n została wyznaczona empiryczna poprawka.²¹

1.7. Test Andersona-Darlinga

Pod tą nazwą funkcjonuje w literaturze odmiana testu wykorzystująca funkcję wagi postaci:

$$\psi(t) = (t(t-1))^{-1}.$$

Jeśli, jak poprzednio, testowany jest ciąg U^n , dla którego elementów zakładamy, że posiadają rozkład równomierny na przedziale $(0;1)$, to po posortowaniu ich rosnąco możemy obliczyć statystykę testową następująco:

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n \left((2i-1) \ln u_{(i)} + (2n+1-2i) \ln (1-u_{(i)}) \right).^{22}$$

Do obliczania $p_{val} = \Pr(A_n^2 \geq x)$ można posłużyć się metodą opracowaną przez Marsaglię²³.

1.8. Test dla dwóch liczb

Dla dwóch, czy trzech zmiennych losowych nie może być mowy o wykorzystaniu któregośkolwiek z omówionych powyżej testów, czasami jednak zachodzi konieczność by mając dwie lub trzy wartości z przedziału $[0;1]$ przekształcić je w jedną. Prezentowaną procedurę można uogólniać na większe zbiory, lecz jest to bardzo uciążliwe i prostszym rozwiązaniem jest zastosowanie któregoś z testów omówionych powyżej.

Zadane liczby u_1 i u_2 traktujemy jako współrzędne punktu na płaszczyźnie, a następnie obliczamy pole obszaru będącego częścią wspólną kwadratu jednostkowego – $(0;0):(1;1)$ i koła o środku $(0;0)$ i promieniu $\sqrt{u_1^2 + u_2^2}$.

Wartość odpowiedzi obliczamy w następujący sposób:

- obliczamy $r^2 = u_1^2 + u_2^2$,
- jeśli $r^2 \leq 1$, to poszukiwaną wartością jest

$$p = \pi r^2 / 4,$$

- jeśli $r^2 > 1$, to obliczamy kąt $\gamma = 2 \arccos(1/r)$, a poszukiwaną wartością jest

$$p = \frac{r^2}{2} \left(\frac{\pi}{2} - \gamma + \sin \gamma \right).$$

²¹ P. L'Ecuyer, R. Simard, *ProbDist: A Software Library of Probability Distributions and Goodness-of-Fit Statistics in ANSI C*, wyd. cyt., s. 11

²² P. L'Ecuyer, R. Simard, *ProbDist: A Software Library of Probability Distributions and Goodness-of-Fit Statistics in ANSI C*, wyd. cyt., s. 28

²³ G. Marsaglia, J.C.W. Marsaglia, *Evaluating the Anderson-Darling Distribution*, Journal of Statistical Software, 9, 2, 2004, s. 1-5

Jeżeli u_1 i u_2 są niezależnymi zmiennymi losowymi o rozkładzie równomiernym na przedziale $[0;1]$, to p również ma rozkład równomierny na przedziale $[0;1]$.

1.9. Test dla trzech liczb

Zadane liczby u_1, u_2, u_3 traktujemy jako współrzędne punktu w przestrzeni trójwymiarowej, a następnie objętość obszaru będącego częścią wspólną sześcianu jednostkowego $(0;0;0):(1;1;1)$ i kuli o środku $(0;0;0)$ i promieniu $\sqrt{u_1^2 + u_2^2 + u_3^2}$.

Wartość odpowiedzi obliczamy w następujący sposób:

- obliczamy $r = \sqrt{u_1^2 + u_2^2 + u_3^2}$,
- jeśli $r \leq 1$, to poszukiwaną wartością jest

$$p = \pi r^3 / 6,$$

- jeśli $1 < r \leq \sqrt{2}$, to poszukiwaną wartością jest

$$p = (-4r^3 + 9r^2 - 3)\pi/12,$$

- jeśli $r > \sqrt{2}$, to poszukiwaną wartością jest

$$p = 1 - \frac{\pi(\sqrt{3} - r)^3}{6} \cdot J,$$

gdzie $J = \frac{12 - \pi(15 - 8\sqrt{2})}{2\pi(\sqrt{3} - \sqrt{2})^3} \approx 2,077781119$ jest poprawką.

Formuła określająca poprawkę nie jest dokładna, w rzeczywistości nie jest ona nawet przybliżeniem prawdziwej. Zdecydowaliśmy się na jej użycie z trzech powodów:

- oddaje charakter rzeczywistej formuły,
- jest od niej zdecydowanie prostsza,
- dotyczy zaledwie 3,5% przypadków (dokładnie $1 - (15 - 8\sqrt{2})\pi/12$).

W związku z zastosowanym wybiegiem nie można, jak w poprzednim punkcie, sformułować twierdzenia o rozkładzie otrzymywanych wartości p , tym niemniej łatwo zobaczyć, że musi on być bliski rozkładowi równomiernemu na przedziale $[0;1]$.

2. Testy dla ciągów bloków bitów

Opisane w tym punkcie testy przeznaczone są do badania ciągów bloków d -bitowych, choć część z nich zaprojektowana została dla ciągów liczb rzeczywistych z przedziału $[0;1]$, to przy zachowaniu pewnej ostrożności możliwe jest ich wykorzystanie poprzez przejście pomiędzy tymi notacjami.

Podstawową stosowaną przez nas dla tych testów operacją jest zbudowanie ciągu d -bitowych bloków $B^{d,n} = b_1, b_2, \dots, b_n$ z badanego ciągu binarnego s^{dn} , przez co będziemy rozumieć wyodrębnienie ciągu nienakładających się, następujących po sobie d -bitowych fragmentów ciągu s^{dn} :

$$B^{d,n} = \underbrace{(s_1, s_2, \dots, s_d)}_{b_1}, \underbrace{(s_{d+1}, s_{d+2}, \dots, s_{2d})}_{b_2}, \dots, \underbrace{(s_{d(n-1)+1}, s_{d(n-1)+2}, \dots, s_{nd})}_{b_n}.$$

Również transformacja ciągu liczbowego U^n na ciąg bloków przebiega bardzo intuicyjnie:

$$B^{d,n} = \lfloor u_1 2^d \rfloor, \lfloor u_2 2^d \rfloor, \dots, \lfloor u_n 2^d \rfloor.$$

Problematyczna jest natomiast próba uzyskania ciągu liczb rzeczywistych w oparciu o ciąg bloków. Pomijając, nierealizowalną na komputerach cyfrowych, nieskończoną precyzję liczb rzeczywistych, należy zwrócić uwagę, że wykonanie przekształcenia odwrotnego do powyższego, czyli wyznaczenie ciągu ilorazów:

$$\frac{b_1}{2^d}, \frac{b_2}{2^d}, \dots, \frac{b_n}{2^d},$$

nie tylko nie realizuje przekształcenia na przedział $[0;1)$, ale również, dla mniejszych d , nie pozwoli uzyskać rozkładu zbliżonego do równomiernego na przedziale $[0;1)$, przez co wypaczone zostaną wyniki testów. W ogólności rekomendujemy korzystanie z ciągu ilorazów postaci:

$$\frac{b_1}{2^d - 1}, \frac{b_2}{2^d - 1}, \dots, \frac{b_n}{2^d - 1}$$

i uwzględnienie w teście skończonej długości bloku albo stosowanie wystarczająco dużych d , by usprawiedliwione było założenie ciągłości rozkładu.

2.1. Testy bazujące na procedurach ogólnych

2.1.1. Test częstości

Test polega na zbadaniu równomierności rozkładu bloków d -bitowych przy użyciu testu zgodności chi-kwadrat dla zliczeń (patrz punkt II.1.1.).

Test ten został przez nas potraktowany jako szczególny przypadek testu zgodności Pearsona dla wektorów nienakładających się (patrz punkt II.3.1.).

2.1.2. Test zgodności rozkładu

Testy opisane w punktach 1.2. do 1.7. można z powodzeniem wykorzystać do testowania ciągów binarnych. Ich charakterystykom został poświęcony rozdział IV, zaś w rozdziale V zbadaliśmy zakres ich używalności w badaniu ciągów binarnych.

2.2. Testy momentów

Jest to rodzina powszechnie znanych testów, służących do sprawdzania hipotez dotyczących wartości kolejnych momentów rozkładu, z którego pochodzi badana próbka. Zwykle spotyka się te-

sty badające momenty pierwszego i drugiego rzędu, odpowiada to badaniu wartości oczekiwanej i wariancji.

2.2.1. Test momentu pierwszego rzędu dla ciągu liczb

Dla ciągu liczb $U^n = u_1, u_2, \dots, u_n$ obliczamy średnią arytmetyczną jego elementów:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n u_i.$$

Przy założeniu równomierności i niezależności rozkładu poszczególnych u_i , dystrybucja zmiennej losowej $\bar{X}_n$ przybliżona jest następująco:

$$\Pr(\bar{X}_n < x) = \Phi\left(\left(x - \frac{1}{2}\right)\sqrt{12n}\right) - \frac{1}{20n} \varphi^{(3)}\left(\left(x - \frac{1}{2}\right)\sqrt{12n}\right),^{24}$$

gdzie $\Phi(x)$ jest dystrybuantą rozkładu $N(0;1)$, a $\varphi^{(3)}(x) = \frac{x(3-x^2)}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ jest trzecią pochodną

funkcji gęstości tego rozkładu. Przybliżenie to jest wystarczająco dokładne dla n rzędu 10^3 , przy czym, co można zauważyć porównując z kolejnym punktem, istotna jest dokładność arytmetyki komputera, w której wykonywane są obliczenia.

2.2.2. Test momentu pierwszego rzędu dla ciągu bitów

Ciąg bloków $B^{d,n} = b_1, b_2, \dots, b_n$, dla $d \geq 1$, które traktujemy jako binarne reprezentacje liczb naturalnych. Jeśli wartość i -tego bloku oznaczmy jako b_i , to przyjmując równomierność i niezależność rozkładu bitów w badanym ciągu, możemy uznać, że b_i tworzą ciąg niezależnych zmiennych losowych o rozkładzie równomiernym na zbiorze $\{0, 1, \dots, 2^d - 1\}$.

Choć statystyka testowa określona jest identycznie jak w teście dla ciągu liczb:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n b_i,$$

to jednak, w związku ze zmianą zbioru, a zwłaszcza jego dyskretnością zmianie ulega dystrybucja:

$$\Pr(\bar{X}_n < x) = \Phi\left(\left(x - \frac{2^d - 1}{2}\right)\sqrt{\frac{12n}{2^{2d} - 1}}\right) - \frac{2^{2d} + 1}{2^{2d} - 1} \frac{1}{20n} \varphi^{(3)}\left(\left(x - \frac{2^d - 1}{2}\right)\sqrt{\frac{12n}{2^{2d} - 1}}\right).^{25}$$

Dla $d > 5$ można, z błędem poniżej 1%, stosować przybliżenie rozkładem ciągłym, dzieląc każdy b_i przez $2^d - 1$.

²⁴ R. Wieczorkowski, R. Zieliński, *Komputerowe generatory liczb losowych*, WNT, Warszawa, 1997, s. 118

²⁵ K. Mańk, *Dokładne dystrybuanty statystyk w testach momentów 1 i 2 rzędu*, Biuletyn WAT, Kryptologia IV, Warszawa, 2004, s. 51

2.2.3. Test momentu drugiego rzędu dla ciągu liczb

Analogicznie można zbudować test dla momentu drugiego rzędu, obliczając średnią z kwadratów elementów badanego ciągu liczb:

$$\bar{X}_n^2 = \frac{1}{n} \sum_{i=1}^n u_i^2.$$

Statystyka testowa ma rozkład określony następująco:

$$\Pr(\bar{X}_n^2 < x) = \Phi\left(\left(x - \frac{1}{3}\right)\sqrt{\frac{45n}{4}}\right) - \frac{0,1065}{\sqrt{n}} \varphi^{(2)}\left(\left(x - \frac{1}{3}\right)\sqrt{\frac{45n}{4}}\right),^{26}$$

gdzie $\varphi^{(2)}(x) = \frac{x^2 - 1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ jest drugą pochodną standardowego rozkładu normalnego.

2.2.4. Test momentu drugiego rzędu dla ciągu bitów

Podobnie, jak dla testu momentu pierwszego rzędu, wyznaczamy statystykę testową:

$$\bar{X}_n^2 = \frac{1}{n} \sum_{i=1}^n b_i^2,$$

której rozkład dla $n \geq 10^3$ może być opisany formułą:

$$\begin{aligned} \Pr(\bar{X}_n^2 < x) = & \Phi(y) - \frac{(2^d - 2)(2^d + 2)(32 \cdot 2^d - 47)}{(8 \cdot 2^d - 11)\sqrt{(2^d - 1)(2^d + 1)(2 \cdot 2^d - 1)(8 \cdot 2^d - 11)}} \cdot \frac{\sqrt{5}}{21\sqrt{n}} \cdot \varphi^{(2)}(y) + \\ & - \frac{(128 \cdot 2^{5d} - 506 \cdot 2^{4d} + 1430 \cdot 2^{3d} - 1745 \cdot 2^{2d} - 1258 \cdot 2^d + 2581)}{(2^d - 1)(2^d + 1)(2 \cdot 2^d - 1)(8 \cdot 2^d - 11)^2} \cdot \frac{1}{28n} \cdot \varphi^{(3)}(y), \end{aligned}$$

gdzie $\Phi(x)$, $\varphi^{(2)}(x)$ oraz $\varphi^{(3)}(x)$ jak wyżej, zaś

$$y = \left(6x - (2^d - 1)(2 \cdot 2^d - 1)\right) \sqrt{\frac{5n}{(2^d - 1)(2^d + 1)(2 \cdot 2^d - 1)(8 \cdot 2^d - 11)}}.^{27}$$

Jeśli badane są bardzo długie ciągi, to nawet dla dużych wartości d nie należy stosować przybliżenia rozkładem ciągłym.

2.3. Test średniej

Dla ciągu U^n wyznaczana jest suma

$$\bar{u}_n = \sum_{i=1}^n u_i,$$

której rozkład opisuje formuła:

$$\Pr(\bar{u}_n \leq z) = \frac{1}{n!} \sum_{i=0}^{\lfloor z \rfloor} (-1)^i \binom{n}{i} (z - i)^n, \quad 0 \leq z \leq n.$$

²⁶ R. Wieczorkowski, R. Zieliński, *Komputerowe generatory liczb losowych*, WNT, Warszawa, 1997, s. 118

²⁷ K. Mańk, *Dokładne dystrybuanty statystyk w testach momentów 1 i 2 rzędu*, wyd. cyt., s. 61

Procedura testowa polega na przeprowadzeniu testu zgodności dla zbioru średnich otrzymanego dla N ciągów U^n , przez co przebieg tego testu znacząco różni się od testu momentu pierwszego rzędu.

2.4. Test autokorelacji

Test polega na obliczeniu współczynnika korelacji pomiędzy badanym ciągiem, a jego cyklicznym przesunięciem o t pozycji. Statystykę testową (współczynnik korelacji)

$$S_{obs} = \frac{n \sum_{i=1}^n u_i u_{i+t} - \left(\sum_{i=1}^n u_i \right)^2}{n \sum_{i=1}^n u_i^2 - \left(\sum_{i=1}^n u_i \right)^2}$$

traktuje się jako zmienną losową o rozkładzie normalnym, z wartością oczekiwaną

$$\mu = -1/(n-1)$$

i wariancją

$$\sigma^2 \approx \frac{n^2}{(n-1)^2(n-2)} - \frac{24}{5n^2}.$$

Należy mieć na uwadze, że dla $t = n/2$ wariancja ulega zmianie wobec dublowania się składników w jednej z sum określających statystykę testową:

$$\sum_{i=1}^n u_i u_{i+n/2} = 2 \sum_{i=1}^{n/2} u_i u_{i+n/2}.$$

2.5. Test korelacji

Jest to jedna z odmian testu autokorelacyjnego. Test wyznacza korelację między ciągiem U^n , a jego niecyklicznym przesunięciem o k pozycji:

$$\hat{\rho}_k = \frac{1}{n-k} \sum_{i=1}^{n-k} u_i u_{i+k} - \frac{1}{4}.$$

Dla dużych n zmienna losowa $\hat{\rho}_k \sqrt{12(n-k)}$ ma w przybliżeniu rozkład $N(0;1)$.

2.6. Test iloczynów

Ciąg U^n dzielimy na n rozłącznych fragmentów długości t i dla każdego z nich obliczamy iloczyn elementów

$$v_i = u_{(i-1)t+1} \cdot u_{(i-1)t+2} \cdot \dots \cdot u_{it}, \quad i = 1..n.$$

Rozkład otrzymanych iloczynów porównywany jest przy pomocy testu Andersona-Darlinga z rozkładem teoretycznym:

$$\Pr(v_i \leq x) = x \sum_{j=0}^{t-1} \frac{(-\ln x)^j}{j!}, \quad 0 \leq x \leq 1.$$

Przy wielokrotnym powtarzaniu testu autorzy proponują przeprowadzenie testu zgodności Andersona-Darlinga dla otrzymanych p_{val} .

2.7. Test sumy logarytmów

Dla ciągu U^n obliczana jest statystyka:

$$S_{obs} = -\sum_{i=1}^n \ln u_i,$$

która jako suma n niezależnych zmiennych losowych o rozkładzie wykładniczym ma rozkład chi-kwadrat z $2n$ stopniami swobody.

Test ten jest równoważny testowi iloczynów przeprowadzonemu dla pojedynczego ciągu.

2.8. Test częstości dla podprzedziału

Dla zadanego przedziału $[\alpha; \beta)$, gdzie $0 \leq \alpha < \beta \leq 1$ zliczamy W_i , $i = 1..n$, liczbę elementów t -elementowego fragmentu ciągu U^m , które do tego przedziału należą. Każde W_i ma rozkład Bernoulliego z parametrami t oraz $p = \beta - \alpha$. Zaobserwowany rozkład W_i jest następnie porównywany z teoretycznym przy użyciu testu zgodności chi-kwadrat.

2.9. Test kolizji indeksów maksymalnych elementów

Każdemu t -elementowemu fragmentowi ciągu U^m przyporządkowujemy J_i , $i = 1..n$ – numer pozycji maksymalnego elementu. Następnie dla ciągu J_i przeprowadzamy test kolizji, opisany w punkcie II.2.23.

Jeśli procedura testowa zostanie powtórzona wiele razy, to dla zaobserwowanego zbioru liczby kolizji można przeprowadzić test zgodności chi-kwadrat.

2.10. Test sum do g

Ciąg U^n dzielony jest na nienakładające się, następujące po sobie fragmenty minimalnej długości, przy czym suma elementów każdego fragmentu musi przekraczać założoną z góry wartość g . Ciąg J_i , $i = 1..N$, długości tych fragmentów jest następnie badany przy użyciu testu zgodności chi-kwadrat.

Rozkład teoretyczny dany jest następująco:

$$\Pr(J = j) = \frac{j - g + 1}{(j + 1)!} \sum_{r=0}^{\lceil g \rceil} (-1)^r \binom{j+1}{r} (g-r)^j, \quad 1 \leq g \leq j \in \mathbb{N}.^{28}$$

Formułę tę można efektywnie obliczać wykorzystując poniższe zależności:

$$\Pr_g(J = j) = \frac{j - g + 1}{j + 1} \left(\Pr_{g-1}(J = j - 1) + \frac{g}{j - g} \Pr_g(J = j - 1) \right), \quad 1 \leq g < j,$$

²⁸ G. Ugrin-Sparac, *Stochastic Investigation of Pseudo-Random Number Generators*, Computing 46, 1991, s. 54

$$\Pr_g(J = g) = \frac{1}{(g+1)!}, \quad 0 \leq g \in \mathbb{N}.^{29}$$

2.11. Test odstępów

W teście bada się długość kolejnych fragmentów $u_i, u_{i+1}, \dots, u_{i+r}$, w których tylko $u_{i+r} \in [\alpha; \beta)$, dla ustalonych liczb rzeczywistych $0 \leq \alpha < \beta \leq 1$. Fragment taki określa lukę długości r pomiędzy dwoma kolejnymi wystąpieniami liczby z przedziału $[\alpha; \beta)$. Aby określić długość ostatniej luki, należy ciąg potraktować jako cykliczny, czyli po dojściu do ostatniej liczby przejść na początek i znaleźć odległość do pierwszego wystąpienia liczby z przedziału $[\alpha; \beta)$ w ciągu. Dla powstałego w ten sposób ciągu luk buduje się test zgodności chi-kwadrat, wykorzystując wzory określające prawdopodobieństwo, że luka będzie miała długość r :

$$p_r = p(1-p)^r$$

oraz prawdopodobieństwo, że luka będzie miała długość co najmniej t

$$p_t = (1-p)^t,$$

gdzie $p = \beta - \alpha$, jest prawdopodobieństwem, że element ciągu należy do przedziału $[\alpha; \beta)$.

Do określenia oczekiwanej liczby luk długości r należy posłużyć się formułą Np_r , gdzie N jest zaobserwowaną liczbą luk, nie zaś $np p_r$, choć np jest wartością oczekiwaną liczby luk w ciągu długości n .

2.12. Test sumy logarytmów odstępów

Określamy ciąg $U^{(n,m)} = u_{(0)}, u_{(1)}, u_{(2)}, \dots, u_{(n)}, u_{(n+1)}, \dots, u_{(n+m)}$, w którym $u_{(0)} = 0$ i jeśli $m > 0$, to $u_{(n+i)} = 1 + u_{(i-1)}$ dla $i = 1..m$, zaś $u_{(1)}, u_{(2)}, \dots, u_{(n)}$ jest posortowanym niemalejąco ciągiem $U^n = u_1, u_2, \dots, u_n$. Zachodzi $u_{(0)} \leq u_{(1)} \leq \dots \leq u_{(n+m)}$.

Na podstawie ciągu $U^{(n,m)}$ definiujemy ciąg odstępów:

$$G_m^{n+1} = g_0, g_1, \dots, g_n,$$

gdzie

$$g_i = u_{(i+m)} - u_{(i)}, \quad i = 0..n.$$

Określamy statystyki testowe:

$$L_{m,n}^c = \sum_{i=0}^n \ln(n \cdot g_i) = (n+1) \ln n + \sum_{i=0}^n \ln(g_i) = (n+1) \ln n + \ln \left(\prod_{i=0}^n g_i \right),$$

²⁹ G. Ugrin-Sparac, *On a Distribution Encountered in the Renewal Process Based on Uniform Distribution*, Glasnik Matematicki, vol. 25 (45), 1990, s. 225

$$L_{m,n} = \sum_{i=0}^{n-m+1} \ln(n \cdot g_i),$$

odpowiadające odpowiednio cyklicznemu zapętlению ciągu i jego brakowi. Jak pokazaliśmy powyżej dla $L_{m,n}^c$ statystyki te dają się efektywnie obliczać dzięki własnościom funkcji logarytm.

Obie statystyki mają w przybliżeniu rozkład normalny z parametrami:

$$E[L_{m,n}^c] = -(n+1)(R(m,n) - \ln n),$$

$$D^2[L_{m,n}^c] = (n+1)(Q(m,n) + n \cdot Q(1,n) - 2(m-1)(m \cdot Q(1,m-1) + 1) + (2m(m-1) - n)\pi^2/6),^{30}$$

$$E[L_{m,n}] = -(n-m+2)(R(m,n) - \ln n),$$

$$D^2[L_{m,n}] \approx D^2[L_{m,n}^c],^{31}$$

gdzie

$$R(j,k) = \sum_{i=j}^k 1/i,$$

$$Q(j,k) = \sum_{i=j}^k 1/i^2.$$

Jeśli ciąg liczb $U^n = u_1, u_2, \dots, u_n$ wytwarzany jest przez generator o precyzji d bitów, to należy przyjąć, że $\ln 0 = -(d+1)\ln 2$.

2.13. Test sumy kwadratów odstępów

Różnica w stosunku do poprzedniej pary testów polega jedynie na zmianie statystyk testowych na następujące:

$$S_{m,n}^c = \sum_{i=0}^n (n \cdot g_i)^2 \text{ oraz } S_{m,n} = \sum_{i=0}^{n-m+1} (n \cdot g_i)^2,$$

które również mają asymptotycznie rozkłady normalne, z parametrami:

$$E[S_{m,n}^c] = \frac{m(m+1)n^2}{n+2},$$

$$D^2[S_{m,n}^c] = \frac{2m(m+1)n^4(n(2m+1) - 3m^2 + m + 2)}{3(n+2)^2(n+3)(n+4)},$$

$$E[S_{m,n}] = \frac{m(m+1)(n-m+2)n^2}{(n+1)(n+2)},$$

$$D^2[S_{m,n}] = \frac{m(m+1)n^4}{3(n+1)^2(n+2)^2(n+3)(n+4)}.$$

³⁰ P. L'Ecuyer, *Tests Based on Sum-Functions of Spacings for Uniform Random Numbers*, Journal of Statistical Computation and Simulation, 59, 1997, s. 256

³¹ L. Holst, *Asymptotic Normality of Sum-Functions of Spacings*, The Annals of Probability, 7, 6, s. 1066-1072

$$\left(\begin{aligned} &2n^3(2m+1) + n^2(8m^3 - 15m^2 + 3m + 16) + n(-12m^4 + 48m^3 - 27m^2 - 37m + 34) + \\ &-30m^4 + 58m^3 + 6m^2 - 54m + 20 \end{aligned} \right).^{32}$$

2.14. Test odstępów dni urodzin

W ramach roku o długości M dni wybieramy, na podstawie badanego ciągu, n dat urodzin, które następnie sortujemy w kolejności niemalejącej, tworząc w ten sposób ciąg: I_i , $i = 1..n$. Z kolei dla tego ciągu wyznaczamy ciąg odstępów postaci: $I_1, I_2 - I_1, I_3 - I_2, \dots, I_n - I_{n-1}$.

Niech J będzie liczbą wartości, które występują w ciągu odstępów więcej niż raz, wówczas J ma asymptotycznie rozkład Poissona z parametrem $\lambda = n^3/4M$.

2.15. Test największej liczby elementów w przedziale

Statystyka testowa równa jest największej liczbie elementów ciągu $U^{(n,0)}$ mieszczącej się w przedziale o długości D :

$$S_D = \max_{i=0..n-1} \max_{j=i..n} \{j - i + 1 : u_{(j)} - u_{(i)} \leq D\}.$$

W zależności od uzyskanej wartości statystyki, żądane prawdopodobieństwo można obliczyć za pomocą jednej z poniższych formuł:

- dla $m \geq \left\lfloor \frac{n+1}{2} \right\rfloor$ i $D \leq 0,5$

$$\Pr(S_D \geq m) = 2 \sum_{i=m+1}^n \binom{n}{i} D^i (1-D)^{n-i} + (D(1-D) - n + m + 1) \binom{n}{m} D^m (1-D)^{n-m},$$

- dla $m \geq \left(\left\lfloor \frac{n+1}{2} \right\rfloor - nD \right) / \sqrt{n}$

$$\Pr\left(\frac{S_D - nD}{\sqrt{n}} \geq m\right) \approx \frac{m}{D\sqrt{D(1-D)}} \varphi\left(\frac{m}{\sqrt{D(1-D)}}\right) + 2\left(1 - \Phi\left(\frac{m}{\sqrt{D(1-D)}}\right)\right),^{33}$$

- dla $m \leq (n+1)D$ można przyjąć $\Pr(S_D \geq m) = 1$.

W dokumentacji projektu „SSJ: Simulation Stochastique en Java”³⁴ znaleźć można również taką postać pierwszej z formuł:

$$\Pr(S_D \geq m) = 2 \sum_{i=m}^n \binom{n}{i} D^i (1-D)^{n-i} + \left(\frac{m}{D} - n - 1\right) \binom{n}{m} D^m (1-D)^{n-m},$$

która w przeciwieństwie do oryginalnej daje wyłącznie dodatnie wartości.

³² P. L’Ecuyer, dz. cyt., s. 256-257

³³ N. Cressie, *The Asymptotic Distribution of the Scan Statistic under Uniformity*, Annals of Probability, 8, 4, 1980, s. 833

³⁴ R. Simard, *SSJ User’s Guide – Package GoF*, 2008, s. 9



Stwierdziliśmy, że przytoczone powyżej formuły są, z naszego punktu widzenia, bezużyteczne. Obie nie pozwalają wyznaczać prawdopodobieństw dla pełnego zakresu wartości statystyki testowej. W pierwszym przypadku wynika to z ograniczeń wskazanych przez autorów, w drugim z jej niedokładności, prowadzącej do uzyskiwania, dla części przypadków, wartości prawdopodobieństwa większych od 1. W ramach realizacji pracy p.k. ROTOR^{35,36}, w związku z niedostatkami badania przy użyciu podstawowego testu, zaproponowaliśmy wykonywanie testu w postaci zintegrowanej procedury dwupoziomowej, w której test drugiego poziomu bada zgodność zaobserwowanego rozkładu wartości statystyki testowej z rozkładami wyznaczonymi w ramach wykonanych opracowań i zapisanymi w postaci tablic stałych.

2.16. Uniwersalny test Maurera

Test jest określony przez trzy parametry d , Q i K .

Ciąg bloków $B^{d,n} = b_1, b_2, \dots, b_n$, dla $d \geq 1$, dzielony jest na dwie części. Pierwsza z nich, o długości Q bloków, wykorzystywana jest do zainicjalizowania wykorzystywanego w teście wektora T o długości 2^d , zawierającego ostatnie zanotowane położenie każdego z możliwych 2^d bloków długości d . K jest natomiast długością sekwencji testowej, przyjmuje się $K = \lfloor n/d \rfloor - Q \geq 33 \cdot 2^d$, na podstawie której obliczana jest statystyka testowa postaci

$$S_{obs} = \frac{1}{K} \sum_{i=Q+1}^{Q+K} \lg(i - T_{b_i}).$$

Maurer pokazał, że S_{obs} ma asymptotycznie rozkład normalny, z wartością oczekiwaną

$$\mu_d = E[\lg(i - T_{b_i})] = 2^{-d} \sum_{j=1}^{\infty} (1 - 2^{-d})^{j-1} \lg j$$

i odchyleniem standardowym

$$\sigma_d = c \cdot \sqrt{\frac{D_d^2 [\lg(i - T_{d_i})]}{K}},$$

gdzie

$$D_d^2 [\lg(i - T_{d_i})] = 2^{-d} \sum_{j=1}^{\infty} (1 - 2^{-d})^{j-1} (\lg j)^2 - \mu_d^2,^{37}$$

$$c = \sqrt{h(d) + \frac{e(d)}{K}}.^{38}$$

³⁵ K. Mańk, *Implementacja i analiza testu największej liczby elementów w przedziale*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2019

³⁶ K. Mańk, *Dostosowanie testu największej liczby elementów w przedziale do wykorzystania w pakiecie testów losowości*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2019

³⁷ U.M. Maurer, *A Universal Statistical Test for Random Bit Generators*, Journal of Cryptology, 5, 1992, Springer-Verlag, s. 89-105

³⁸ J.S. Coron, D. Naccache, *An Accurate Evaluation of Maurer's Universal Test*, Proceedings of SAC '98 (Lecture Notes in Computer Science), Springer-Verlag, 1998, s. 57-71



Stablicowane wartości μ_d , D_d^2 , $h(d)$, $e(d)$ dla $d = 1..16$ znaleźć można np. w [Ma08].³⁹

2.17. Test Corona

J. Coron rozważał uogólnienie testu Maurera polegające na wprowadzeniu dowolnej funkcji g do statystyki testowej:

$$S_{obs}^C = \frac{1}{K} \sum_{i=Q+1}^{Q+K} g(i - T_{b_i}).^{40}$$

W szczególności pokazał, że funkcja

$$g(k) = \frac{1}{\ln 2} \sum_{i=1}^{k-1} \frac{1}{i}$$

pozwała uzyskać statystykę bliższą entropii źródła badanego ciągu.

Tak jak w poprzednim przypadku, Coron założył, że statystyka testowa ma w przybliżeniu rozkład normalny z wartością oczekiwaną $\mu_d^C = d$ i odchyleniem standardowym

$$\sigma_d^C = c^C \cdot \sqrt{\frac{D_d^2[g(i - T_{d_i})]}{K}},$$

gdzie $c^C = \sqrt{h^C(d) + \frac{e^C(d)}{K}}$, zaś stablicowane wartości $D_d^2[g(i - T_{d_i})]$, $h^C(d)$ i $e^C(d)$ dla $d = 3..16$ znaleźć można w oryginalnej pracy.⁴¹

2.18. Stos książek

W tym punkcie omawiamy rodzinę testów, analizujących własności stosu książek, zaproponowaną w publikacjach [RM03]⁴² i [RSS03]⁴³, choć jej początków można poszukiwać już w [R80]⁴⁴.

2.18.1. Test stosu książek

Położenie książki o wartości a w i -tym kroku opisuje $w_i(a)$, $i = 1..n$, gdzie 0 odpowiada wartości początkowej.

³⁹ K. Mańk, *Wyznaczenie dokładnych wartości parametrów dla testu Maurera*, Biuletyn WAT, LVII, 4, Warszawa, 2008, s. 117-149

⁴⁰ J.S. Coron, *On the Security of Random Sources*, Gemplus Corporate Product R&D Division, Technical Report IT02-1998

⁴¹ J.S. Coron, *On the Security of Random Sources*, Gemplus Corporate Product R&D Division, Technical Report IT02-1998, s. 9

⁴² B.Ya. Ryabko, V.A. Monarev, *Using Information Theory Approach to Randomness Testing*, Cryptology ePrint Archive, 2003, pozycja 2003/127

⁴³ B.Ya. Ryabko, V. S. Stognienko, Yu. I. Shokin, *A new test for randomness and its application to some cryptographic problems*, Journal of Statistical Planning and Inference, 2003

⁴⁴ B.Ya. Ryabko, *Information Compression by a Book Stack*, Problems of Information Transmission, v.16, N 4, 1980, s.16-21

Rozpoczynamy od stosu 2^d „książek”, uporządkowanych rosnąco zgodnie z ich wartością, odpowiadającą możliwym wartościom bloków bitów wydzielanych z badanego ciągu, przyjmujemy:

$$w_0(a) = a, \text{ dla } a = 0..2^d - 1.$$

W kolejnych krokach położenie książek modyfikowane jest według formuły:

$$w_i(a) = \begin{cases} 1, & \text{dla } b_i = a, \\ w_{i-1}(a) + 1, & \text{dla } w_{i-1}(a) < w_{i-1}(b_i), \\ w_{i-1}(a), & \text{dla } w_{i-1}(a) > w_{i-1}(b_i), \end{cases}$$

gdzie $i = 1..n$.

Zbiór położenia książek dzielimy na k przedziałów:

$$A_1 = \{0, \dots, K_1\}, A_2 = \{K_1 + 1, \dots, K_2\}, \dots, A_k = \{K_{k-1} + 1, \dots, 2^d - 1\}.$$

Niech $v_j = \#\{i = 1..n : w_{i-1}(b_i) \in A_j\}$ dla $j = 1..k$ będzie liczbą przypadków, gdy książka wyciągana jest na szczyt stosu z pozycji należącej do zbioru A_j .

Zaobserwowany rozkład wartości v_j porównywany jest następnie z rozkładem teoretycznym przy użyciu testu zgodności chi-kwadrat, przy założeniu, że rozkład wartości $w_{i-1}(b_i)$ powinien być rozkładem równomiernym.

2.18.2. Test porządku

Niech $\lambda_i(a) = \#\{j = 1..i : b_j = a\}$ będzie liczbą wystąpień wartości a pośród i początkowych bloków b_i .

Rozpoczynamy od stosu 2^d „książek”, uporządkowanych rosnąco zgodnie z ich wartością, odpowiadającą możliwym wartościom bloków bitów wydzielanych z badanego ciągu, przyjmujemy:

$$w_0(a) = a, \text{ dla } a = 0..2^d - 1.$$

W kolejnych krokach położenie książek modyfikowane jest w taki sposób, aby zachowana została zależność:

$$w_i(b_\alpha) < w_i(b_\beta) \Rightarrow \lambda_i(b_\alpha) \geq \lambda_i(b_\beta).$$

Dalsza część testu przebiega identycznie, jak w poprzednim przypadku.

Autorzy proponują aby d dobierać tak by

$$d \approx 2 \lg n.$$

W przypadku doboru podziału położenia zalecają przeprowadzenie własnych testów w celu osiągnięcia możliwie małej liczności próbek.

2.19. Test zgodności liczby komórek z dokładnie b punktami (Quadratic Box-Test)

Test jest rozwinięciem testu badającego liczbę komórek z dokładnie b punktami, opisanego w punkcie II.3.1.

Przyjmujemy, że badany jest ciąg binarny dzielony na n d -bitowych nienakładających się bloków. Każda z możliwych wartości bloku stanowi oddzielną komórkę (klasę), których liczba wynosi $k = d^l$.

Niech K_b będzie liczbą komórek z dokładnie b punktami (obserwacjami). Oznaczmy

$$\lambda = n/k, \quad p_b = \frac{\lambda^b}{b!} e^{-\lambda},$$

wówczas

$$E[K_b] \approx k \cdot p_b, \quad \text{Cov}[K_b, K_h] \approx k \cdot \sigma_{bh},$$

gdzie:

$$\sigma_{bb} = p_b \left(1 - p_b \left(1 + \frac{(\lambda - b)^2}{\lambda} \right) \right),$$

$$\sigma_{bh} = -p_b p_h \left(1 + \frac{(\lambda - b)(\lambda - h)}{\lambda} \right),^{45}$$

są elementami granicznej macierzy kowariancji B .⁴⁶

Rozpatrujemy wektor liczb komórek z poszczególnymi liczbami obserwacji:

$$K = [K_{b_1}, K_{b_2}, \dots, K_{b_l}],$$

odpowiadający tylko wybranemu podzbiorowi możliwych wartości liczby punktów w komórce.

Znormalizowany wektor K ma postać:

$$\underline{K} = \left[\frac{K_{b_1} - E[K_{b_1}]}{\sqrt{k}}, \frac{K_{b_2} - E[K_{b_2}]}{\sqrt{k}}, \dots, \frac{K_{b_l} - E[K_{b_l}]}{\sqrt{k}} \right].$$

Niech $B^{-1} = [\sigma^{i,j}]$ będzie macierzą odwrotną do granicznej macierzy kowariancji. Dla dużych k statystyka testowa:

$$Q = \underline{K} \times B^{-1} \times \underline{K} =$$

$$= \frac{1}{k} \sum_{i,j=1}^l \sigma^{i,j} (K_{b_i} - E[K_{b_i}]) (K_{b_j} - E[K_{b_j}]),$$

ma asymptotycznie rozkład chi-kwadrat z l stopniami swobody.⁴⁷

⁴⁵ I. Semaev, *Local limit theorem for large deviations and statistical box-tests*, Cryptology ePrint Archive, 2011, pozycja 2011/298, s. 19

⁴⁶ I. Semaev, M.M. Hassanzadeh, *New Statistical Box-Test and It's Power*, Cryptology ePrint Archive, 2011, pozycja 2011/297, s. 2

⁴⁷ I. Semaev, M.M. Hassanzadeh, *New Statistical Box-Test and It's Power*, Cryptology ePrint Archive, 2011, pozycja 2011/297, s. 5

Dla małych k , do obliczenia prawdopodobieństwa odpowiadającego otrzymanej wartości statystyki Q , można posłużyć się opisaną poniżej procedurą.⁴⁸

Niech $K = \sum_{i=1}^l K_{b_i}$ oznacza liczbę komórek łącznie dla wszystkich wybranych wartości liczby punktów w komórce, zaś $O = \sum_{i=1}^l b_i K_{b_i}$ łączną liczbę przypadających na nie obserwacji. Przyjmujemy:

$$n_1 = \left\lfloor n - \sum_{i=1}^l b_i \left(E[K_{b_i}] - \sqrt{Q\sigma_{ii}k} \right) \right\rfloor,$$

$$k_1 = \left\lfloor k - \sum_{i=1}^l \left(E[K_{b_i}] - \sqrt{Q\sigma_{ii}k} \right) \right\rfloor.$$

Rekurencyjnie obliczamy, dla wszystkich $n_2 \leq n_1$ i $k_2 \leq k_1$, prawdopodobieństwa:

$$P\{K(n_2, k_2) = 0\} = P\{K(n_2 - j, k_2 - 1) = 0\} \cdot P\{K(j, 1) = 0\} \cdot \sum_{i=0}^{n_2} \binom{n_2}{i} \frac{(k_2 - 1)^{n_2 - i}}{k_2^{n_2}},$$

przy czym:

$$P\{K(n, 1) = 0\} = \begin{cases} 0, & n \in \{b_1, \dots, b_l\}, \\ 1, & n \notin \{b_1, \dots, b_l\}. \end{cases}$$

Powyższa formuła upraszcza się dla wyboru wektora K postaci $[K_0, K_1, \dots, K_{l-1}]$ do:

$$P\{K(n_2, k_2) = 0\} = \sum_{i=l}^{n_2 - lk + l} \binom{n_2}{i} \frac{(k_2 - 1)^{n_2 - i}}{k_2^{n_2}} P\{K(n_2 - j, k_2 - 1) = 0\}.$$

Poszukiwanym prawdopodobieństwem jest suma prawdopodobieństw odpowiadających wszystkim możliwym wektorom $K = [K_{b_1}, K_{b_2}, \dots, K_{b_l}]$, którym odpowiada wartość statystyki testowej nie większa od zaobserwowanej.

Rozpoczynamy od wyznaczenia rozkładu macierzy B^{-1} na iloczyn macierzy trójkątnej górnej i jej transpozycji $B^{-1} = G \times G^T$, do czego można wykorzystać poniższy algorytm.

⁴⁸ I. Semaev, M.M. Hassanzadeh, *New Statistical Box-Test and It's Power*, Cryptology ePrint Archive, 2011, pozycja 2011/297, s. 13

```

 $G = B^{-1}$ 
for  $w = l$  to 1 do {
    for  $i$  from 1 to  $w - 1$  do {
        for  $j$  from 1 to  $w - 1$  do
             $g_{ij} = g_{ij} - \frac{g_{iw}g_{jw}}{g_{ww}}$ 
         $g_{wi} = 0$ 
    }
    for  $i$  from 1 to  $w$  do
         $g_{iw} = \frac{g_{iw}}{\sqrt{g_{ww}}}$ 
}
    
```

Ponieważ zachodzi:

$$X \times B^{-1} \times X = (X \times G) \times (X \times G)^T = (X \times G) \circ (X \times G),$$

to można sprowadzić obliczanie statystyki testowej do następującej sumy kwadratów:

$$kQ = \left((K_{b_1} - E[K_{b_1}])g_{11} \right)^2 + \left((K_{b_1} - E[K_{b_1}])g_{12} + (K_{b_2} - E[K_{b_2}])g_{22} \right)^2 + \dots + \left((K_{b_1} - E[K_{b_1}])g_{1l} + \dots + (K_{b_l} - E[K_{b_l}])g_{ll} \right)^2.$$

Na jej podstawie, iteracyjnie, wykorzystując kolejne składniki, możemy wyznaczać zakresy wartości dla kolejnych K_{b_i} . Dla K_{b_1} , na podstawie pierwszego składnika, mamy:

$$\max \left\{ E[K_{b_1}] - \frac{\sqrt{kQ}}{g_{11}}, 0 \right\} \leq K_{b_1} \leq E[K_{b_1}] + \frac{\sqrt{kQ}}{g_{11}}.$$

Na podstawie drugiego składnika dostajemy:

$$\left| (K_{b_1} - E[K_{b_1}])g_{12} + (K_{b_2} - E[K_{b_2}])g_{22} \right| \leq \sqrt{kQ - \left((K_{b_1} - E[K_{b_1}])g_{11} \right)^2},$$

z czego możemy wyznaczyć zakres wartości K_{b_2} i tak dalej.

Dodatkowo możemy ograniczyć zakres poszukiwań, biorąc pod uwagę oczywiste relacje:

$$\sum_{i=1}^l K_{b_i} \leq k \quad \text{oraz} \quad \sum_{i=1}^l b_i K_{b_i} \leq n.$$

Jeśli wybrano wektor K postaci $[K_0, K_1, \dots, K_{l-1}]$, to dostajemy dodatkowy warunek:

$$\sum_{i=1}^l b_i K_{b_i} \leq n - l \left(k - \sum_{i=1}^l K_{b_i} \right).$$

2.20. Test pochodnej ciągu liczb

Test, a w zasadzie pomysł na nową testowaną własność, został zaproponowany w [BA14]⁴⁹.

Pochodną ciągu liczb naturalnych $B^{d,n}$, nazywać będziemy ciąg $n-1$ liczb naturalnych danych następująco:

$$|b_2 - b_1|, |b_3 - b_2|, \dots, |b_n - b_{n-1}|.$$

k -tą pochodną ciągu liczb $B^{d,n}$ nazywać będziemy pochodną $k-1$ pochodnej tego ciągu.

Średnią wartością k -tego odstępu (IAI - *Interval Average Indices*) nazywać będziemy średnią uzyskaną dla ciągu stanowiącego k -tą pochodną i oznaczać $IAI_k(B^{d,n})$.

W szczególności mamy:

$$IAI_1(B^{d,n}) = \frac{1}{n-1} \sum_{i=1}^{n-1} |b_{i+1} - b_i|.$$

Autorzy przeprowadzili serię eksperymentów, na podstawie których poczynili następujące obserwacje:

$$E[IAI_1(B^{d,n})] = \frac{2^d}{3} \pm \varepsilon,$$

gdzie

- $\varepsilon = -\frac{2^d}{4}$ dla $d = 1$,
- $\varepsilon \leq \left\lfloor \frac{2^d}{10} \right\rfloor$ dla $d > 1$,
- $\lim_{d \rightarrow \infty} \varepsilon = 0$;

$$E[IAI_k(B^{d,n})] = \frac{2^d}{K_k} \pm \varepsilon,$$

gdzie K_k są stałymi podanymi w tabeli II.1 i $\lim_{d \rightarrow \infty} \varepsilon = 0$.⁵⁰

Tabela II.1. Stałe dla wartości oczekiwanych k -tego odstępu

k	1	2	3	4	5	6	7	8	9	10
K_k	3	4	5,08 (5,094)	6,44 (6,449)	7,98 (8,025)	9,8 (8,89)	11,73 (12,62)	13,96 (13,55)	16,45	19,25

Autorzy proponują aby samo testowanie odbywało się na zasadzie porównania otrzymywanych wyników z jakąś grupą kontrolną (np. zweryfikowanym generatorem losowym).

⁴⁹ S. Bouftass, A. Azhari, *On a new properties of random number sequences, a randomness test and a new RC4's key scheduling algorithm*, Cryptology ePrint Archive, 2014, pozycja 2014/466

⁵⁰ S. Bouftass, A. Azhari, *On a new properties of random number sequences, a randomness test and a new RC4's key scheduling algorithm*, s. 2-3

W ramach realizacji pracy p.k. ROTOR⁵¹ udało się nam wyznaczyć wiele parametrów tego testu:

- wartość oczekiwaną i wariancję $IAI_1(B^{d,n})$:

$$E[IAI_1(B^{d,n})] = \frac{1}{3} \left(2^d - \frac{1}{2^d} \right),$$

$$D^2[IAI_1(B^{d,n})] = \frac{1}{90(n-1)^2} \left((6n-7)2^{2d} + 5 - (3n-1)\frac{2}{2^{2d}} \right),$$

- wartość oczekiwaną i wariancję $IAI_2(B^{d,n})$:

$$E[IAI_2(B^{d,n})] = \frac{2^d}{4},$$

$$D^2[IAI_2(B^{d,n})] = \frac{1}{2160(n-2)^2} \left((109n-252) \cdot 2^{2d} + 30(15n-32) - (107n-260)\frac{24}{2^{2d}} + (n-3)\frac{1280}{2^{4d}} \right),$$

- wartości oczekiwane dla $IAI_3(B^{d,n})$, $IAI_4(B^{d,n})$ i $IAI_5(B^{d,n})$,
- wartości stałych K_k dla $k \leq 8$, które w tabeli II.1 podane zostały w nawiasach.

2.21. Test kolekcjonera

Dla ciągu $[k \cdot u_i]$ badamy długość rozłącznych odcinków potrzebnych do uzyskania pełnego zestawu wartości, w tym przypadku liczb całkowitych z zakresu $0..k-1$. Prawdopodobieństwo, że kolekcja będzie miała długość r , obliczyć można ze wzorów:

$$p_r = \frac{k!}{k^r} \left\{ \begin{matrix} r-1 \\ k-1 \end{matrix} \right\} = \frac{1}{k^{r-1}} \sum_{i=0}^{k-2} (-1)^i \binom{k-1}{i} (k-1-i)^{r-1}, \text{ dla } r \geq k,^{52}$$

zaś, że kolekcja będzie miała długość co najmniej t :

$$p_t = 1 - \frac{k!}{k^{t-1}} \left\{ \begin{matrix} t-1 \\ k \end{matrix} \right\}, \text{ dla } t \geq d.$$

Do budowy statystyki chi-kwadrat należy wykorzystać zaobserwowaną dla danego ciągu liczbę kolekcji, nie zaś jej wartość oczekiwaną.

W naszym przypadku sensowne wydaje się ograniczenie do k będących potęgami 2 – badanie ciągów $B^{d,n}$.

2.22. Test serii rosnących / malejących

2.22.1. Test serii z macierzą kowariancji

Ciąg U^n badany jest w tym teście pod kątem rozkładu długości serii rosnących bądź malejących, przy czym badanie przeprowadza się dla nich niezależnie. Przez serię rozumiemy każdą sekwencję elementów ciągu zachowującą monotoniczność. Niech v_i dla $i = 1..t-1$ będą za-

⁵¹ K. Mańk, *Wyznaczenie rozkładu średniej wartości odstępu w teście pochodnej ciągu liczb*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2017

⁵² R. Wieczorkowski, R. Zieliński, *Komputerowe generatory liczb losowych*, WNT, Warszawa, 1997, s. 123

obserwowanymi liczbami serii rosnących (malejących) długości i , zaś v_i – zaobserwowaną liczbą serii długości co najmniej t . Statystyka testowa ma postać

$$S_{obs} = \sum_{1 \leq i, j \leq t} (v_i - E_i)(v_j - E_j) c_{ij}^{-1},$$

gdzie E_i są wartościami oczekiwanymi odpowiadającymi v_i , zaś c_{ij}^{-1} – elementami macierzy odwrotnej do macierzy kowariancji C dla wektora v_i , $i = 1..t$. Statystyka testowa ma asymptotycznie rozkład chi-kwadrat z t stopniami swobody.

Sposób wyznaczania wartości oczekiwanych i macierzy kowariancji przedstawiony jest w [K97]⁵³. Niech E_i^* będzie wartością oczekiwaną liczby serii długości co najmniej i , zaś macierz kowariancji C ma postać:

$$C = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,t}^* \\ c_{1,2} & c_{2,2} & \cdots & c_{2,t}^* \\ \vdots & \vdots & \ddots & \vdots \\ c_{1,t}^* & c_{2,t}^* & \cdots & c_{t,t}^{**} \end{bmatrix},$$

gdyż korzystamy z symetryczności macierzy kowariancji. Mamy:

$$E_i^* = (n+1) \frac{i}{(i+1)!} - \frac{i-1}{i!},$$

$$E_i = E_i^* - E_{i+1}^*,$$

oraz

$$c_{i,j}^{**} = E_{\max\{i,j\}}^* + f(i, j),$$

$$c_{i,j}^* = c_{i,j}^{**} - c_{i+1,j}^{**},$$

$$c_{i,j} = c_{i,j}^* - c_{i,j+1}^*,$$

gdzie

$$f(i, j) = (n+1) \left(\frac{(i+j)(1-ij) + ij}{(i+1)!(j+1)!} - \frac{2(i+j)}{(i+j+1)!} \right) + \\ + \frac{2(i+j-1)}{(i+j)!} + \frac{\left((i+j)^2 - i - j - 2 \right) ij - (i+j)^2 - i^2 j^2 + 1}{(i+1)!(j+1)!}.$$

Otrzymywana macierz C ma postać

$$C = n \cdot C_1 + C_2,$$

zaś macierz odwrotna wyraża się formułą

$$C^{-1} = \frac{1}{n} C_1^{-1} - \frac{1}{n^2} C_1^{-1} C_2 C_1^{-1} + \frac{1}{n^3} C_1^{-1} C_2 C_1^{-1} C_2 C_1^{-1} - \dots,$$

przy czym dla dużych n istotne są jedynie dwa lub trzy pierwsze wyrazy.

⁵³ D.E. Knuth, *Sztuka programowania, t.2. Algorytmy seminumeryczne*, WNT, 2002, s. 70-73

Efektywna implementacja tego testu zawierać powinna zawczasu obliczone macierze C_1^{-1} , $C_1^{-1}C_2C_1^{-1}$ oraz $C_1^{-1}C_2C_1^{-1}C_2C_1^{-1}$, odpowiadające dopuszczalnym wartościom liczby klas t .

2.22.2. Test serii z pomijaniem

Test serii rosnących / malejących posiada wariant, pozwalający uniknąć stosowania macierzy kowariancji. Sposobem na to jest pominięcie elementu następującego po serii, dzięki temu długości kolejnych serii przestają być skorelowane i do zliczeń można zastosować standardowy test zgodności chi-kwadrat.

Wartości oczekiwane liczby serii poszczególnych długości obliczyć można na podstawie poniższych formuł:

$$E_i^* = \frac{R}{i!},$$

$$E_i = E_i^* - E_{i+1}^*,^{54}$$

gdzie R jest liczbą zaobserwowanych serii.

2.23. Test kolizji

Procedura testowa polega na wyznaczeniu liczby kolizji, jakie pojawią się podczas próby podziału zbioru n -elementowego na m kategorii, przez kolizję rozumiemy próbę umieszczenia kolejnego elementu w kategorii, do której już jakiś element należy.

Prawdopodobieństwo wystąpienia r kolizji wyraża się następująco:

$$p_r = \frac{m!}{(m-n+r)!m^n} \left\{ \begin{matrix} n \\ n-r \end{matrix} \right\}.^{55}$$

Na podstawie powyższej formuły konstruuje się testy całkowitej liczby kolizji oraz zgodności liczby kolizji przy podziale ciągu na fragmenty ustalonej długości.

2.24. Test ściskania

Procedura testowa polega na ustaleniu liczby kroków N , które należy wykonać, aby wykonując $k_i = \lceil u_i \cdot k_{i-1} \rceil$ otrzymać $k_N = 1$, przy czym N jest najmniejszym indeksem, dla którego to zachodzi, a k_0 jest ustaloną stałą. Następnie zaobserwowany rozkład wartości N porównywany jest przy użyciu testu chi-kwadrat z rozkładem teoretycznym.

Niestety nie są znane metody analityczne i dla ustalonych d i k_0 rozkład wyznacza się symulacyjnie. W ramach niniejszej pracy wyznaczono empiryczne tabele dla bloków $d = 10..32$ i 64-bitowych oraz $k_0 = 2^d$.

Test ściskania został wprowadzony w pakiecie DIEHARD.⁵⁶

⁵⁴ D.E. Knuth, *Sztuka programowania*, t.2. *Algorytmy seminumeryczne*, WNT, 2002, s. 608

⁵⁵ D.E. Knuth, *Sztuka programowania*, t.2. *Algorytmy seminumeryczne*, WNT, 2002, s. 74

⁵⁶ G. Marsaglia, *DIEHARD Battery of Tests of Randomness*, 1995, <http://stat.fsu.edu/pub/diehard/>

2.25. Test oparty na grze w kości

W trakcie testu wykonywane jest N gier w kości, wg zasad Craps, tzn.:

- rzuca się dwiema kośćmi, a o wyniku decyduje suma oczek,
- jeśli w pierwszej kolejce zostanie wyrzucone:
 - 7 lub 11 – gracz wygrywa i gra się kończy,
 - 2, 3 lub 12 – gracz przegrywa i gra się kończy,
 - 4, 5, 6, 8, 9 lub 10 – gra toczy się dalej,
- jeśli w kolejnych kolejkach zostanie wyrzucone:
 - 7 – gracz przegrywa i gra się kończy,
 - ta sama liczba oczek, co w pierwszej kolejce – gracz wygrywa i gra się kończy,
 - inna liczba oczek – gra toczy się dalej.

Zliczana jest liczba gier zakończonych zwycięstwem, dla której obliczane jest prawdopodobieństwo zgodnie z rozkładem normalnym o wartości oczekiwanej $N \cdot p$ i wariancji $N \cdot p \cdot (1 - p)$, gdzie p jest prawdopodobieństwem wygranej w pojedynczej grze.

Równocześnie zliczana jest także liczb kolejek konieczna do rozstrzygnięcia pojedynczej gry. Niech v_i dla $i = 1..t-1$ oznacza zaobserwowaną liczbę gier, w których do rozstrzygnięcia doszło w i -tej kolejce, zaś v_t – liczbę gier, w których do rozstrzygnięcia doszło w co najmniej t -tej kolejce. Dla ciągu obserwacji v_i buduje się test zgodności chi-kwadrat, w którym wartości oczekiwane wyraża się przy pomocy wyznaczonego zestawu prawdopodobieństw.

Test został zaproponowany w pakiecie DIEHARD w wariancie dla kostek sześciościennych, w ramach realizacji projektu ROTOR Tomasz Czechowski zaproponował wariant dla kostek ośmiościennych, który usuwa konieczność odrzucania trzybitowych bloków o wartościach 0 i 7 w oryginalnej wersji testu.⁵⁷

2.26. Test największego wspólnego dzielnika

Na podstawie badanego ciągu liczb U^n , dla $n = 10^7$, wyznaczane są dwa ciągi:

- największych wspólnych dzielników kolejnych nienakładających się par liczb – g ,
- liczb kroków algorytmu Euklidesa koniecznych od ich wyznaczenia – k .

Dla ciągu obserwacji największych wspólnych dzielników stosowany jest test zgodności chi-kwadrat z rozkładem teoretycznym danym formułą:

$$\Pr(\gcd(a, b) = g) = \pi^2 / 6g^2. \quad ^{58}$$

Dla liczby kroków musiałyby być stosowane tablice wyznaczone empirycznie.

⁵⁷ T. Czechowski, *Wykonanie implementacji i analiza możliwości uogólnienia testu opartego na grze Craps*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2018

⁵⁸ G. Marsaglia, W.W. Tsang, *Some difficult-to-pass tests of randomness*, Journal of Statistical Software 7, 3, 2002, s. 3

2.27. Test wypełniania drzewa binarnego na podstawie ciągu liczb

Dla badanego ciągu liczb $U^n = u_1, u_2, \dots, u_n$, powtarzana jest procedura umieszczania ich w drzewie binarnym o wysokości h . W każdej iteracji drzewo jest opróżniane, a następnie kolejne elementy umieszczane są w nim, aż do momentu, gdy wpis nie jest możliwy z powodu zajęcia liścia przez wcześniejszy element. Wyznaczane są dwie statystyki testowe:

- liczba elementów, jaką udało się umieścić w drzewie – zaobserwowany rozkład porównywany jest przy użyciu testu Pearsona z rozkładem wyznaczonym empirycznie,
- numer liścia, dla którego wystąpiła kolizja – zaobserwowany rozkład porównywany jest przy użyciu testu Pearsona z rozkładem równomiernym.

Jeżeli zamiast ciągu liczb użyty zostanie ciąg bloków $B^{d,n}$ dla małych d , wówczas rozkład numerów liści, dla których wystąpiła kolizja przestaje być rozkładem równomiernym.

Na potrzeby pracy, najpierw wyznaczone zostały empiryczne przybliżenia funkcji prawdopodobieństwa dla obu rozkładów dla $d = 4..32$ oraz ∞ i wysokości drzewa $h = 3..11$. W późniejszym czasie udało nam się pokonać początkowe trudności i wyprowadzić formuły analityczne. Obszerniejsze omówienie wyprowadzenia przedstawione zostało w artykule [Ma25]⁵⁹, poniżej przedstawiamy tylko istotne formuły.

Niech $P_{h,l,k}(A, B)$ będzie prawdopodobieństwem wystąpienia kolizji na l -tym liściu, gdy w drzewie wysokości h umieszczono już k elementów będących liczbami naturalnymi z zakresu $\{A, \dots, B-1\}$. Mamy rekurencyjne formuły:

$$P_{h,l,k}(A, B) = \sum_{a=A}^{B-1} \frac{1}{B-A} \sum_{m=\max\{h-1, k-2^{h-1}\}}^{\min\{k-1, 2^{h-1}-1\}} P_{h-1,l,m}(A, a) \cdot S_{h-1,k-m-1}(a, B) \cdot \frac{(a-A)^{m+1} (B-a)^{k-m-1}}{(B-A)^k} \cdot \binom{k-1}{m},$$

dla $l \in \{0, \dots, 2^{h-2} - 1\}$ oraz

$$P_{h,l+2^{h-2},k}(A, B) = \sum_{a=A}^{B-1} \frac{1}{B-A} \sum_{m=\max\{h-1, k-2^{h-1}\}}^{\min\{k-1, 2^{h-1}-1\}} P_{h-1,l,m}(a, B) \cdot S_{h-1,k-m-1}(A, a) \cdot \frac{(a-A)^{k-m-1} (B-a)^{m+1}}{(B-A)^k} \cdot \binom{k-1}{m},$$

dla $l + 2^{h-2} \in \{2^{h-2}, \dots, 2^{h-1} - 1\}$.

Na ich podstawie, dla danego h możemy obliczyć interesujące nas prawdopodobieństwa:

- $P_{h,l,*}(A, B) = \sum_{k=h}^{2^h-1} P_{h,l,k}(A, B)$ – wystąpienia kolizji na l -tym liściu,
- $P_{h,*,k}(A, B) = \sum_{l=0}^{2^{h-1}-1} P_{h,l,k}(A, B)$ – wystąpienia kolizji po umieszczeniu k elementów w drzewie.

⁵⁹ K. Mańk, *Upgrading Frequency Test for Overlapping Vectors and Fill Tree Tests*, IJET, 2/2025

W kolejnym kroku rekurencji potrzebne będzie prawdopodobieństwo umieszczenia k elementów w drzewie:

$$S_{h,k}(A, B) = \sum_{m=k}^{2^h-1} P_{h,*,m}(A, B).$$

Powyższe prawdopodobieństwa dla drzewa wysokości 2 można łatwo wyznaczyć kombinatorycznie. Mamy:

$$P_{2,0,2}(A, B) = \sum_{a=A}^{B-1} \sum_{b=A}^{a-1} \sum_{c=A}^{a-1} \frac{1}{(B-A)^3} = \frac{1 + (2(B-A) - 3)(B-A)}{6(B-A)^2},$$

$$P_{2,1,2}(A, B) = \sum_{a=A}^{B-1} \sum_{b=A}^{B-1} \sum_{c=a}^{B-1} \frac{1}{(B-A)^3} = \frac{1 + (2(B-A) + 3)(B-A)}{6(B-A)^2},$$

$$P_{2,0,3}(A, B) = \sum_{a=A}^{B-1} \sum_{b=A}^{a-1} \sum_{c=a}^{B-1} \sum_{e=A}^{a-1} \frac{2}{(B-A)^4} = \frac{(B-A)^2 - 1}{6(B-A)^2},$$

$$P_{2,1,3}(A, B) = \sum_{a=A}^{B-1} \sum_{b=A}^{a-1} \sum_{c=a}^{B-1} \sum_{e=a}^{B-1} \frac{2}{(B-A)^4} = \frac{(B-A)^2 - 1}{6(B-A)^2},$$

z czego:

$$S_{2,0}(A, B) = S_{2,1}(A, B) = S_{2,2}(A, B) = 1.$$

$$S_{2,3}(A, B) = P_{2,0,3}(A, B) + P_{2,1,3}(A, B) = \frac{(B-A)^2 - 1}{3(B-A)^2}$$

Po dojściu do założonej wysokości drzewa należy do uzyskanych formuł podstawić $A = 0$ oraz $B = 2^d$.

2.28. Test autokorelacyjny dla ciągu bloków bitów

Testy opisane w punktach II.2.4 i II.5.1 są, z punktu widzenia obliczeń przy użyciu komputera cyfrowego, dwoma skrajnymi przypadkami. Pierwszy zakłada nieskończoną precyzję arytmetyki, choć w praktyce jest to 32 lub 64 bity, w drugim ciąg dzielony jest na bloki jednobitowe.

Wyjściowy ciąg binarny $s^{d(n+t)}$ dzielony jest na rozłączne bloki długości d bitów, które następnie utożsamiamy z liczbami naturalnymi ze zbioru $\{0, 1, \dots, 2^d - 1\}$. W ten sposób otrzymujemy ciąg liczb całkowitych $C^{n+t} = c_1, c_2, \dots, c_{n+t}$. Na podstawie założenia o równomierności i niezależności rozkładu bitów w ciągu wyjściowym, można również założyć, że c_i tworzą ciąg niezależnych zmiennych losowych o rozkładzie równomiernym na zbiorze $\{0, 1, \dots, 2^d - 1\}$.

Wyznaczana jest statystyka testowa:

$$S_{ncykl}^d = \sum_{i=1}^n c_i c_{i+t}, \quad t > 0.$$

Zakładamy, że dla dużych n ma ona, w przybliżeniu, rozkład dany formułą:⁶⁰

$$\Pr\left(\frac{S_{ncykl}^d - m_1}{\sigma} < x\right) = \Phi(x) - \frac{1}{3!} \frac{\mu_3}{\sigma^3} \varphi^{(2)}(x) + \frac{1}{4!} \left(\frac{\mu_4}{\sigma^4} - 3\right) \varphi^{(3)}(x) - \frac{1}{5!} \left(\frac{\mu_5}{\sigma^5} - 10 \frac{\mu_3}{\sigma^3}\right) \varphi^{(4)}(x),$$

gdzie:

$\Phi(x)$ - dystrybuanta standardowego rozkładu normalnego,

$\varphi^{(k)}(x)$ - k -ta pochodna gęstości standardowego rozkładu normalnego,

σ - odchylenie standardowe zmiennej S_{ncykl}^d ,

μ_k - moment centralny rzędu k zmiennej S_{ncykl}^d .

Poszczególne momenty podajemy za [Ma13]:⁶¹

$$m_1 = \frac{(2^d - 1)^2}{4} n,$$

$$\mu_2 = \sigma^2 = \frac{(2^d - 1)^2 (2^d + 1)}{144} ((13 \cdot 2^d - 11)n - 6(2^d - 1)t),$$

$$\mu_3 = \frac{1}{24} (2^d - 1)^4 (2^d + 1)^2 (n - t),$$

$$\mu_4 = \frac{(2^d - 1)^2 (2^d + 1)}{2^8 3^3 5^2} \left[\begin{aligned} &25(2^d - 1)^2 (2^d + 1)(13 \cdot 2^d - 11)^2 n^2 + \\ &-6(50(2^d - 1)^3 (2^d + 1)(13 \cdot 2^d - 11)t - A_2)n + \\ &+20(2^d - 1)(45(2^d - 1)^3 (2^d + 1)t - A_3)t \end{aligned} \right],$$

$$A_2 = 327 \cdot 2^{5d} + 167 \cdot 2^{4d} - 2606 \cdot 2^{3d} + 2514 \cdot 2^{2d} + 367 \cdot 2^d - 753,$$

$$A_3 = 215 \cdot 2^{4d} + 130 \cdot 2^{3d} - 792 \cdot 2^{2d} + 110 \cdot 2^d + 313,$$

$$\mu_5 = \frac{(2^d - 1)^4 (2^d + 1)^2}{2^6 3^3 5} \left[\begin{aligned} &25(2^d - 1)^2 (2^d + 1)(13 \cdot 2^d - 11)^2 n^2 + \\ &- (25(2^d - 1)^2 (2^d + 1)(19 \cdot 2^d - 17)t - A_9)n + \\ &+ 2(75(2^d - 1)^3 (2^d + 1)t + A_8)t \end{aligned} \right],$$

$$A_8 = 59 \cdot 2^{4d} - 330 \cdot 2^{3d} + 668 \cdot 2^{2d} - 330 \cdot 2^d - 91,$$

$$A_9 = 148 \cdot 2^{4d} - 480 \cdot 2^{3d} + 856 \cdot 2^{2d} - 480 \cdot 2^d - 92.$$

⁶⁰ K. Mańk, *Test autokorelacyjny dla ciągu binarnego*, Studia Bezpieczeństwa Narodowego, IV, 6, Kryptologia i Cyberbezpieczeństwo, Warszawa, 2014, s. 213

⁶¹ K. Mańk, *Test autokorelacyjny dla ciągu binarnego*, Studia Bezpieczeństwa Narodowego, IV, 6, Kryptologia i Cyberbezpieczeństwo, Warszawa, 2014, s. 205-207

3. Testy dla ciągów nienakładających się wektorów

W tym punkcie omawiamy dużą grupę testów, które badany ciąg binarny przekształcają najpierw na ciąg bloków, z których formowane są następnie wektory nie posiadające wspólnych elementów.

3.1. Testy dla zliczeń liczby punktów w komórkach

Badamy równomierność rozkładu i niezależność kolejnych liczb w ciągu $U^m = u_1, u_2, \dots, u_m$, gdzie $u_i \in [0;1)$ poprzez sprawdzenie równomierności rozkładu punktów w hiperprzestrzennej kostce jednostkowej. W tym celu badany ciąg dzielimy na n rozłącznych t -elementowych wektorów, które traktujemy jako współrzędne w t -wymiarowej przestrzeni. Jednocześnie przedział $[0;1)$ dzielimy na k podprzedziałów równej długości, co daje podział hiperkostki na $K = k^t$ identycznych komórek – hiperkostek.

Przy założeniu równomierności i niezależności rozkładu liczb w badanym ciągu rozkład punktów w hiperkostce jednostkowej jest również równomierny, co daje identyczną wartość oczekiwaną liczby punktów w każdej komórce:

$$\lambda = n/K.$$

W pracy [ESW02]⁶² rozważana jest statystyka testowa ogólnej postaci

$$S_{obs} = \sum_{i=0}^{K-1} f_{n,K}(v_i),$$

gdzie

v_i , $i = 0..K-1$ jest liczbą punktów, które znalazły się w i -tej kostce,

$f_{n,K}(x)$ jest funkcją rzeczywistą, parametryzowaną przez n i K , rozważane postaci tej funkcji przedstawione są w tabeli II.2.

Warto zwrócić uwagę, że test zgodności Pearsona jest szczególnym przypadkiem rodziny testów potęgowo-dywergencyjnych, dla $\delta = 1$, zaś statystyka G^2 odpowiada przypadkowi $\delta \rightarrow 0$. Z kolei, tak zdefiniowany test entropii nie jest odpowiednikiem uniwersalnego testu Maurera. Inne związki pomiędzy tymi statystykami to:

- $H = \lg(K) - G^2 / (2n \ln 2),$
- $L_0 = K - W_1 = K - n + C,$
- $W_b = \sum_{i=b}^n L_i,$

⁶² P. L'Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, SIAM Journal on Scientific Computing, 24, 2, 2002, s. 652-668

$$\bullet \quad C = \sum_{i=2}^n W_i. {}^{63}$$

Tabela II.2. Przykłady postaci funkcji $f_{n,K}(x)$ ⁶⁴

S_{obs}	$f_{n,K}(x)$	nazwa testu
D_δ	$2x((x/\lambda)^\delta - 1)/\delta(1+\delta)$	potęgowo-dywergencyjny (Power Divergence)
χ^2	$(x - \lambda)^2/\lambda$	zgodności Pearsona (Pearson's)
G^2	$2x \ln(x/\lambda)$	logarytmu prawdopodobieństwa (Loglikelihood)
$-H$	$(x/n) \lg(x/n)$	entropii (Negative Entropy)
L_b	$I[x = b]$	liczby komórek z dokładnie b punktami
W_b	$I[x \geq b]$	liczby komórek z co najmniej b punktami
L_0	$I[x = 0]$	liczby pustych komórek
C	$(x-1)I[x > 1]$ albo $\max\{0, x-1\}$	kolizji

W celu wyznaczenia rozkładów i ich parametrów dla określonych powyżej statystyk testowych wyróżniono trzy przypadki:

- gęsty, gdy $n \rightarrow \infty$ i $\lambda \rightarrow \infty$ przy ustalonym K ,
- rzadki, gdy $n \rightarrow \infty$ i $K \rightarrow \infty$, ale λ jest stałe (lub ma skończoną granicę),
- bardzo rzadki, gdy $\lambda \rightarrow 0$.

Dokładne wzory określające wartość oczekiwaną i wariancję statystyki mają postać:

$$E[S_{obs}] = K\mu = \sum_{i=0}^n \binom{n}{i} \frac{(K-1)^{n-i}}{K^{n-1}} f_{n,K}(i),$$

$$D^2[S_{obs}] = \sum_{i=0}^n \binom{n}{i} \frac{(K-1)^{n-i}}{K^{n-1}} (f_{n,K}(i) - \mu)^2 + \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} \binom{n-i}{i} \frac{(K-1)(K-2)^{n-2i}}{K^{n-1}} (f_{n,K}(i) - \mu)^2 +$$

$$+ 2 \sum_{i=0}^n \sum_{j=0}^{\min\{n-i, i-1\}} \binom{n}{i} \binom{n-i}{j} \frac{(K-1)(K-2)^{n-i-j}}{K^{n-1}} (f_{n,K}(i) - \mu)(f_{n,K}(j) - \mu),$$

gdzie

$$\mu = E[f_{n,K}(v_i)] = \sum_{i=0}^n \binom{n}{i} \frac{(K-1)^{n-i}}{K^n} f_{n,K}(i).$$

⁶³ P. L'Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, SIAM Journal on Scientific Computing, 24, 2, 2002, s. 655

⁶⁴ tamże, s. 656

Niestety, powyższe sumy dają się łatwo wyznaczać tylko dla małych λ , a więc dla przypadku „bardzo rzadkiego”. Wówczas kolejne wyrazy zanikają wykładniczo wraz ze wzrostem $i + j$. Dla dużych wartości zarówno n i λ (przypadek „gęsty”), można posłużyć się następującymi przybliżeniami:

$$E[D_\delta] = (K-1) + \frac{(\delta-1)(K-1)}{2n} \left(\frac{\delta(K-1)}{2} - \frac{K+1}{3} \right) + o\left(\frac{1}{n}\right),$$

$$D^2[D_\delta] = 2(K-1) \left(1 + \frac{1}{n} \left((\delta-1) \left(\frac{(\delta+5)(K-2)}{3} + (\delta-2)(K-1) \right) - 1 \right) \right) + o\left(\frac{1}{n}\right).^{65}$$

Dla dużych n i K dobre przybliżenie uzyskuje się również wykorzystując model, w którym zamiast stałej liczby n wektorów przyjmuje się zmienną losową o rozkładzie Poissona, której wartość oczekiwana wynosi n . Wartość oczekiwana i wariancja wyrażają się wówczas formułami:

$$E[D_\delta] = K\mu = K \sum_{i=0}^n \frac{\lambda^i e^{-\lambda}}{i!} f_{n,K}(i),$$

$$D^2[S_{obs}] = K \sum_{i=0}^n \frac{\lambda^i e^{-\lambda}}{i!} (f_{n,K}(i) - \mu)^2. ^{66}$$

Rozkład zmiennych losowych opisujących poszczególne statystyki jest bardziej zróżnicowany i zależy nie tylko od typu statystyki, ale też od wartości λ .

Dla przypadku „gęstego” statystyka testowa:

$$D_\delta^{(C)} = \frac{D_\delta - E[D_\delta] + (K-1)\sigma_C}{\sigma_C}, ^{67}$$

gdzie $\sigma_C^2 = \frac{D^2[D_\delta]}{2(K-1)}$, ma asymptotycznie rozkład chi-kwadrat z $K-1$ stopniami swobody. Jeśli użyjemy modelu wykorzystującego rozkład Poissona, to liczba stopni swobody wynosić powinna K .

Dla przypadku „rzadkiego”, w obu modelach, statystyka testowa

$$D_\delta^{(N)} = \frac{D_\delta - E[D_\delta]}{\sqrt{D^2[D_\delta]}} ^{67}$$

ma asymptotycznie rozkład $N(0;1)$.

⁶⁵ T.R.C. Read, N.A.C. Cressie, *Goodness-of-Fit Statistics for Discrete Multivariate Data*, Springer Series in Statistics, Springer-Verlag, Nowy Jork, 1988, s. 65

⁶⁶ P. L’Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, SIAM Journal on Scientific Computing, 24, 2, 2002, s. 656

⁶⁷ P. L’Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, SIAM Journal on Scientific Computing, 24, 2, 2002, s. 657

W przypadku „rzadkim” dla pozostałych statystyk można posłużyć się poniższymi twierdzeniami:⁶⁷

- dla $b \geq 2$ niech $\frac{n^b}{K^{b-1}b!} \rightarrow \lambda_b$, wówczas statystyki L_b i W_b mają asymptotycznie rozkład Poissona z parametrem λ_b ,

- C zbiega do L_2 ,
- jeśli $\frac{n}{K} - \ln(K) \rightarrow \gamma_0$, to L_0 ma asymptotycznie rozkład Poissona z parametrem $e^{-\gamma_0}$,
- jeśli $K \rightarrow \infty$ i $\frac{n}{K} \rightarrow \lambda_0 > 0$, wówczas $S_{obs} = L_b$, W_b lub C , po ustandaryzowaniu

$$\frac{S_{obs} - E[S_{obs}]}{\sqrt{D^2[S_{obs}]}}$$

ma asymptotycznie rozkład $N(0;1)$,

- dokładny rozkład liczby kolizji opisuje formuła:

$$\Pr(C = c) = \Pr(L_0 = K - n + c) = \frac{K!}{(K - n + c)!K^n} \left\{ \frac{n}{n - c} \right\},^{68}$$

- wartość oczekiwana liczby kolizji wynosi

$$E[C] = K \left(\frac{n}{K} - 1 + \left(1 - \frac{1}{K} \right)^n \right) \approx \frac{n^2}{2K}.$$

3.2. Test m najbliższych par

Jak w poprzednim punkcie badany ciąg U^m dzielimy na n rozłącznych t -elementowych wektorów $W_i = [w_1, w_2, \dots, w_t] = [u_{t(i-1)+1}, u_{t(i-1)+2}, \dots, u_{ti}]$ dla $i = 1..n$, które utożsamiamy ze współrzędnymi punktów w t -wymiarowej przestrzeni.

Dla pary punktów W_i oraz W_j definiujemy odległość między nimi:⁶⁹

$$D_{i,j} = \|W_i - W_j\|_p,$$

gdzie

$$\|W\|_p = \begin{cases} \sqrt[p]{\sum_{j=1}^t \left(\min\{|w_j|, 1 - |w_j|\} \right)^p}, & 1 \leq p < \infty, \\ \max_{j=1..t} \left\{ \min\{|w_j|, 1 - |w_j|\} \right\} & p = \infty. \end{cases}$$

⁶⁸ patrz punkt II.2.23.

⁶⁹ P. L'Ecuyer, R. Simard, *TestU01: A C Library for Empirical Testing of Random Number Generators*, 2002, s. 14

Oznaczmy $\lambda(n) = n(n-1)V_t(1)/2$, gdzie $V_t(r)$ jest objętością t -wymiarowej kuli o promieniu r według miary $\|W\|_p$. Wzory na $V_t(r)$ są stosunkowo proste w dwóch przypadkach:

- dla $p = 2$ mamy $V_t(r) = \begin{cases} \pi^h r^t / h!, & \text{dla } t = 2h, \\ 2^h \pi^{h-1} r^t / t!!, & \text{dla } t = 2h-1, \end{cases}$
- dla $p = \infty$ mamy $V_t(r) = (2r)^t$.

Niech $Y(\tau)$, dla $\tau \geq 0$, oznacza liczbę par punktów $(W_i; W_j)_{i < j}$, dla których $D_{i,j} \leq \sqrt{\tau/\lambda(n)}$.

Wówczas dla dowolnego $\tau_1 > 0$ i wystarczająco dużego n , proces $\{Y(\tau), 0 \leq \tau \leq \tau_1\}$ zbiega do procesu Poissona z intensywnością 1.⁷⁰

Dla $\tau \leq \lambda(n)/2^t$ mamy:

$$E[Y(\tau)] = \tau \text{ i } D^2[Y(\tau)] = \tau - 2\tau^2/n(n-1).$$

Niech $T_i = \inf\{\tau \geq 0 : Y(\tau) \geq i\}$ dla $i = 1..n$ oraz $T_0 = 0$, określa punkty skokowe rozkładu $Y(\tau)$. Wówczas:

$$W_i^* = 1 - e^{T_{i-1} - T_i}, \text{ dla } i = 1..m \ll n,$$

są niezależnymi zmiennymi losowymi o rozkładzie równomiernym na przedziale $[0;1)$. Ich rozkład badany jest następnie przy pomocy testu Andersona-Darlinga.

W przypadku stosowania testu drugiego poziomu, autorzy zalecają następujące podejścia:

- zastosowanie testu zgodności dla otrzymanych prawdopodobieństw,
- połączenie W_i^* ze wszystkich iteracji testu w jeden zbiór, dla którego dopiero przeprowadza się test zgodności z rozkładem równomiernym Andersona-Darlinga,
- stworzenie Y – superpozycji procesów $Y(\tau)$ i dla nich przeprowadzenie dalszej części procedury testowej przy ustalonej wartości $\tau_1 > 0$; wówczas T_i mają rozkład równomierny na przedziale $[0; \tau_1]$, a J – liczba skoków Y w przedziale $[0; \tau_1]$, ma w przybliżeniu rozkład Poissona z wartością oczekiwaną $\tau_1 \cdot N$ – liczba powtórzeń testu.

Ostatnie podejścia zaleca się jedynie w przypadkach, gdy liczby u_i są reprezentowane z precyzją co najmniej 32 bitów, w przeciwnym wypadku będzie to znacząco wypaczało wyniki.

Zaleca się by $n \geq 4m^2\sqrt{N}$ i $t \leq 8$.

W pracy [Ma22]⁷¹ pokazujemy metodę znaczącego przyspieszenia testu.



⁷⁰ P. L'Ecuyer, J.F. Cordeau, R. Simard, *Close-Point Spatial Tests and Their Application to Random Number Generators*, Operations Research, 48, 2, 2000, s. 309

⁷¹ K. Mańk, *Speeding Up Minimum Distance Randomness Tests*, Journal of Telecommunications and Information Technology, 2022, vol. 2, s. 103-109

3.3. Test zgodności bitowej

Test różni się od poprzedniego sposobem wyznaczania odległości między parą punktów. Odległość pomiędzy punktami W_i oraz W_j wynosi $D_{i,j} = 2^{-b_{i,j}}$, gdzie $b_{i,j}$ jest maksymalną liczbą zgodnych początkowych bitów w binarnych rozwinięciach współrzędnych punktów. Odpowiada to takiej maksymalnej wartości b , że przy podziale hipersześcianu jednostkowego na 2^{tb} identycznych hipersześciennej komórek, oba punkty znajdują się w tej samej komórce.

Przy równomiernym rozkładzie punktów W_i w hipersześcianie jednostkowym mamy:

$$\Pr(b_{i,j} \geq b) = 2^{-tb}$$

oraz

$$q_b = \Pr(D \leq 2^{-b}) \approx 1 - (1 - 2^{-tb})^{n(n-1)/2},$$

gdzie $D = \min_{1 \leq i < j \leq n} \{D_{i,j}\}$. Dla $D = 2^{-b}$ otrzymujemy ponadto:

$$p_{val}^L = q_b \text{ oraz } p_{val}^R = 1 - q_{b-1}.$$

Jeśli procedurę testową powtórzymy N -krotnie, to w teście drugiego poziomu możemy posłużyć się również formułą:

$$\Pr\left(\min_{i=1..N} \{D_i\} \leq 2^{-b}\right) \approx 1 - (1 - 2^{-tb})^{Nn(n-1)/2}.$$

3.4. Test Bickela-Breimana

Test jest zbliżony przebiegiem do testu m najbliższych par, z punktu II.3.2. Dla każdego punktu wyznaczamy D_i – odległość do najbliższego innego punktu, a następnie $V_i(D_i)$ – objętość t -wymiarowej hiperkuli odpowiadającą tej odległości.

Statystyka testowa ma postać:

$$S_{obs} = \sum_{i=1}^n \left(W_{(i)}^* - i/n\right)^2,$$

gdzie $W_{(i)}^*$ są uporządkowanymi niemalejąco wartościami $W_i^* = 1 - e^{-n \cdot V_i(D_i)}$.

Poszczególne W_i^* mają w przybliżeniu rozkład równomierny na przedziale $[0;1)$, w teście mierzone jest więc ich odchylenie od równomierności.

Niestety statystyka testowa ma bardzo skomplikowany rozkład, wariancję udało się wyznaczyć jedynie dla przypadków $t=1$ oraz $t \rightarrow \infty$, zaś asymptotyczny rozkład został określony dla $n, t \rightarrow \infty$.⁷² Dla praktycznych przypadków znaleźć można symulacyjnie wyznaczone tablice interpolacyjne.⁷³

⁷² P.J. Bickel, L. Breiman, *Sums of Functions of Nearest Neighbor Distances, Moment Bounds, Limit Theorems and a Goodness of Fit Test*, The Annals of Probability, 11, 1, 1983, s. 185-214

⁷³ P. L'Ecuyer, J.F. Cordeau, R. Simard, *Close-Point Spatial Tests and Their Application to Random Number Generators*, wyd. cyt., s. 311

W pracy [Ma22]⁷⁴ pokazujemy metodę znaczącego przyspieszenia testu.

3.5. Test największy- z - t

Ciąg U^n dzielimy na N nienakładających się fragmentów po t liczb, a następnie z każdego z nich wybieramy największą. Do tak otrzymanego ciągu stosujemy test Kołmogorowa-Smirnowa, przyjmując $F(x) = x^t$ jako dystrybuantę teoretyczną lub też test równomierności, biorąc t -te potęgi każdego elementu.⁷⁵

3.6. Test permutacji

Ciąg U^n dzielimy na N nienakładających się t -elementowych wektorów, a następnie zliczamy wystąpienia każdego z $t!$ możliwych uporządkowań. Dla zliczeń, przyjmując równe prawdopodobieństwa wystąpienia każdej z permutacji, przeprowadzamy standardowy test zgodności chi-kwadrat.

W teście zakłada się, że wszystkie liczby w każdym t -elementowym wektorze są parami różne. Założenie to może prowadzić do istotnych błędów, gdy posługujemy się skończoną arytmetyką komputera, w takiej sytuacji musimy bądź dostosować do tego test, bądź zabezpieczyć implementację przed takim zdarzeniem. Kwestię tę omawiamy nieco szerzej w punkcie III.7.1.2.

3.7. Test parkingowy

Procedurę testową można wyobrazić sobie jako określenie liczby zakończonych powodzeniem prób parkowania latających spodków o promieniu jednostkowym na kwadracie o boku 100. Próba kończy się niepowodzeniem, jeśli w kole o promieniu 0,5 wokół punktu, którego współrzędne określają dwie kolejne liczby z badanego ciągu, znajduje się już zaparkowany wcześniej spodek, natomiast obwód spodka może wykraczać poza krawędzie parkingu. Współrzędne środka koła są liczbami rzeczywistymi z przedziału $[0;100)$.

Na podstawie eksperymentów określono, że dla liczby prób wynoszącej 12000, przeciętnie udaje się zaparkować 3523 spodki, z odchyleniem standardowym wynoszącym 23,9. Przyjęto, że statystyka testowa będąca liczbą sukcesów ma rozkład normalny o powyższych parametrach.⁷⁶

⁷⁴ K. Mańk, *Speeding Up Minimum Distance Randomness Tests*, Journal of Telecommunications and Information Technology, 2022, vol. 2, s. 103-109

⁷⁵ D.E. Knuth, *Sztuka programowania, t.2. Algorytmy seminumeryczne*, WNT, 2002, s. 73

⁷⁶ G. Marsaglia, *DIEHARD Battery of Tests of Randomness*, 1995, <http://stat.fsu.edu/pub/diehard/>

4. Testy dla ciągów nakładających się wektorów

W tym punkcie omawiamy testy operujące na wektorach t -elementowych, jednak odmienienie niż to było w poprzednim punkcie zakładamy, że do konstrukcji dwóch kolejnych wektorów wykorzystano tę samą sekwencję $t - 1$ bloków. Rozpatrywane są więc ciągi wektorów postaci:

$$(B_1, B_2, \dots, B_t), (B_2, B_3, \dots, B_{t+1}), (B_3, B_4, \dots, B_{t+2}), \dots, (B_{n-t+1}, B_{n-t+2}, \dots, B_n),$$

których elementy są d -bitowymi blokami.

4.1. Testy dla zliczeń liczby punktów w komórkach

Analogicznie jak dla nienakładających się wektorów (patrz. punkt II.3.1), badany będzie rozkład t -elementowych wektorów w t -wymiarowej kostce jednostkowej, dzielonej na k^t identycznych komórek. Zmianie ulega jednak sposób tworzenia wektorów. Przyjmujemy, że badany ciąg ma charakter cykliczny, tzn. $u_{n+i} = u_i$ dla $i = 1..t-1$, a następnie tworzymy ciąg wektorów testowych postaci $U_i^{(t)} = (u_i, u_{i+1}, \dots, u_{i+t-1})$, $i = 1..n$. Dla tego ciągu można, w zasadzie, przeprowadzić identyczne testy, jak dla wektorów nienakładających się, jedynym ograniczeniem jest trudność wyznaczenia rozkładów dla tych statystyk.

4.1.1. Test potęgowo-dywergencyjny dla przypadku gęstego

Dla przypadku „gęstego” i testu potęgowo-dywergencyjnego znany jest trick polegający na złożeniu dwóch statystyk. Niech v_i dla $i = 1..k^t$ oznaczają liczbę punktów, określonych przy użyciu wektorów $U_i^{(t)}$, które znalazły się w i -tej komórce. Statystyka

$$\tilde{D}_\delta^{(t)} = D_\delta^{(t)} - D_\delta^{(t-1)},$$

gdzie

$$D_\delta^{(t)} = \frac{2}{\delta(1+\delta)} \sum_{i=0}^{k^t-1} v_i \left((v_i/\lambda)^\delta - 1 \right) \text{ i } \delta > -1,$$

ma asymptotycznie rozkład chi-kwadrat z $k^t - k^{t-1}$ stopniami swobody.⁷⁷

Szczególnym przypadkiem omawianego testu jest test częstości nakładających się bloków, dla którego udało się uzyskać dokładną formułę określającą statystykę testową, co omawiamy w punkcie II.4.1.4.

4.1.2. Test potęgowo-dywergencyjny dla przypadku rzadkiego

Przypadek rzadki jest trudniejszy. Niech, dla ustalonego t , będą określone dwie statystyki Pearsona:

$$X_{(t)}^2 = \sum_{i=0}^{k^t-1} \frac{(v_i - \lambda_t)^2}{\lambda_t} \text{ i } X_{(t-1)}^2 = \sum_{i=0}^{k^{t-1}-1} \frac{(v_{i-1} - \lambda_{t-1})^2}{\lambda_{t-1}},$$

⁷⁷ P. L'Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, SIAM Journal on Scientific Computing, 24, 2, 2002, s. 659

gdzie, jak wyżej, v_{i_t} jest liczbą punktów w komórce, przy podziale t -wymiarowej kostki jednostkowej na k^t identycznych komórek, zaś $\lambda_t = n/k^t$ odpowiadającą temu przypadkowi wartością oczekiwaną.

Zmienna losowa

$$\frac{X_{(t)}^2 - X_{(t-1)}^2 - (k^t - k^{t-1})}{\sqrt{2(k^t - k^{t-1})}}$$

ma asymptotycznie rozkład $N(0;1)$.⁷⁸

4.1.3. Test liczby wolnych komórek

Wyniki Marsagli⁷⁹ dotyczące L_0 omówione są w punkcie II.4.3.

W [ES07a] podano, iż eksperymenty dawały poprawne wyniki po przyjęciu dla K_0 rozkładu Poissona, jak dla wektorów nienakładających się dla $\lambda \leq 1$.⁸⁰

4.1.4. Test częstości dla nakładających się wektorów

Jest to wariant testu potęgowo-dywergencyjnego, dla przypadku gęstego, oparty na rezultatach otrzymanych w [AKM04]⁸¹, [A04]⁸² i [Ma13]⁸³.

Badany będzie rozkład t -elementowych wektorów w t -wymiarowej kostce jednostkowej, dzielonej na k^t identycznych komórek.

Niech $v = [v_0, \dots, v_{k^t-1}]$ będzie wektorem zliczeń punktów, określonych przy użyciu wektorów $U_i^{(t)}$, które znalazły się w poszczególnych komórkach. Statystyka

$$S_{d,t}^\forall = \frac{1}{n} \sum_{w=1}^t \frac{1}{w} \sum_{i=1}^{L(w,k,t)} (\psi^{w,i} \circ v)^2,$$

gdzie sumowanie odbywa się po wszystkich niezerowych wartościach własnych w i ich wszystkich wektorach własnych $\psi^{w,i}$, ma asymptotycznie rozkład chi-kwadrat z $(k-1) \cdot k^{t-1}$ stopniami swobody.⁸⁴

⁷⁸ P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, wyd. cyt., s. 107

⁷⁹ G. Marsaglia, A. Zaman, *Monkey Tests for Random Number Generators*, Computers and Mathematics with Applications, 9, 1993, s. 3

⁸⁰ P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, wyd. cyt., s. 107

⁸¹ A. Alhakim, J. Kawczak, S. Molchanov, *On the class of nilpotent Markov chains, I: The spectrum of covariance operator*, Markov Processes and Related Fields 10, 2004, s. 629-652

⁸² A. Alhakim, *On the Eigenvalues and Eigenvectors of an Overlapping Markov Chain*, Probability Theory and Related Fields, 128, 2004, s. 589-605

⁸³ K. Mańk, *Test częstości dla nakładających się wektorów*, Cyberprzestępczość i ochrona informacji – Bezpieczeństwo w internecie tom II, Wydawnictwo Wyższej Szkoły Menedżerskiej w Warszawie, ISBN 978-83-7520-131-4, 2013

⁸⁴ A. Alhakim, J. Kawczak, S. Molchanov, *On the class of nilpotent Markov chains, I: The spectrum of covariance operator*, Markov Processes and Related Fields 10, 2004, s. 645

Liczbę wektorów własnych $\psi^{w,i}$ określa funkcja:

$$L(w, k, t) = (k-1)^{\min\{2, t-w+1\}} \cdot k^{\max\{0, t-w-1\}}.$$

W [A04] znaleźć można propozycję testu wykorzystującą jedynie wektory własne dla wartości własnej $w = 1$, wraz z algorytmem ich szybkiego wyznaczania.⁸⁵

Obliczanie obu wariantów statystyki testowej daje się najefektywniej zaimplementować dla przypadku, gdy $k = 2^d$. Sprowadza się ono wówczas do obliczenia, odpowiednio przekształconej, transformaty Walsha. Pozwala to zmniejszyć liczbę wykonywanych operacji z $2^{dt} (2^{dt} + 1)$ mnożeń i $2^{2dt} - 1$ dodawań w metodzie z macierzą kowariancji, do zaledwie $dt \cdot 2^{dt}$ dodawań przy wykorzystaniu szybkiej transformaty Walsha.⁸⁶

Macierzą Sylwestera-Walsha-Hadamarda nazywać będziemy macierz H_n wymiaru $2^n \times 2^n$, określoną następującym wzorem rekurencyjnym:

$$H_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad H_n = H_1 \otimes H_{n-1} = \begin{bmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{bmatrix}, \quad n = 2, 3, \dots$$

Poniżej prezentujemy schemat macierzy H_4 , wraz z oznaczeniem sposobu wyznaczania wektorów własnych dla wartości własnych 1 i 2. Kolumny odpowiadające wartości własnej 1 wprost są poszukiwanymi wektorami. Dla wartości własnej 2, należy oznaczone tym samym numerem kolumny zsumować, a następnie podzielić przez $\sqrt{2}$.

$H_4 =$	2				1				2				1			
	3				9				2				3			

wartość własna
nr wektora dla tej wartości

Kolejny przykład konstrukcji dotyczy przypadku $d = 2$ i $t = 3$, w którym posługujemy się macierzą H_6 . Znaczenie liczb w obu wierszach jest identyczne, jak powyżej.

	3				2				3				2				1				2				1			
	3				9				2				3				36				9				33			
	2				1				3				2				1				2				1			
	8				30				2				6				21				5				18			
	2				1				3				2				1				2				1			
	4				15				1				3				6				1				3			

36 wektorów własnych dla wartości własnej 1 danych jest wprost. Dla wartości własnej 2, wynik sumowania każdej z 9 par należy podzielić przez $\sqrt{2}$, a dla wartości własnej 3, wynik sumowania każdej z trzech trójek należy podzielić przez $\sqrt{3}$.

⁸⁵ A. Alhakim, *On the Eigenvalues and Eigenvectors of an Overlapping Markov Chain*, Probability Theory and Related Fields, 128, 2004, s. 594-595

⁸⁶ K. Mańk, *Test częstości dla nakładających się wektorów*, Cyberprzestępczość i ochrona informacji – Bezpieczeństwo w internecie tom II, Wydawnictwo Wyższej Szkoły Menedżerskiej w Warszawie, ISBN 978-83-7520-131-4, 2013

Ostatni przykład konstrukcji dotyczy przypadku $d = 1$ i $t = 6$, w którym również wykorzystać należy macierz H_6 , jednak przyporządkowanie wektorów jest inne.

	6	6	5	6	4	5	4	6	3	4	3	5	3	4	3	6	2	3	2	4	2	3	2
	1	1	1	1	2	1	1	1	4	2	3	1	2	1	1	1	8	4	7	2	6	3	5
5	2	3	2	4	2	3	2	6	1	2	1	3	1	2	1	4	1	2	1	3	1	2	1
1	4	2	3	1	2	1	1	1	16	8	15	4	14	7	13	2	12	6	11	3	10	5	9
5	1	2	1	3	1	2	1	4	1	2	1	3	1	2	1	3	1	2	1				
1	8	4	7	2	6	3	5	1	4	2	3	1	2	1	1								

Tutaj, jak zawsze, wektory dla wartości własnej 1 dane są wprost, przy czym jest ich 16. 8 wektorów własnych, dla wartości własnej 2, otrzymamy sumując po dwa oznaczone wektory i wynik dzieląc przez $\sqrt{2}$, dla wartości własnej 3 mamy cztery trójki wektorów, których sumę dzielimy przez $\sqrt{3}$. Następnie bierzemy dwie czwórki, których sumy dzielimy przez 2, otrzymując oba wektory własne dla 4. Kończymy sumując piątkę i szóstkę wektorów, dające po jednym wektorze własnym dla wartości własnych 5 i 6, przy czym otrzymane sumy podzielić należy odpowiednio przez $\sqrt{5}$ i $\sqrt{6}$.

Konstrukcja jest bardzo prosta w implementacji i wprost powiązana z wartościami d oraz t .

4.1.5. Test entropii

Test ten można zastosować dla przypadku wektorów elementów jednobitowych, tzn. $d = 1$.

Najpierw sprawdzana jest częstość występowania wszystkich t oraz $(t+1)$ -bitowych nakładających się wektorów w ciągu rozszerzonym o $t-1$ i t bitów. Oznaczmy te częstości: v_{i_t} , $i = 0..2^t - 1$ – dla bloków t -bitowych oraz $v_{i_{t+1}}$, $i = 0..2^{t+1} - 1$ – dla bloków o długości $t+1$.

Dla obu długości bloku obliczamy entropie odpowiadające uzyskanym częstościom:

$$\varphi^{(t)} = \frac{1}{n} \sum_{i=0}^{2^t-1} v_{i_t} \ln v_{i_t} - \ln n,$$

$$\varphi^{(t+1)} = \frac{1}{n} \sum_{i=0}^{2^{t+1}-1} v_{i_{t+1}} \ln v_{i_{t+1}} - \ln n.$$

Podane powyżej formuły zostały nieco zmienione w stosunku do oryginalnych, przez co nie wymagają uprzedniego przeliczenia liczby wystąpień na częstości wystąpień, czyli podzielenia wszystkich v_i przez n .

Statystyka testowa postaci

$$S_{obs} = 2n \left(\ln 2 - \varphi^{(t)} + \varphi^{(t+1)} \right),$$

ma w przybliżeniu rozkład chi-kwadrat z 2^t stopniami swobody.⁸⁷

⁸⁷ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001

Zaleca się, by $t \leq \lg n - 2$.

4.2. Test nakładających się permutacji

Test został zaproponowany przez George'a Marsaglię w pakiecie DIEHARD.⁸⁸

Dla ciągu liczbowego U^n określamy ciąg t pozycyjnych permutacji, przy czym dwie kolejne mają wspólne $t - 1$ liczb z badanego ciągu. Przy założeniu, że każda z t kolejnych liczb jest inna, permutacja określana jest jako kombinacja kolejności występujących w niej liczb w porządku rosnącym.

Dla każdej z permutacji S_i , $i = 0..t!-1$, zliczana jest liczba jej wystąpień v_i , a następnie zestaw ten badany jest przy użyciu testu zgodności chi-kwadrat. Ponieważ kolejne elementy ciągu permutacji są silnie skorelowane, więc wykorzystuje się statystykę testową postaci:

$$S_{obs} = \sum_{i,j} (v_i - \mu_i) c_{i,j}^{-1} (v_j - \mu_j),$$

gdzie

- μ_i jest wartością oczekiwaną zmiennej losowej v_i ,
- $c_{i,j}^{-1}$ są elementami słabej macierzy odwrotnej Moore'a–Penrose'a macierzy kowariancji $t!$ -elementowego wektora zmiennych v_i .

Statystyka testowa ma asymptotycznie rozkład chi-kwadrat z $t! - (t-1)!$ stopniami swobody.

Na potrzeby pakietu DIEHARD macierz $[c_{i,j}^{-1}]$ wyznaczono empirycznie, toteż szybko pojawiły się poważne przesłanki poddające w wątpliwość poprawność, użytej w teście, macierzy kowariancji, jak i liczby stopni swobody. Pierwsze, w pracach Wegenkittla^{89,90} oraz ich kontynuacjach opublikowanych wraz z Ecuyerem i Simardem^{91,92}. Ostateczne wyniki pojawiły się jednak dopiero w 10 lat po publikacji DIEHARDa, w pracy Alhakima, Kawczaka i Molchanova.⁹³

Z perspektywy tego testu, głównymi osiągnięciami Alhakima, Kawczaka i Molchanova było podanie formuły, pozwalającej wyznaczać macierz kowariancji dla wektora liczby odwiedzin dla poszczególnych stanów stacjonarnego procesu Markowa oraz określenie rzędu tej macierzy. W ten sposób można otrzymać, wykorzystywaną w teście, macierz kowariancji dla scentralizowanego wektora liczby wystąpień poszczególnych permutacji, której rząd równy jest liczbie stopni swobody statystyki testowej.

⁸⁸ G. Marsaglia, *DIEHARD Battery of Tests of Randomness*, 1995, <http://stat.fsu.edu/pub/diehard/>

⁸⁹ S. Wegenkittl, *Empirical Testing of Pseudorandom Number Generators*, praca dyplomowa, Naturwissenschaftlichen Fakultät der Universität Salzburg, 1996

⁹⁰ S. Wegenkittl, *Generalized ϕ -Divergence and Frequency Analysis in Markov Chains*, rozprawa doktorska, Naturwissenschaftlichen Fakultät der Universität Salzburg, 1998

⁹¹ P. L'Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, SIAM Journal on Scientific Computing, 24, 2, 2002, s. 652-668

⁹² P. L'Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, 1998

⁹³ A. Alhakim, J. Kawczak, S. Molchanov, *On the class of nilpotent Markov chains, I: The spectrum of covariance operator*, Markov Processes and Related Fields 10, 2004, s. 629-652

Na podstawie wspomnianej pracy udało nam się wydedukować formułę opisującą macierz kowariancji $C_t = [c_{ij}]_{t! \times t!}$, gdzie

$$c_{ii} = \pi_i \left(1 + 2(P_{ii}^1 + P_{ii}^2 + \dots + P_{ii}^{t-1}) - (2t-1)\pi_i \right), \quad i = 1..t!,$$

$$c_{ij} = \pi_i \left(P_{ij}^1 + P_{ij}^2 + \dots + P_{ij}^{t-1} \right) + \pi_j \left(P_{ji}^1 + P_{ji}^2 + \dots + P_{ji}^{t-1} \right) - (2t-1)\pi_i\pi_j, \quad i, j = 1..t!, \quad i \neq j,$$

zaś

- $\Pi = [\pi_i]_{t!}$ jest wektorem prawdopodobieństw wystąpienia poszczególnych permutacji (prawdopodobieństw granicznych znalezienia się w poszczególnych stanach),
- $P_{ij}^1, P_{ij}^2, \dots, P_{ij}^{t-1}$ są prawdopodobieństwami przejścia z permutacji (stanu) i do j w $1, 2, \dots, t-1$ krokach.

Jeśli przyjąć, tak jak to czyniono we wspomnianej pracy, że wszystkie prawdopodobieństwa graniczne są równe, to powyższe formuły znacznie się upraszczają, nie pozwala to jednak rozpatrywać przypadków dyskretnych, co okazuje się poważnym ograniczeniem.

Na potrzeby tej pracy wyznaczyliśmy wartości oczekiwane oraz słabe macierze odwrotne dla $t = 3, \dots, 6$ oraz długości bloku reprezentującego element permutacji $d = 1, \dots, 31$ i przypadku ciągłego.

4.3. Testy wolnych cel

Dla ciągu n nakładających się t -elementowych wektorów, których elementy są d -bitowymi blokami, wyznaczana jest statystyka testowa określająca liczbę tych spośród 2^{dt} wektorów, które w ciągu nie wystąpiły. Przyjmuje się, że dla odpowiednio dużych $d \cdot t$ i n , ma ona rozkład normalny.

Wyznaczenie parametrów testu nie nastręcza już dziś takich problemów, jak w momencie opracowywania pakietu DIEHARD. Już w roku 2009 z tematem tym zmierzył się Tomasz Jarocki.⁹⁴ Opracował on metodę wyznaczania wszystkich postaci słów dla zadanej liczby znaków w słowie t i liczby bitów reprezentujących każdy ze znaków d , a następnie wyznaczania dla każdego z nich macierzy stochastycznej. Po podniesieniu tej macierzy do potęgi, równej liczbie słów wyodrębnianych z sekwencji, odczytuje się wartość poszukiwanego prawdopodobieństwa niewystąpienia słowa danej postaci. Po ich skompletowaniu wyznacza się wartość oczekiwaną liczbę słów, które w sekwencji nie wystąpiły. Tą samą metodą, rozpatrując jedynie słowa o dwa razy większej liczbie znaków, wyznacza się moment drugiego rzędu, a więc i wariancję.

Aplikacja bazująca na pomysłe Jarockiego, wykonana w ramach realizacji tej pracy pozwoliła na wyznaczenie wartości oczekiwanych i odchyłeń standardowych statystyki testowej dla niemal wszystkich kombinacji t i d , których iloczyn nie przekracza 32 (nie wykonano obliczeń dla sześciu par: $t = 10$ i $d = 3$ oraz $t = 12..16$ i $d = 2$), oraz długości sekwencji 2^k dla $k = 7..30$.

⁹⁴ T. Jarocki, *Wyznaczenie dokładnych momentów pierwszego i drugiego rzędu w testach wolnych cel Marsagli*, praca dyplomowa, WAT, WCY, Kryptologia, 2009

4.4. Test „policz jedyńki w ciągu bloków”

Test został zaproponowany w pakiecie DIEHARD. Jest to wariacja testu niezależności wag Hamminga zastosowana dla nakładających się wektorów.

Na podstawie ciągu bloków $B^{d,n} = b_1, b_2, \dots, b_n$, tworzony jest najpierw ciąg liczb $z_i = Hwt(b_i)$ (wag Hamminga kolejnych bloków). Z kolei na jego podstawie tworzone są dwa ciągi nakładających się wektorów t - i $(t-1)$ -elementowych, dla którego wyznacza się częstości występowania poszczególnych wektorów.

W teście nie można wykorzystać prostej statystyki Pearsona ze względu na silne korelacje pomiędzy kolejnymi wektorami w ciągach, więc stosuje się wybieg opisany m.in. w [MOV96].⁹⁵

Niech $v_i^t, i = 1..(d+1)^t$ będą zaobserwowanymi częstościami wszystkich możliwych wektorów t -elementowych, zaś $v_i^{t-1}, i = 1..(d+1)^{t-1}$ – częstościami dla wektorów $(t-1)$ -elementowych. Statystyka testowa:

$$S_{obs} = \sum_{i=1}^{(d+1)^t} \frac{(v_i^t - E_i^t)^2}{E_i^t} - \sum_{i=1}^{(d+1)^{t-1}} \frac{(v_i^{t-1} - E_i^{t-1})^2}{E_i^{t-1}},$$

gdzie E_i^t są wartościami oczekiwanymi odpowiadającymi zmiennym losowym v_i^t , ma rozkład chi-kwadrat z $(d+1)^t - (d+1)^{t-1}$ stopniami swobody.

W razie zbyt małych wartości oczekiwanych E_i^t , należy zastosować grupowanie klas, odpowiednio modyfikując powyższe sumy i liczbę stopni swobody.

5. Testy dla ciągów bitów

Ostatnią grupą omawianych przez nas testów będą te działające wprost na ciąg bitów.

5.1. Test autokorelacyjny

Dla binarnego ciągu s^n obliczana jest seria empirycznych korelacji z jego niecyklicznym przesunięciem o $t = 1.. \lfloor n/2 \rfloor$ pozycji:

$$A_t = \sum_{i=1}^{n-t} s_i \oplus s_{i+t}.$$

A_d mają rozkład Bernoulliego z parametrami $n-t$ oraz $1/2$, więc dla dużych wartości różnicy $n-t$ zmienna losowa

$$\frac{2A_t - (n-t)}{\sqrt{n-t}}$$

ma w przybliżeniu standardowy rozkład normalny.

⁹⁵ A. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, s. 181

W przygotowanej implementacji wykorzystaliśmy statystykę wyznaczoną dla testu momentu pierwszego rzędu dla ciągu bloków jednobitowych (patrz II.2.2.2).

5.2. Test liczby serii

Test pochodzi z pakietu STS, bada zgodność zaobserwowanej liczby serii jedynek lub zer z rozkładem teoretycznym. Swoistą nowością jest tu podwójne odniesienie się do testu częstości bitów – testu serii nie przeprowadza się, jeśli ciąg nie przeszedł tego pierwszego oraz do wyznaczenia statystyki testowej wykorzystywana jest zaobserwowana dla danego ciągu częstość występowania jedynek – oznaczana tu jako π , nie jest zaś jak zazwyczaj przyjmowana z góry wartość 0,5.

Dla S_{obs} – liczby serii w badanym ciągu pokazano, że dla dużych n ma ona w przybliżeniu rozkład normalny o wartości oczekiwanej $2n\pi(1-\pi)$ i odchyleniu standardowym $2\sqrt{n\pi(1-\pi)}$.

Zalecana minimalna długość ciągu to $n = 100$ bitów.

Znaleźć można dokładniejsze wartości, zarówno dla wartości oczekiwanej, jak i odchylenia standardowego:

$$E[S_{obs}] = 2n\pi(1-\pi) + \pi^2 + (1-\pi)^2,^{96}$$

$$\sqrt{D^2[S_{obs}]} = 2\sqrt{n\pi(1-\pi)(1-3\pi(1-\pi))}.^{97}$$

5.3. Test serii z macierzą kowariancji

Przedstawiamy dwa warianty testu, w pierwszym serie zer i jedynek zliczane są oddzielnie, w drugim – łącznie.

Dla ciągu binarnego długości n wyznaczamy liczbę serii zer i, oddzielnie, jedynek, poszczególnych długości. Niech v_i^0 , $i = 1, \dots, K-1$ będą liczbami serii zer długości i , zaś v_K^0 liczbą serii zer długości co najmniej K . Dla jedynek mamy odpowiednio v_i^1 dla $i = 1, \dots, K-1$ oraz v_K^1 .

Mood⁹⁸ pokazał, że zmienne losowe:

$$x_i = \frac{v_i^0 - R_i}{\sqrt{n}}, \quad i = 1, \dots, K-1,$$

$$x_K = \frac{v_K^0 - R_K}{\sqrt{n}},$$

$$x_{K+i} = \frac{v_i^1 - R_i}{\sqrt{n}}, \quad i = 1..K-1,$$

$$x_{2K} = \frac{2n^0 - n}{2\sqrt{n}},$$

⁹⁶ A.M. Mood, *The Distribution Theory of Runs*, Annals of Mathematical Statistics, 11, 1940, s. 389

⁹⁷ J. Wishart, H.O. Hirshfeld, *A Theorem Concerning the Distribution of Joins Between Line Segments*, Journal of the London Mathematical Society, 11, 1936, s. 227

⁹⁸ A.M. Mood, *The Distribution Theory of Runs*, Annals of Mathematical Statistics, 11, 1940, s. 389

gdzie:

$R_i = \frac{n-i+3}{2^{i+2}}$ jest oczekiwaną liczbą serii zer albo jedynek długości i ,

$R_K = \frac{n-K+2}{2^{K+1}}$ jest oczekiwaną liczbą serii zer albo jedynek długości co najmniej K ,

n^0 jest liczbą zer w ciągu,

mają asymptotycznie rozkład normalny, z wartością oczekiwaną 0 i macierzą wariancji i kowariancji $[\sigma_{i,j}]_{2K \times 2K}$ postaci:

$$\begin{aligned}\sigma_{i,i} &= \sigma_{K+i,K+i} = 1/2^{i+2} - (2i-1)/2^{2i+4}, \\ \sigma_{i,j} &= \sigma_{K+i,K+j} = -(i+j-1)/2^{i+j+4}, \quad i \neq j, \\ \sigma_{i,K} &= -(K+i)/2^{K+i+3}, \\ \sigma_{i,K+j} &= -(i+j-5)/2^{i+j+4}, \\ \sigma_{i,2K} &= -\sigma_{K+i,2K} = (i-2)/2^{i+3}, \\ \sigma_{K,K} &= 1/2^{K+1} - (2K+1)/2^{2K+2}, \\ \sigma_{K,K+i} &= -(K+i-4)/2^{K+i+3}, \\ \sigma_{K,2K} &= (K-1)/2^{K+2}, \\ \sigma_{2K,2K} &= 1/4,\end{aligned}$$

dla $i, j = 1..K-1$.

Statystyka testowa

$$Q^1 = \sum_{i=1}^{2K} \sum_{j=1}^{2K} \sigma^{i,j} x_i x_j,$$

gdzie $[\sigma^{i,j}]_{2K \times 2K}$ jest macierzą odwrotną do powyższej macierzy kowariancji, ma asymptotycznie rozkład chi-kwadrat z $2K$ stopniami swobody.

W drugim wariancie wyznaczamy liczbę serii, łącznie zer i jedynek, poszczególnych długości. Niech v_i , $i = 1..K-1$ będą liczbami serii długości i , zaś v_K – liczbą serii długości co najmniej K . Podobnie, jak poprzednio mamy:

$$\begin{aligned}x_i &= \frac{v_i - E[v_i]}{\sqrt{n}}, \quad i = 1..K-1, \\ x_K &= \frac{v_K - E[v_K]}{\sqrt{n}},\end{aligned}$$

gdzie

$E[v_i] = \frac{n-i+3}{2^{i+1}}$ jest oczekiwaną liczbą serii długości i ,

$E[v_K] = \frac{n-K+2}{2^K}$ jest oczekiwaną liczbą serii długości co najmniej K ,

mają asymptotycznie rozkład normalny, z wartością oczekiwaną 0 i macierzą wariancji i kowariancji $[\sigma_{i,j}]_{K \times K}$ postaci:

$$\begin{aligned}\sigma_{i,j} &= \sigma_{j,i} = (3-i-j)/2^{i+j+2}, \\ \sigma_{i,i} &= 1/2^{i+1} + (3-2i)/2^{2i+2}, \\ \sigma_{i,K} &= \sigma_{K,i} = (2-K-i)/2^{K+i+1}, \\ \sigma_{K,K} &= 1/2^K + (1-2K)/2^{2K},\end{aligned}$$

gdzie $i, j = 1..K-1$ i $i \neq j$.

Statystyka testowa

$$Q = \frac{1}{n} \sum_{i,j=1}^K \sigma^{i,j} x_i x_j,$$

gdzie $[\sigma^{i,j}]_{K \times K}$ jest macierzą odwrotną do powyższej macierzy kowariancji, ma asymptotycznie rozkład chi-kwadrat z K stopniami swobody.⁹⁹



Korelacje pomiędzy liczbami serii poszczególnych długości są na tyle znaczące, że nawet dla bardzo długich ciągów nie jest możliwe wykorzystanie prostego testu zgodności chi-kwadrat.¹⁰⁰

5.4. Test do liczby serii N

Badany ciąg binarny s^n przeglądany jest dopóty, dopóki nie zostanie zaobserwowanych po N serii zer i jedynek. Niech v_i^0 , $i = 1..k-1$ będzie zaobserwowaną liczbą serii zer długości i , zaś v_k^0 zaobserwowaną liczbą serii zer długości co najmniej k , analogicznie mamy v_i^1 oraz v_k^1 dla jedynek. k wybierane jest tak, aby uzyskać założoną minimalną wartość oczekiwaną dla każdego v_i – $minEv$:

$$k = 1 + \lfloor \lg(N/minEv) \rfloor.$$

Pierwszy test bada, przy użyciu testu chi-kwadrat, zgodność zaobserwowanych rozkładów v_i^0 oraz v_i^1 z rozkładem teoretycznym. Wyznaczana jest statystyka:

$$S_{obs} = \sum_{i=1}^k \frac{(v_i^0 - Np_i)^2 + (v_i^1 - Np_i)^2}{Np_i(1-p_i)},$$

gdzie

$$p_i = \begin{cases} 2^{-i}, & i = 1..k-1, \\ 2^{-k+1}, & i = k, \end{cases}$$

jest prawdopodobieństwem wystąpienia serii o długości i lub o długości co najmniej k .

⁹⁹ A.M. Mood, *The Distribution Theory of Runs*, Annals of Mathematical Statistics, 11, 1940, s. 390

¹⁰⁰ K. Mańk, *Testy losowości wykorzystujące statystykę chi-kwadrat*, referat w ramach XXXVII Ogólnopolskiej Konferencji Zastosowań Matematyki, Zakopane-Kościelisko, 2008

S_{obs} ma asymptotycznie rozkład chi-kwadrat z $2(k-1)$ stopniami swobody.

Drugi test polega na sprawdzeniu zgodności długości ciągu n , koniecznej do zaobserwowania po N serii zer i jedynek, z rozkładem normalnym o średniej $4N$ i odchyleniu standardowym $2\sqrt{N}$.

5.5. Testy losowych wycieczek

Badany ciąg binarny s^{NL} dzielony jest na N rozłącznych fragmentów długości L bitów. Dla każdego z nich wyznaczany jest ciąg sum częściowych postaci:

$$S_l = \sum_{i=1}^l (2s_i - 1), \quad l = 1..L \text{ i } 2 \mid L,$$

na którego podstawie wyznaczane jest sześć statystyk:

- liczba jedynek w ciągu

$$H = L/2 + S_L/2 = \sum_{i=1}^L s_i,$$

- maksymalne odchylenie ścieżki

$$M = \max_{l=1..L} \{S_l\},$$

- łączny „czas” odchylenia w prawo

$$J = 2 \sum_{i=1}^{L/2} I[S_{2i-1} > 0],$$

- moment pierwszego osiągnięcia wartości y

$$P_y = \min \{l : S_l = y\},$$

- liczba powrotów do 0

$$R = \sum_{i=1}^{L/2} I[S_{2i} = 0],$$

- liczba przejść przez 0 ze zmianą strony

$$C = \sum_{i=3}^L I[S_{i-2} S_i < 0].$$

Niech $p_{L,2k-L} = 2^{-L} \binom{L}{k}$, wówczas rozkłady tych statystyk określone są następująco:

- $\Pr(H = k) = \Pr(S_L = 2k - L) = p_{L,2k-L},$
- $\Pr(M = k) = p_{L,k} + p_{L,k+1}, \quad 0 \leq k \leq L,$
- $\Pr(J = k) = p_{k,0} \cdot p_{L-k,0}, \quad 0 \leq k \leq L \text{ i } 2 \mid k,$
- $\Pr(P_y = k) = (y/k) p_{k,y},$
- $\Pr(R = k) = p_{L-k,k}, \quad 0 \leq k \leq L/2,$

- $\Pr(C = k) = 2p_{L-1,2k+1}, 0 \leq k \leq (L-1)/2.$

Procedurę testową kończy porównanie, przy pomocy testu zgodności chi-kwadrat, rozkładów otrzymanych wartości z odpowiadającymi rozkładami teoretycznymi.¹⁰¹

5.6. Testy losowych wycieczek z liniowym wyznaczaniem kierunku

Procedura testowa jest identyczna z opisaną w poprzednim punkcie, z tą różnicą, że badany jest ciąg y_i otrzymany z wyjściowego na podstawie liniowego przekształcenia:

$$y_i = \sum_{j=0}^{t-1} c_j s_{i+j}, t > 0,$$

gdzie $c_j \in \{0;1\}$, $j = 0..t-1$ są współczynnikami wyznaczanej kombinacji liniowej. Jeśli tylko $c_0 = 1$, to otrzymujemy dokładnie poprzedni przypadek.

c_j można również utożsamić ze współczynnikami wielomianu charakterystycznego, widać wówczas, że tak skonstruowana procedura ma na celu wykrywanie raczej krótkozakresowych zależności liniowych w badanym ciągu. Należy pamiętać, że liczba możliwych kombinacji liniowych rośnie wykładniczo wraz ze wzrostem ich rozpiętości.

5.7. Testy złożoności liniowej ciągu

Są to trzy testy, wykorzystujące złożoność liniową i profil złożoności liniowej, otrzymywane przy użyciu algorytmu Berlekampa-Masseya.

5.7.1. Test liczby skoków w profilu liniowym

Test polega na zliczeniu liczby skoków w profilu liniowym. Przyjmuje się, że dla dużych n liczba skoków (oznaczana J) ma w przybliżeniu rozkład normalny z wartością oczekiwaną

$$E[J] = \frac{n}{4} + \frac{9 - (-1)^n}{24} - \frac{1}{3 \cdot 2^n}$$

i wariancją

$$D^2[J] = \frac{n}{8} - \frac{3 + (-1)^n}{17 + (-1)^n} + \frac{n}{3 \cdot 2^{n+1}} + \frac{4 - 3(-1)^n}{6 - 3(-1)^n} \frac{1}{2^n} - \frac{1}{9 \cdot 2^{2n}}.^{102}$$

5.7.2. Test rozkładu wysokości skoków w profilu liniowym

Drugi test polega na zbadaniu rozkładu zaobserwowanej liczby skoków poszczególnych wysokości. Wydaje się, że autorzy przyjęli, wartość oczekiwaną liczby skoków wysokości k jako

$$n/2^{k+2}.$$

¹⁰¹ P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, Universite de Montreal, 2002

¹⁰² H. Niederreiter, *The Linear Complexity Profile and the Jump Complexity of Keystream Sequences*, Lecture Notes in Computer Science, 473, 1991, s. 178-179

Elementarnymi metodami można wyznaczyć lepszą, od przyjętej przez autorów, postać wyrażenia dla wartości oczekiwanej liczby skoków wysokości k :

$$\frac{\left\lfloor \frac{n-k}{2} \right\rfloor + 2}{2^{k+1}}, \text{ dla } k \leq n.$$

Różnica w otrzymywanych wartościach ma charakter bezwzględny, np. dla parzystego n i $k=1$ w ten sposób otrzymamy wartość o $1/4$ większą, tak więc błąd względny będzie dążył do zera wraz z wydłużaniem ciągu.

5.7.3. Test rozkładu złożoności liniowej podciągów

Test bada złożoność liniową ciągu. Ciąg dzielony jest na N rozłącznych fragmentów długości M , a dla każdego z nich, przy użyciu algorytmu Berlekampa-Masseya, wyznaczana jest jego złożoność liniowa L_i , $i = 1..N$. Następnie obliczana jest

$$T_i = (-1)^M (L_i - \mu) + 2/9,$$

gdzie

$$\mu = \frac{M}{2} + \frac{9 + (-1)^{M+1}}{36} - \frac{3M + 2}{9 \cdot 2^M}$$

jest wartością oczekiwaną złożoności liniowej ciągu o długości M .

Formuła określająca T_i została tak zaprojektowana, by, po pominięciu trzeciego członu w wyrażeniu dla μ , otrzymywać liczby całkowite, dla których prawdopodobieństwa uzyskania wyrażają się następująco:

$$\Pr(T_i = k) = \begin{cases} 1/2, & k = 0, \\ 1/2^{2^k}, & k > 0, \\ 1/2^{-2^{k+1}}, & k < 0. \end{cases}$$

Na tej podstawie buduje się test zgodności chi-kwadrat.

5.8. Test złożoności Lempela-Ziva

Autorzy wykorzystali metodę przyrostowego generowania słownika, wykorzystaną w dowodzie twierdzenia opublikowanego w pracy Ziva i Lempela.¹⁰³

Niech statystyką testową S_{obs} będzie liczbą różnych wzorców (długością słownika) w ciągu o długości $n = 2^k$, wówczas dla dużych n przyjęto, że ma ona rozkład normalny ze średnią $n/\lg n = n/k$ i odchyleniem standardowym $\sqrt{0,266n/\lg^3 n} = \sqrt{0,266n/k^3}$.¹⁰⁴

¹⁰³ J. Ziv, A. Lempel, *Compression of Individual Sequences via Variable-Rate Coding*, IEEE on Information Theory, 24, 5, 1978, s. 533

¹⁰⁴ P. Kirschenhofer, H. Prodinger, W. Szpankowski, *Digital Search Trees Again Revisited: The Internal Path Length Perspective*, SIAM Journal on Computation, 23, 1994, s. 598-616

W pracy [SUH04] znaleźć można znaczące poprawki dotyczące rozkładu S_{obs} i jego parametrów. Autorzy wykazali, że analizowany rozkład ma charakter asymetryczny, wyznaczyli dla niego średnią: $\mu = 69588,09$ oraz parę wariancji dla wartości mniejszych i większych od μ : $\sigma_L^2 = 75,574336518$ i $\sigma_R^2 = 72,42178447$. Zwrócili oni również uwagę na dyskretny charakter rozpatrywanego rozkładu, co skutkuje koniecznością korekty, stosowanej w pakiecie, metody badania równomierności rozkładu uzyskiwanych prawdopodobieństw p_{val} .¹⁰⁵

Wyniki uzyskane w trakcie realizacji projektu p.k. ROTOR¹⁰⁶ potwierdziły słuszność wycofania tego testu z pakietu NIST¹⁰⁷, tak więc nie był on przez nas rozpatrywany w dalszej części pracy.

5.9. Testy spektralne

5.9.1. Test spektralny

Procedura testowa polega na sprawdzeniu liczby pików w dyskretnej transformacie Fouriera (DTF) ciągu $(-1)^{s_i}$, których amplituda mieści się w górnych 5% wartości – korzysta się z faktu, iż dla ciągu o długości n przeciętnie 95% elementów DTF ma moduł nieprzekraczający $T = \sqrt{-n \cdot \ln(0,05)} \approx \sqrt{2,995732274 \cdot n}$.¹⁰⁸

Statystyka testowa ma postać:

$$S_{obs} = \frac{2N - 0,95 \cdot n}{\sqrt{0,0125 \cdot n}},$$

gdzie N jest liczbą elementów DTF o numerach 1 do $n/2$, których moduł nie przekracza T .

Powyższy opis uwzględnia poprawki wynikające z prac [SUH04]¹⁰⁹ i [PRS12]¹¹⁰.

Zaleca się badanie ciągów o długości $n = 2^k$, gdzie $k = 8..20$, jak również stosowanie bardzo małej liczby powtórzeń przy badaniu drugiego poziomu.

5.9.2. Test spektralny sumy kwadratów modułów – wariant 1

Test ten również bada wartości $|f_l|$ – modułów elementów dyskretnej transformaty Fouriera, statystyka testowa ma jednak postać:

$$S_{obs} = \frac{1}{n} \sum_{l=1}^{n/4} |f_l|^2.$$

¹⁰⁵ K. Song-Ju, K. Umeno, A. Hasegawa, dz. cyt., s. 11-13

¹⁰⁶ P. Szabuńko, *Wykonanie implementacji i analiza teoretyczna testu złożoności sekwencji wykorzystującego algorytm kompresji Lempel-Ziv*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2018

¹⁰⁷ L. Bassham, *Validation Testing and NIST Statistical Test Suite*, 2004

¹⁰⁸ K. Song-Ju, K. Umeno, A. Hasegawa, *Corrections of the NIST Statistical Test Suite for Randomness*, Cryptology ePrint Archive, Report 2004/018, s. 9

¹⁰⁹ K. Song-Ju, K. Umeno, A. Hasegawa, *Corrections of the NIST Statistical Test Suite for Randomness*, Cryptology ePrint Archive, Report 2004/018, s. 9

¹¹⁰ F. Pareschi, R. Rovatti, G. Setti, *On statistical tests for randomness included in the NIST SP800-22 test suite and based on the binomial distribution*, IEEE Transactions on Information Forensics and Security 7(2), 2012

Ponieważ dla $l \neq 0$ zmienna losowa $|f_l|^2/n$ ma wartość oczekiwaną równą 1, zaś wariancję $1 - 2/n$, więc dla odpowiednio dużych n przyjąć można, że S_{obs} ma rozkład normalny ze średnią $n/4$ i odchyleniem standardowym $\sqrt{n-2}/2$.

Zaleca się badanie ciągów o długości $n = 2^k$, gdzie $k = 4..26$ oraz niestosowanie badania przy użyciu testu drugiego poziomu.

5.9.3. Test spektralny sumy kwadratów modułów – wariant 2

Dla N ciągów, każdy o długości n bitów, wyznaczamy, jak poprzednio, $|f_l^{(j)}|$ dla $l = 1..n/4$, gdzie $j = 1..N$ jest numerem ciągu. Następnie dla każdego l wyznaczana jest suma:

$$S_l = \frac{1}{n} \sum_{j=1}^N |f_l^{(j)}|^2, \quad l = 1..n/4.$$

Dla dużych N każda S_l ma w przybliżeniu rozkład normalny ze średnią N i odchyleniem standardowym $\sqrt{N(1-2/n)}$. Rozkład uzyskanej serii wartości S_l porównywany jest następnie z rozkładem teoretycznym.

Jak poprzednio zaleca się $n = 2^k$ i $k = 4..26$ oraz $N \geq 2^k$.

Wyniki uzyskane w trakcie realizacji projektu p.k. ROTOR¹¹¹ wskazały, że jedynie wariant 2 testu sumy kwadratów modułów powinien być wykorzystywany podczas masywnych badań generatorów, tak więc jedynie on był przez nas rozpatrywany w dalszej części pracy.

5.10. Test okresowości w ciągu

Test wykorzystuje algorytm opisany w pracy [GO81]¹¹², oparty na tzw. ciągach korelacyjnych. Dla ciągu binarnego $s^n = s_1, s_2, \dots, s_n$, ciąg korelacyjny $c^n = c_0, c_1, \dots, c_{n-1}$ jest również ciągiem binarnym takim, że $c_p = 1$ wtedy i tylko wtedy, gdy p jest okresem ciągu s^n , tzn. $s_i = s_{i+p}$ dla $i = 1..n-p$, w szczególności zawsze zachodzi $c_0 = 1$.

Badany ciąg dzielony jest na N rozłącznych fragmentów, każdy długości M bitów. Dla każdego określany jest następnie odpowiadający mu ciąg korelacyjny. Po zliczeniu wystąpień każdego z możliwych ciągów korelacyjnych ich rozkład porównywany jest z rozkładem teoretycznym przy użyciu testu zgodności chi-kwadrat. Sposób wyznaczenia rozkładu teoretycznego znaleźć można we wspomnianej wyżej pracy [GO81]¹¹² oraz w pracy [RR03]¹¹³. Zaleca się, by $2 \leq M \leq 31$.

¹¹¹ K. Mańk, *Implementacja testów opartych o dyskretną transformatę Fouriera*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2018

¹¹² L.J. Guibas, A.M. Odlyzko, *Periods in Strings*, Journal of Combinatorial Theory, 20, 1981, s. 19-42

¹¹³ E. Rivals, S. Rahmann, *Combinatorics of Periods in Strings*, Journal of Combinatorial Theory, 104, 2003, s. 95-113

5.11. Test najdłuższej serii jedynek w bloku

Test bada rozkład wartości długości najdłuższej serii jedynek w blokach ustalonej długości M , na które dzielony jest badany ciąg. Statystyka testowa ma postać:

$$S_{obs} = \sum_{i=0}^K \frac{(v_i - N\pi_i)^2}{N\pi_i},$$

gdzie

N – liczba bloków długości M , na które dzielony jest badany ciąg,

v_i – zaobserwowana liczba długości najdłuższych serii jedynek z i -tej klasy,

$N\pi_i$ – wartość oczekiwana dla powyższej,

$K+1$ – liczba klas, na jaką podzielone zostały obserwowane długości.

S_{obs} ma asymptotycznie rozkład chi-kwadrat z K stopniami swobody.

Podziału na klasy dokonuje się tak, by spełnione były założenia statystyki chi-kwadrat dotyczące wartości oczekiwanej dla każdej z klas.

Wartości π_i można utożsamiać z prawdopodobieństwami wystąpienia w pojedynczym bloku najdłuższej serii jedynek o długości odpowiadającej i -tej klasie.

Niech r będzie liczbą jedynek w danym bloku, wówczas prawdopodobieństwo, że długość ν najdłuższej serii jedynek w tym bloku jest nie większa niż η , wyraża się następująco:

$$\Pr(\nu \leq \eta \mid r) = \frac{1}{\binom{M}{r}} \sum_{i=0}^R (-1)^i \binom{M-r+1}{i} \binom{M-i(\eta+1)}{M-r},$$

gdzie $R = \min\left(M-r+1, \left\lfloor \frac{r}{\eta+1} \right\rfloor\right)$.

Pozwala to wyznaczyć prawdopodobieństwo bezwarunkowe:

$$\Pr(\nu \leq \eta) = \frac{1}{2^M} \sum_{r=0}^M \binom{M}{r} \Pr(\nu \leq \eta \mid r).$$

5.12. Testy wag Hamminga

5.12.1. Test częstości wag Hamminga

Badany ciąg dzielony jest na N rozłącznych fragmentów długości M bitów. Dla każdej z $M+1$ możliwych wag Hamminga zliczana jest liczba fragmentów odpowiadająca tej wartości – v_i , dla $i = 0..M$. Zaobserwowany rozkład wartości v_i jest następnie porównywany z rozkładem teoretycznym za pomocą testu zgodności chi-kwadrat.

Prawdopodobieństwo, że losowo wybrany blok długości M bitów ma wagę Hamminga równą w , określone jest następująco:

$$\Pr(Hwt(s^M) = w) = \binom{M}{w} \cdot 2^{-M}.$$

5.12.2. Test wag Hamminga

Badany ciąg dzielony jest na N rozłącznych fragmentów długości M bitów. Dla każdego wyznaczana jest jego waga Hamminga – π_i , $i = 1..N$, a następnie obliczana jest statystyka testowa:

$$S_{obs} = \frac{4}{M} \sum_{i=1}^N (\pi_i - M/2)^2,$$

która ma w przybliżeniu rozkład chi-kwadrat z N stopniami swobody.

M powinno być duże, aby rozkład poszczególnych π_i , który jest rozkładem Bernoulliego, można było przybliżać rozkładem normalnym.

Stosowanie nieparzystych M wprowadza, zbiegające do zera, obciążenie statystyki testowej.

5.12.3. Test korelacji wag Hamminga

Jak poprzednio wyznaczane są wagi Hamminga dla N rozłącznych fragmentów ciągu, każdy długości M bitów – π_i . Na ich podstawie obliczana jest empiryczna korelacja pomiędzy następującymi po sobie fragmentami:

$$\hat{\rho}_{obs} = \frac{4}{M \sqrt{N-1}} \sum_{i=1}^{N-1} (\pi_i - M/2)(\pi_{i+1} - M/2).$$

Dla dużych M i N , $\hat{\rho}_{obs}$ ma, w przybliżeniu, standardowy rozkład normalny.

5.12.4. Testy niezależności wag Hamminga

Ciąg dzielony jest na $2N$ rozłącznych fragmentów po M bitów, dla których wyznaczane są wagi Hamminga π_i , $i = 1..2N$. Pierwszy z testów polega na przeprowadzeniu testu częstości chi-kwadrat dla N par wag postaci $(\pi_{2j-1}; \pi_{2j})$, $j = 1..N$.

W drugim teście zaobserwowane częstości występowania poszczególnych par zapisywane są w macierzy stopnia $M + 1$. Następnie dla ustalonego $d = 1..[(M + 1)/2]$ macierz ta dzielona jest na trzy obszary według następującego schematu:

$$\left[\begin{array}{cc|c|cc} \hline & & & & \\ \hline & Y_2 & & & Y_1 \\ \hline & & & & \\ & & Y_3 & & \}^D \\ & & & & \\ \hline & Y_1 & & & Y_2 \\ \hline & & & & \\ & & & & \\ \hline & & & & \\ \hline \end{array} \right],$$

gdzie D jest liczbą centralnych wierszy i kolumn, określoną następująco:

$$D = \begin{cases} 2d-1, & 2 \mid M, \\ 2d-2, & 2 \nmid M. \end{cases}$$

Dla każdego obszaru obliczana jest suma elementów do niego należących – liczba par wag odpowiadających danemu obszarowi. Zachodzi $Y_3 = N - Y_1 - Y_2$. Dla tych trzech wartości stosuje się test zgodności chi-kwadrat.

Szczególnym przypadkiem jest $d = 1$ i M nieparzyste, gdyż wówczas $D = 0$ i $Y_3 = 0$, w takim przypadku test zgodności przeprowadzany jest jedynie dla Y_1 i Y_2 , a statystyka testowa ma 1 stopień swobody.

5.13. Test rzędu macierzy binarnej

Test bada występowanie w ciągu zależności liniowych. Ciąg dzielony jest na nienakładające się Q bitowe bloki, traktowane jako kolejne wiersze macierzy o wymiarach $M \times Q$, dla których wyznaczany jest rząd. Następnie zliczenia wystąpień poszczególnych wartości rzędu porównywane są z rozkładem teoretycznym przy użyciu testu zgodności chi-kwadrat. Potrzebne wartości oczekiwane wyznaczane są przy pomocy prawdopodobieństwa, że losowo wybrana macierz wymiaru $M \times Q$ posiada rząd r :

$$p_r = 2^{r(M+Q-r)-MQ} \prod_{i=0}^{r-1} \frac{(1-2^{i-Q})(1-2^{i-M})}{1-2^{i-r}}.$$

5.14. Test nienakładających się wzorców

Ciąg dzielony jest na N rozłącznych fragmentów długości M , a następnie dla każdego z nich zliczana jest v_i , gdzie $i = 1..N$ – liczba wystąpień ustalonego d -bitowego bloku – wzorca. Wzorce muszą być acykliczne, co zapewnia to, że żadne dwa jego wystąpienia się nie nałożą, a dzięki temu upraszcza się statystyka testowa.

Przyjmuje się, że zmienne losowe opisujące v_i mają asymptotycznie rozkład normalny o wartości oczekiwanej $\mu = (M - d + 1)/2^d$ i wariancji $\sigma^2 = M \left(\frac{1}{2^d} - \frac{2d-1}{2^{2d}} \right)$, więc statystyka testowa

$$S_{obs} = \sum_{i=1}^N \frac{(v_i - \mu)^2}{\sigma^2}$$

ma, w przybliżeniu, rozkład chi-kwadrat z N stopniami swobody.

5.15. Test nakładających się wzorców

Badany ciąg dzielony jest na N rozłącznych fragmentów, każdy długości M . Dla każdego z nich sprawdzana jest liczba wystąpień ustalonego d -bitowego wzorca. Z założenia wzorec ma charakter cykliczny, co sprowadza się w praktyce do ciągu d jedynek albo zer, toteż dopuszcza się

dowolne nakładanie się wystąpień wzorca w obrębie ciągu. Wyróżnia się $K + 1$ kategorii tak, że v_0 oznacza liczbę fragmentów, w których wzorec nie wystąpił ani razu, v_1 – liczbę z pojedynczym wystąpieniem, itd., aż do v_K , które oznacza liczbę fragmentów, w których wzorec wystąpił co najmniej K razy. Na tej podstawie obliczana jest statystyka testowa

$$S_{obs} = \sum_{i=0}^K \frac{(v_i - N\pi_i)^2}{N\pi_i},$$

która ma asymptotycznie rozkład chi-kwadrat z K stopniami swobody. Prawdopodobieństwa π_i , i -krotnego wystąpienia we fragmencie wzorca, wyznaczone są na podstawie rekurencyjnej formuły

$$\pi_0(m) = \begin{cases} 2, & m = -1, \\ 1, & 0 \leq m \leq d-1, \\ \pi_0(m-1) - 2^{-d-1}\pi_0(m-d-1), & d \leq m \leq M, \end{cases}$$

$$\pi_1(m) = \begin{cases} 0, & m \leq d-1, \\ 2^{-d}, & d \leq m \leq d+1, \\ 2^{-d-2} \sum_{i=-1}^{m-d-1} \pi_0(i) \cdot \pi_0(m-d-2-i), & d+2 \leq m \leq M, \end{cases}$$

$$\pi_a(m) = 2^{-1} \cdot \pi_{a-1}(m-1) + 2^{-d-2} \cdot \sum_{i=-1}^{m-2d-a} \pi_0(i) \cdot \pi_{a-1}(m-d-2-i), \quad a = 2..K-1,$$

$$\pi_K(m) = 1 - \sum_{a=0}^{K-1} \pi_a(m),$$

Gdzie bierze się wartości dla $m = M$.¹¹⁴

Dla $M > 1000$ każdorazowe wyznaczanie wektora wartości $\pi_i(M)$ okazuje się na tyle czasochłonne, że w przygotowanej implementacji zostały one stablicowane dla wybranych kombinacji M i d .

5.16. Test wypełniania drzewa binarnego na podstawie ciągu bitów

Podstawą testu jest procedura błędzenia po drzewie binarnym o wysokości h . W każdej iteracji drzewo jest opróżniane, a następnie kolejne bity ciągu interpretowane są jako wskaźnik zagłębiania w lewe albo prawe poddrzewo, aż do momentu znalezienia jeszcze nie odwiedzonego węzła lub liścia, po czym następuje powrót do korzenia. Wyznaczane są dwie statystyki testowe:

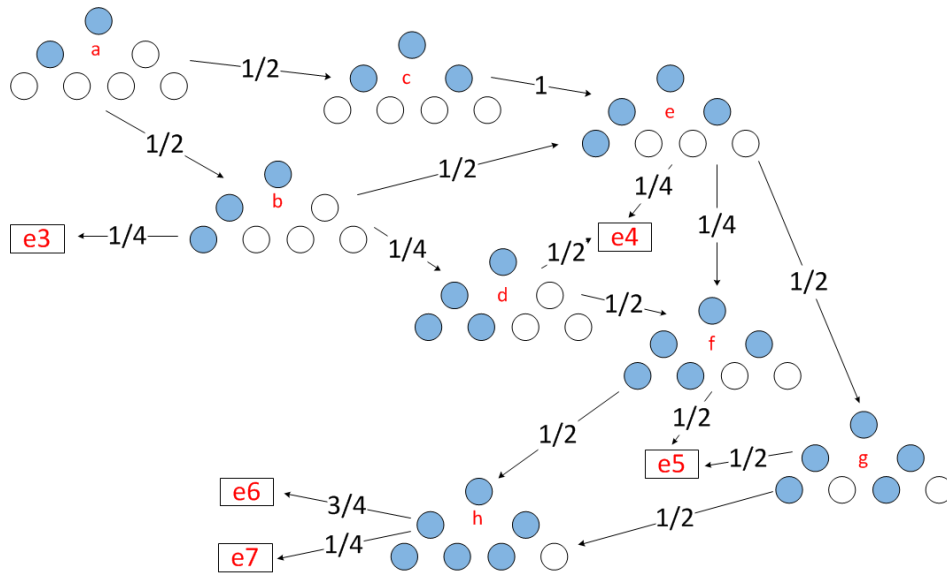
- liczba różnych węzłów lub liści, jakie udało się odwiedzić – zaobserwowany rozkład porównywany jest przy użyciu testu Pearsona z rozkładem teoretycznym,
- numer liścia, dla którego wystąpiła kolizja – zaobserwowany rozkład porównywany jest przy użyciu testu Pearsona z rozkładem równomiernym.

¹¹⁴ K. Hamano, T. Kaneko, *Correction of Overlapping Template Matching Test Included in NIST Randomness Test Suite*, IEICE Transactions of Electronics, Communications and Computer Sciences 2007, E90-A(9), s. 1789

Poniżej prezentujemy, nieco dokładniej niż w [Ma25]¹¹⁵, metodę wyznaczenia teoretycznego rozkładu liczby węzłów (na potrzeby analizy liście również traktować będziemy jako węzły, chyba że będzie wyraźnie zaznaczone inaczej) wypełnionych przed wystąpieniem kolizji – oryginalnie rozkład ten został wyznaczony empirycznie.

Punktem wyjściowym naszych rozważań była analiza przeprowadzona dla drzewa o wysokości 3. W celu redukcji liczby rozważanych stanów założyliśmy, że w każdym kroku żadne poddrzewo nie może być przeważone w prawo – tzn. lewa gałąź nie może mieć mniej wypełnionych węzłów. Jeśli miałyby tak się stać, to gałęzie zamieniane są miejscami.

Na rysunku II.1. przedstawiliśmy graf przejść dla wyróżnionych przez nas stanów wypełnienia drzewa. Krawędzie grafu opisane są prawdopodobieństwami przejścia pomiędzy stanami, zaś dodatkowe stany e3, e4, e5, e6 i e7 odpowiadają wystąpieniu kolizji przy wypełnieniu, odpowiednio, 3, 4, 5, 6 i 7 węzłów drzewa.



rys. II.1. Graf przejść stanów drzewa o wysokości 3

Bazując na powyższym grafie utworzyliśmy macierz prawdopodobieństw przejść T . Na krawędziach opisaliśmy ją stanami zgodnie z rysunkiem II.1.

$$T = \begin{array}{c|cccccccccccccc} & a & b & c & d & e & f & g & h & e3 & e4 & e5 & e6 & e7 \\ \hline a & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 \\ c & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ d & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ e & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 0 & 1/4 & 0 & 0 & 0 \\ f & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ g & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3/4 & 1/4 \\ e3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ e4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ e5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ e6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ e7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}.$$

¹¹⁵ K. Mańk, *Upgrading Frequency Test for Overlapping Vectors and Fill Tree Tests*, IJET, 2/2025

Macierz graniczna, równa piątej potędze macierzy T , zaprezentowana jest poniżej:

$$T^5 =$$

	a	b	c	d	e	f	g	h	e3	e4	e5	e6	e7
a	0	0	0	0	0	0	0	0	1/8	1/4	5/16	15/64	5/64
b	0	0	0	0	0	0	0	0	1/4	1/4	1/4	3/16	1/16
c	0	0	0	0	0	0	0	0	0	1/4	3/8	9/32	3/32
d	0	0	0	0	0	0	0	0	0	1/2	1/4	3/16	1/16
e	0	0	0	0	0	0	0	0	0	1/4	3/8	9/32	3/32
f	0	0	0	0	0	0	0	0	0	0	1/2	3/8	1/8
g	0	0	0	0	0	0	0	0	0	0	1/2	3/8	1/8
h	0	0	0	0	0	0	0	0	0	0	0	3/4	1/4
e3	0	0	0	0	0	0	0	0	1	0	0	0	0
e4	0	0	0	0	0	0	0	0	0	1	0	0	0
e5	0	0	0	0	0	0	0	0	0	0	1	0	0
e6	0	0	0	0	0	0	0	0	0	0	0	1	0
e7	0	0	0	0	0	0	0	0	0	0	0	0	1

Z tej macierzy interesują nas pogrubione prawdopodobieństwa w pierwszym wierszu, będące poszukiwanymi przez nas wartościami.

Na tej samej zasadzie uzyskaliśmy wynik dla drzewa o wysokości 4. Na rysunku II.2 prezentujemy roboczą postać grafu stanów, by pokazać złożoność tak prowadzonej analizy. Nie prezentujemy już samej macierzy, gdyż ma ona stopień 76.



rys. II.2. Graf przejść stanów drzewa o wysokości 4

Obliczając 16 potęgę wspomnianej macierzy, równą macierzy granicznej, uzyskujemy wektor prawdopodobieństw:

$$\frac{1}{2^6}, \frac{3}{2^6}, \frac{45}{2^9}, \frac{535}{2^{12}}, \frac{1335}{2^{13}}, \frac{355}{2^{11}}, \frac{5115}{2^{15}}, \frac{30525}{2^{18}}, \frac{9075}{2^{17}}, \frac{32175}{2^{20}}, \frac{75075}{2^{23}}, \frac{10725}{2^{23}}.$$

Metoda ta staje się nieefektywna dla większych wysokości drzewa, już dla wysokości 6 macierz miałaby stopień 2 598 119, a dla wysokości 11 – niemalże 2^{651} . Potrzebowaliśmy więc podejścia o mniejszych wymaganiach. Pokażemy je, wychodząc od wyników uzyskanych dla drzewa o wysokości 3 i wykorzystując je w obliczeniach dla drzewa o wysokości 4, którego poddrzewa mają wysokość 3.

Wyznaczone wcześniej prawdopodobieństwa dla wysokości 3 możemy wykorzystać do obliczenia prawdopodobieństw, że w drzewie o wysokości 3 i danej wadze (liczbie wypełnionych węzłów) nastąpi kolizja oraz dopełniających je prawdopodobieństw kontynuacji.

Tabela II.3. Prawdopodobieństwa dla drzewa wysokości 3

liczba elementów	0	1	2	3	4	5	6	7
p-two zakończenia	0	0	0	1/8	1/4	5/16	15/64	5/64
p-two konfliktu	0	0	0	1/8	2/7	1/2	3/4	1
p-two kontynuacji	1	1	1	7/8	5/7	1/2	1/4	0

Pogrubiana wartość otrzymana została jako iloraz:

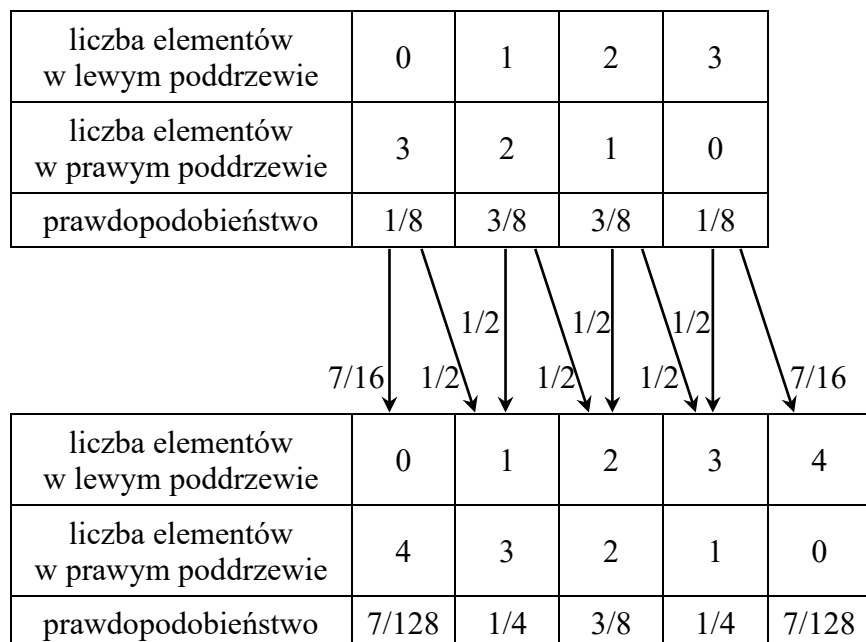
$$\frac{5/16}{5/16 + 15/64 + 5/64}.$$

Obliczenia dla drzewa wysokości 4 rozpoczynamy od wypełnienia go 3 elementami – aż do tego momentu nie istnieje możliwość zajścia kolizji. Możliwe są 4 sposoby rozłożenia tych 3 elementów w poddrzewach. Ich prawdopodobieństwa można wyznaczyć wykorzystując rozkład dwumianowy z liczbą prób 3 i prawdopodobieństwem sukcesu 1/2.

Tabela II.4. Prawdopodobieństwa dla drzewa wysokości 4 o wadze 3

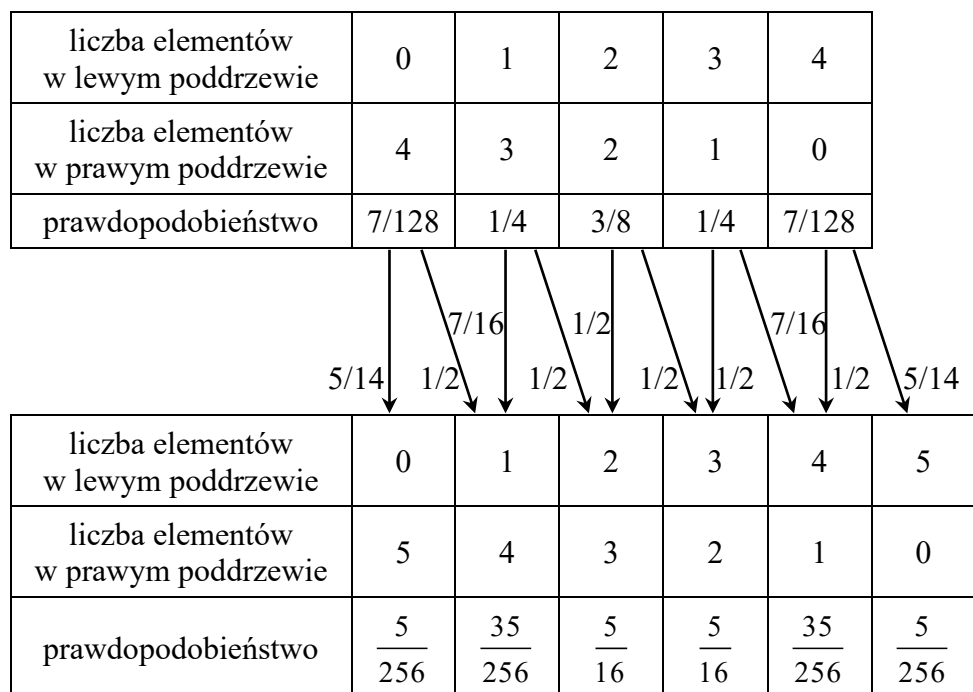
liczba elementów w lewym poddrzewie	0	1	2	3
liczba elementów w prawym poddrzewie	3	2	1	0
prawdopodobieństwo	1/8	3/8	3/8	1/8

Następnie, łącząc dwie powyższe tabele, budujemy graf przejść do drzewa o wadze 4.



rys. II.3. Graf przejść stanów drzewa o wysokości 4 o wadze 3

Łatwo zauważyć, że w ostatnim wierszu brakuje $1/64$, to jest prawdopodobieństwo, że dla drzewa o wysokości 4 konflikt wystąpi po umieszczeniu w nim 4 elementów (3, które zostały umieszczone w poddrzewach, plus korzeń).

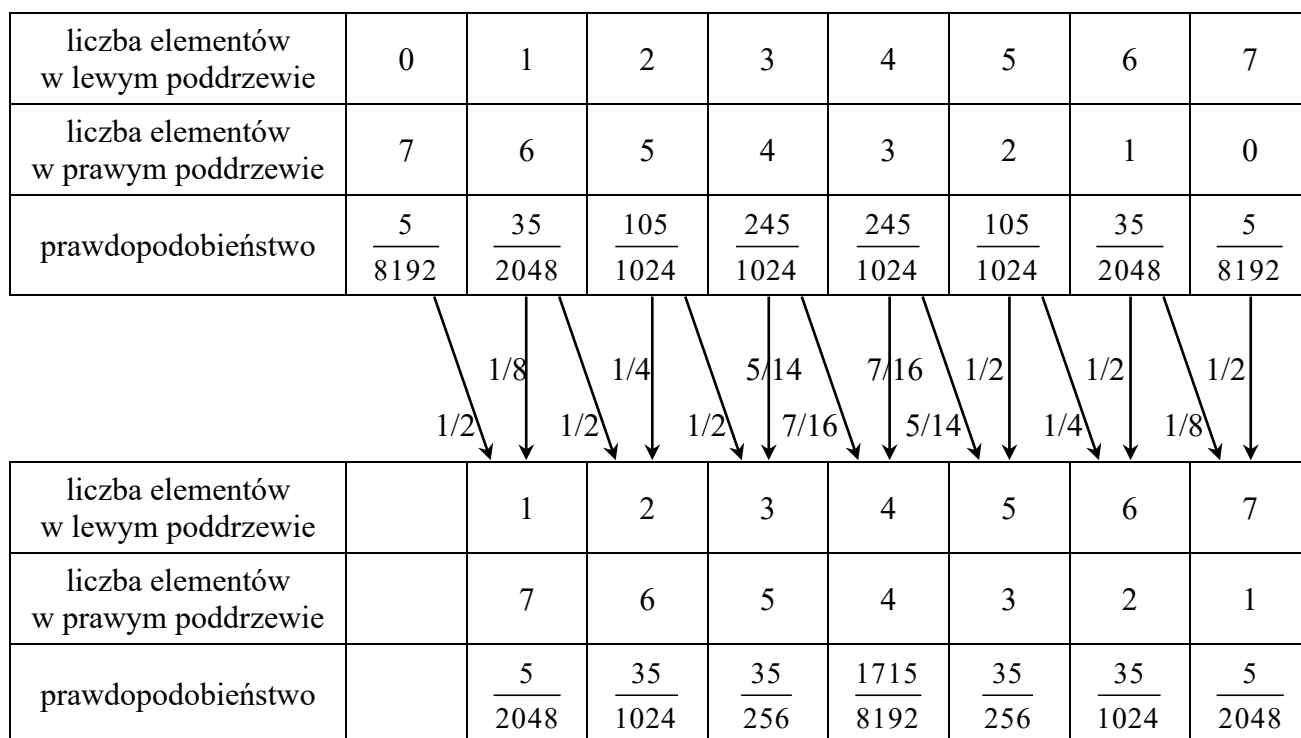


rys. II.4. Graf przejść stanów drzewa o wysokości 4 o wadze 4

Tym razem w ostatnim wierszu brakuje $1/16$, na co składają się prawdopodobieństwa, że dla drzewa o wysokości 4 konflikt wystąpi po umieszczeniu w nim 4 oraz 5 elementów. Ponieważ pierwsze z nich wynosi $1/64$, tak więc dla drugiego dostajemy $3/64$.

Dwa kolejne kroki przebiegają analogicznie. Otrzymujemy dwa kolejne prawdopodobieństwa $45/512$ oraz $535/4096$ i wektor prawdopodobieństw widoczny w górnej tabeli rysunku II.5.

Kolejne kroki przebiegają już nieco inaczej, co wynika z wysycania pojemności poszczególnych poddrzew.



rys. II.5. Graf przejść stanów drzewa o wysokości 4 o wadze 7

Prawdopodobieństwo kolizji dla drzewa o wadze 7 wynosi więc $1335/8192$. W kolejnych sześciu krokach wyznaczamy pozostałe prawdopodobieństwa dla drzewa o wysokości 4.

Na podstawie prawdopodobieństw dla drzewa wysokości 4, możemy wyznaczyć te dla wysokości 5, itd. Tą drogą, niewielkim nakładem obliczeniowym możemy otrzymać dokładne, teoretyczne, stałe dla pierwszej części testu.

Rozdział III

Wybrane pakiety testów losowości

W rozdziale tym przedstawiamy pokrótce wybrane pakiety testów.

1. Knuth II.3.3.2

Tytuł tego podrozdziału odnosi się do drugiego tomu dzieła Donalda E. Knutha *The Art of Computer Programming*, wydanego w 1969 r. Rozdział 3.3.2 uważamy za pierwsze poważne opracowanie z tej dziedziny.

W drugim wydaniu tego tomu z roku 1981¹¹² znaleźć można opisy, częstokroć wzbogacone dogłębną analizą, kilkunastu testów.

1.1. Test równomierności (*Equidistribution Test*)

Test bada, czy elementy ciągu liczbowego U^n są równomiernie rozłożone na przedziale $[0;1)$. Test można wykonać na dwa sposoby. W pierwszym wprost do ciągu U^n stosujemy test Kolmogorowa-Smirnowa (II.1.2), przyjmując $F(x) = x$ jako dystrybuantę teoretyczną. Drugi sposób polega na wykorzystaniu testu zgodności chi-kwadrat (II.1.1) do zbadania równomierności rozkładu ciągu liczb naturalnych postaci $\lfloor k \cdot u_i \rfloor$, gdzie $i = 1..n$ i $k \in \mathbb{N}$ jest wybraną przez nas liczbą – w przypadku współczesnych komputerów binarnych praktycznym wyborem dla k są potęgi 2.

1.2. Test częstości par (*Serial Test*)

Pod tą nazwą pojawiają się dwa testy, będące szczególnymi przypadkami testów częstości dla wektorów nakładających się i nienakładających się, omówionych w punktach II.3.1. oraz II.4.1.

1.3. Test odstępów (*Gap Test*)

Test opisano w punkcie II.2.11.

1.4. Test podziałów (*Partition Test*)

Jest to wariant testu częstości dla wektorów nienakładających się. Rozważany jest ciąg piątek powstały na podstawie badanego ciągu liczbowego

$$\{\lfloor d \cdot u_{5i-4} \rfloor, \lfloor d \cdot u_{5i-3} \rfloor, \lfloor d \cdot u_{5i-2} \rfloor, \lfloor d \cdot u_{5i-1} \rfloor, \lfloor d \cdot u_{5i} \rfloor\}, i = 1.. \lfloor n/5 \rfloor.$$

Knuth zaproponował odmianę testu pokerowego, polegającą na sprawdzaniu jedynie liczby różnych wartości w każdej piątce, nie zaś identyfikowaniu konkretnego „układu”. Do tych zliczeń stosuje się następnie test zgodności chi-kwadrat, w którym prawdopodobieństwo wystąpienia

¹¹² D.E. Knuth, *The Art of Computer Programming*, Addison-Wesley, 1981

w k -tce r różnych liczb określone jest następująco:

$$p_r = \frac{d(d-1)\dots(d-r+1)}{d^k} \left\{ \begin{matrix} k \\ r \end{matrix} \right\},$$

gdzie $\left\{ \begin{matrix} k \\ r \end{matrix} \right\}$ jest liczbą Stirlinga II rodzaju.

1.5. *Test kolekcjonera (Coupon Collector's Test)*

Test opisano w punkcie II.2.21.

1.6. *Test permutacji (Permutation Test)*

Test opisano w punkcie II.3.6.

1.7. *Test serii rosnących / malejących (Run Test)*

1.7.1. Test z macierzą kowariancji

Test opisano w punkcie II.2.22.1.

1.7.2. Test z pomijaniem

Test opisano w punkcie II.2.22.2.

1.8. *Test największy-z-t (Maximum-of-t Test)*

Test opisano w punkcie II.3.5.

1.9. *Test kolizji (Collision Test)*

Test opisano w punkcie II.2.23.

1.10. *Test autokorelacji (Serial Correlation Test)*

Test opisano w punkcie II.2.4.

2. DIEHARD

Zestaw testów nazwanych DIEHARD został zaprezentowany w 1995 roku przez Georgia Marsaglię jako wynik dwóch grantów realizowanych na Florida State University. Został on zamieszczony na CD-ROMie wraz z drugim elementem wykonanej pracy – zestawem 60 zbiorów, każdy o objętości 10^7 bajtów, wygenerowanych przy użyciu rozbudowanego zestawu generatorów pseudolosowych, których opisy i implementacje również znalazły się w tej publikacji.

Znaczącą nowością w przypadku tego zestawu testów był, w przypadku znacznej ich części, empiryczny charakter stosowanych statystyk, dla których nie udało się analitycznie określić rozkładów, a jedynie wyznaczyć eksperymentalnie przybliżone wartości podstawowych parametrów. Aż do powstania zbioru NIST 800-22, DIEHARD był podstawowym zestawem testów do badania generatorów ciągów losowych i pseudolosowych.

Omawiając w rozdziale drugim testy, które pochodzą z pakietu DIEHARD, staraliśmy się przedstawiać je możliwe ogólnie, choć sam pakiet zawiera zazwyczaj implementacje poszczególnych procedur testowych.

Kolejność testów została przyjęta za zbiorem TESTS, zaś poniższe opisy testów powstały również na podstawie [M84]¹¹³, [MZ93]¹¹⁴, [M95]¹¹⁵ i analizy zbiorów źródłowych. Posiłowaliśmy się także [B06]¹¹⁶.

2.1. Test odstępów dni urodzin (*Birthday Spacings Test*)

Test opisano w punkcie II.2.14. Implementacja wykonana została dla $M = 2^{24}$ i $n = 2^9$, co daje $\lambda = 2$. Powyższa sekwencja powtarzana jest 500 razy – dając 500 wartości J , dla których badana jest zgodność z rozkładem teoretycznym przy użyciu testu zgodności chi-kwadrat.

Ze względów technicznych (rok 1995) korzysta się de facto ze słów 32-bitowych, a nie 24, na co wskazuje M , przy czym wykonuje się 9 przebiegów, w pierwszym biorąc z każdego słowa bity 0-23, w drugim 1-24, i tak dalej aż do 8-31, tak więc otrzymuje się 9 wartości p_{val} . Powoduje to, że wymagany jest ciąg o długości $500 \cdot 512 \cdot 32 = 8192000 \text{ b} = 1000 \text{ kB}$.

2.2. Test nakładających się permutacji piątek (*Overlapping 5-Permutation Test*)

Jest to procedura testu nakładających się permutacji, opisanego w punkcie II.4.2.

W implementacji przyjęto, iż badany jest ciąg długości $n = 10^6$ liczb w reprezentacji binarnej o długości $m = 32$ bity. Permutacje są pięcioelementowe, tzn. $M = 5$, co daje 120 możliwych kombinacji. W kodzie programu przyjęto, że wszystkie wartości oczekiwane $\mu_i = n/120$, macierz $[c_{i,j}^{-1}]$, w postaci dwóch części, każda stopnia 60, została wyznaczona empirycznie i zapisana w dodatkowym zbiorze. Błędnie określono liczbę stopni swobody dla statystyki testowej, przyjmując 99 (patrz II.4.2).¹¹⁷

2.3. Test rzędu macierzy binarnej (*Binary Rank Test*)

Test opisano w punkcie II.5.13. W zbiorze DIEHARD zaimplementowano trzy procedury testowe:

- $M = Q = 31$, $N = 40000$, zliczenia dokonuje się dla rzędów: 31, 30, 29 i ≤ 28 , z każdego 32-bitowego słowa 31 najbardziej znaczących bitów tworzy wiersz macierzy,

¹¹³ G. Marsaglia, *A Current View of Random Number Generators*, The Proceedings of the Computer Science and Statistics: 16th Symposium on the Interface, Atlanta, Elsevier Press, 1984

¹¹⁴ G. Marsaglia, A. Zaman, *Monkey Tests for Random Number Generators*, Computers and Mathematics with Applications, 9, 1993

¹¹⁵ G. Marsaglia, *DIEHARD Battery of Tests of Randomness*, 1995, <http://stat.fsu.edu/pub/diehard/>

¹¹⁶ R.G. Brown, *DIEHARDER: A GNU Public Licensed Random Number Tester*, Duke University Physics Department, 2006, <http://www.phy.duke.edu/~rgb>

¹¹⁷ A. Alhakim, J. Kawczak, S. Molchanov, *On the class of nilpotent Markov chains, I: The spectrum of covariance operator*, Markov Processes and Related Fields 10, 2004, s. 629-652

- $M = Q = 32$, $N = 40000$, zliczenia dokonuje się dla rzędów: 32, 31, 30 i ≤ 29 , każde 32-bitowe słowo tworzy wiersz macierzy,
- $M = 6$, $Q = 8$, $N = 100000$, zliczenia dokonuje się dla rzędów: 6, 5 i ≤ 4 , z każdego 32-bitowego słowa wybierany jest określony bajt, który tworzy wiersz macierzy.

2.4. Testy wolnych cel

Test wolnych cel został opisany w punkcie II.4.3. W pakiecie DIEHARD znalazły się cztery procedury tego testu, są to Bitstream, OPSO (*ang. Overlapping-Pairs-Sparse-Occupancy*), OQSO (*ang. Overlapping-Quadruples-Sparse-Occupancy*) i DNA. Oryginalnie udało się wyznaczyć analitycznie jedynie wartości oczekiwane w testach OPSO i OQSO oraz wariancję w pierwszym z nich. Nie ustrzeżono się przy tym pewnych błędów, które wskazaliśmy poniżej.

2.4.1. Test nakładających się słów 20-bitowych (Bitstream Test)

Na podstawie badanego ciągu tworzy się ciąg $N = 2^{21}$ słów o długości 20 bitów, przy czym kolejne słowa przesunięte są, w pierwotnym ciągu, o 1 bit – nakładają się na 19 bitach.

Autorzy przyjęli, że liczba niezaobserwowanych słów ma, w przybliżeniu, rozkład normalny o wartości oczekiwanej 141909 i odchyleniu standardowym 428, w rzeczywistości wynoszą one, odpowiednio, 141909,11 i 427,38.

2.4.2. Test nakładających się par (Overlapping-Pairs-Sparse-Occupancy Test)

Na podstawie ciągu $2^{21} + 1$ liczb o długości 32 bitów tworzony jest ciąg $N = 2^{21}$ słów 20-bitowych, w taki sposób, że po 10 bitów wybieranych jest z każdych dwóch kolejnych liczb w badanym ciągu. Z danej liczby wybierane są tylko kolejno występujące bity, więc w zależności od ich lokalizacji w liczbie, dostajemy 23 warianty testu. Przyjęto, że statystyka testowa ma rozkład normalny o wartości oczekiwanej 141 909 i odchyleniu standardowym 290.

W pracy Marsaglii i Zamana¹¹⁸ znaleźć można wyprowadzenie formuły określającej wartość oczekiwaną oraz bardzo ogólne stwierdzenie co do sposobu wyznaczenia wariancji. Podano jedynie jawną postać przybliżonego wzoru dla prawdopodobieństwa niewystąpienia w ciągu słowa postaci AB:

$$c_1 r_1^n + c_2 r_2^n,$$

gdzie r_1 i r_2 są pierwiastkami równania $r^2 - r + p^2 = 0$, zaś $c_i = 1 / (r_i - 2(p r_i)^2)$.

Ponieważ r_2 jest bardzo małe, więc drugi człon przestaje mieć znaczenie nawet dla niewielkich n , toteż ostatecznie przyjęta formuła go nie uwzględnia. Parametry r_1 oraz c_1 zostały obliczone, przy czym drugi z nich z o wiele mniejszą dokładnością, niż to sugerują autorzy – powinien on wynosić 1,000002861028406195, co pociąga za sobą zmianę podanej wartości prawdopodobień-

¹¹⁸ G. Marsaglia, A. Zaman, *Monkey Tests for Random Number Generators*, Computers and Mathematics with Applications, 9, 1993, s. 4

stwa na 0,135335283235977 i dalej wartości oczekiwanej liczby słów, które nie wystąpiły na 141909,60. Choć podana jest ta sama wartość, to nie odpowiada ona wynikom pośrednim.

Różnice te nie są znaczące, ale dziwi przy tym wszystkim przyjęcie w kodzie wartości 141909, w miejsce poprawnego zaokrąglenia 141 910. Ponadto trzeba zauważyć, że samo zaokrąglenie wartości oczekiwanej i odchylenia standardowego jest niepoprawne.

Posługując się Maple'm można obecnie bardzo łatwo wyznaczyć te parametry. Uzyskane przez nas wyniki pokazały, że przyjęty w oryginalnej pracy¹¹⁹ model daje dokładność wyznaczenia prawdopodobieństw rzędu 10^{-5} . Uzyskane przez nas wartości to 0,13533489616515872 dla słów postaci AB oraz 0,135599222806408 dla słów postaci AA, a wyznaczona na ich podstawie wartość oczekiwana to 141909,19475. Dla odchylenia standardowego otrzymaliśmy wynik 290,39767.

2.4.3. Test nakładających się czwórek (Overlapping–Quadruples–Sparse–Occupancy Test)

Test polega na wyznaczeniu liczby 20-bitowych słów w ciągu długości $N = 2^{21}$ słów, przy czym słowo tworzone jest poprzez wybór określonych 5 bitów z każdej czwórki kolejnych 32-bitowych liczb, tworzących badany ciąg. Przyjęto, że liczba niezaobserwowanych słów ma rozkład normalny o wartości oczekiwanej 141 909 i odchyleniu standardowym 295, przy czym ta ostatnia wartość jest jedynie empirycznym przybliżeniem. Wyznaczona przez nas wartość to 294,64395.

2.4.4. Test DNA (DNA Test)

W Kolejne słowa tworzone są poprzez wybór 2 bitów z każdych kolejnych 10 liczb o długości 32 bitów. Dla liczby niezaobserwowanych słów przyjęto, że ma rozkład normalny o wartości oczekiwanej 141 909, a odchylenie standardowe, na podstawie eksperymentów, przyjęto równe 339. Podana wartość oczekiwana jest inna niż oryginalnie wyznaczona – 141910,5378411.¹²⁰ Otrzymane przez nas wartości to 141909,1782392 i 338,6180217.

2.5. Test „policz jedyńki w ciągu bajtów” (Count–the–1’s Test on a Stream of Bytes)

W pakiecie DIEHARD zaimplementowano wariant testu opisanego w punkcie II.4.4., w którym ciąg bajtów przekształcany jest najpierw w ciąg znaków z pięciznakowego alfabetu:

$$z_i = \begin{cases} A, & Hwt(x_i) \in \{0, 1, 2\}, \\ B, & Hwt(x_i) = 3, \\ C, & Hwt(x_i) = 4, \\ D, & Hwt(x_i) = 5, \\ E, & Hwt(x_i) \in \{6, 7, 8\}, \end{cases}$$

gdzie x_i jest przekształcanym bajtem. Na podstawie ciągu znaków z_i tworzony jest ciąg 256 000 nakładających się pięciznakowych słów oraz 256 001 nakładających się słów czteroznakowych.

¹¹⁹ G. Marsaglia, A. Zaman, *Monkey Tests for Random Number Generators*, Computers and Mathematics with Applications, 9, 1993

¹²⁰ G. Marsaglia, A. Zaman, *Monkey Tests for Random Number Generators*, Computers and Mathematics with Applications, 9, 1993, s. 7

Niech v_i^5 , $i = 1..5^5$ będą zaobserwowanymi częstościami wszystkich możliwych słów pięciorazowych, zaś v_i^4 , $i = 1..5^4$ – częstościami dla słów czterorazowych. Statystyka testowa:

$$S_{obs} = \sum_{i=1}^{3125} \frac{(v_i^5 - E_i^5)^2}{E_i^5} - \sum_{i=1}^{625} \frac{(v_i^4 - E_i^4)^2}{E_i^4},$$

gdzie E_i^d są wartościami oczekiwanymi odpowiadającymi zmiennym losowym v_i^d , ma rozkład chi-kwadrat z $5^5 - 5^4 = 2500$ stopniami swobody.

2.6. Test „policz jedyńki w bajtach” (Count-the-1’s Test for Specific Bytes)

Test przebiega tak jak poprzedni, z wyjątkiem tego, że kolejne bajty są „wyjmowane” po jednym z każdej 32-bitowej liczby tworzącej badany ciąg.

2.7. Test parkingowy (Parking Lot Test)

Test opisano w punkcie II.3.7.

Opis dołączony do pakietu DIEHARD sugeruje, że parkowane są spodki o promieniu 1 oraz nie wspomina o tym, że mogą one „wystawać” poza krawędź parkingu, co istotnie zmienia uzyskiwane wartości.

W pakiecie DIEHARD eksperyment powtarzany jest dziesięciokrotnie, a następnie badany jest rozkład uzyskanych wartości przy użyciu testu Kołmogorowa-Smirnowa.

2.8. Test minimalnej odległości (Minimum Distance Test)

Jest to procedura testu m najbliższych par opisanego w punkcie II.3.2. Przyjęto tu: $n = 8000$, $t = 2$, $p = 2$, $m = 1$ i $N = 100$, przy czym jako test drugiego poziomu wykorzystywany jest test Kołmogorowa-Smirnowa.

Autorzy przyjęli, że kwadrat najmniejszej, w metryce euklidesowej, z odległości pomiędzy 8000 punktów ma, w przybliżeniu, rozkład wykładniczy o średniej 0,995, stąd uzyskiwane, według dystrybucji $1 - e^{-\pi n(n-1)S_{obs}^2/2}$, prawdopodobieństwa powinny mieć, dla losowego źródła, rozkład równomierny na przedziale $[0;1)$.

Zestaw wartości prawdopodobieństw uzyskanych po 100-krotnym powtórzeniu powyższej procedury, badany jest przy użyciu testu Kołmogorowa-Smirnowa.

W pracy [Ma22]¹²¹ pokazujemy metodę znaczącego przyspieszenia testu.

2.9. Test kul (3D Spheres Test)

Jest to kolejna procedura testu m najbliższych par opisanego w punkcie II.3.2. Przyjęto $n = 4000$, $t = 3$, $p = 2$, $m = 1$, $N = 20$ i test Kołmogorowa-Smirnowa w roli testu drugiego poziomu.

¹²¹ K. Mańk, *Speeding Up Minimum Distance Randomness Tests*, Journal of Telecommunications and Information Technology, 2022, vol. 2, s. 103-109

W sześciacie o boku 1000 rozmieszczane jest losowo 4000 punktów, następnie wyznaczana jest najmniejsza odległość pomiędzy dowolnymi dwoma punktami, sześciat której ma w przybliżeniu rozkład wykładniczy o średniej 30.

Zestaw 20 prawdopodobieństw $1 - e^{-2\pi n(n-1)S_{obs}^2/3}$, uzyskanych po 20-krotnym powtórzeniu procedury testowej, badany jest przy użyciu testu Kołmogorowa-Smirnowa.

W pracy [Ma22]¹²² pokazujemy metodę znaczącego przyspieszenia testu.



2.10. Test ściskania (*Squeeze Test*)

Procedura testowa polega na ustaleniu liczby kroków N , które należy wykonać, aby wykonując $k_i = \lceil u_i \cdot k_{i-1} \rceil$, dla $i = 1..N$ i $k_0 = 2^{31} - 1 = 2147483647$, otrzymać $k_N = 1$, przy czym N jest najmniejszym indeksem, dla którego to zachodzi.

Powyższe powtarzane jest 100 000 razy, a rozkład uzyskanych wartości porównywany jest przy użyciu testu chi-kwadrat z rozkładem zapisanym w kodzie programu, który przypuszczalnie został określony eksperymentalnie.

W kontekście niniejszej pracy jest to procedura testu opisanego w punkcie II.2.24.

2.11. Test nakładających się sum (*Overlapping Sums Test*)

Na podstawie badanego ciągu liczb $U^n = u_1, u_2, \dots, u_n$ tworzymy ciąg nakładających się sum częściowych postaci:

$$s_i = \sum_{k=0}^{N-1} u_{k+i}, \text{ dla } i = 1..(n - N + 1),$$

gdzie N jest ustaloną długością sumowanego fragmentu ciągu.

Dla dostatecznie dużych N (w implementacji przyjęto $N = 100$) zmienna losowa opisująca każdą z sum ma, w przybliżeniu, rozkład normalny o średniej $N/2$ i wariancji $N/12$, lecz sumy mające wspólne składniki są, w oczywisty sposób, skorelowane, z kowariancją równą

$$\text{Cov}(s_i, s_j) = (N - r)/12,$$

gdzie $r = \min\{|i - j|, N\}$.

Aby wyeliminować skorelowanie sum autorzy posłużyli się techniką mnożenia przez macierz odwrotną do macierzy kowariancji. W tym celu przyjęli stałą długość ciągu sum, wynoszącą 100 i dla tej konkretnej kombinacji parametrów wyznaczyli postać macierzy kowariancji, a następnie macierzy odwrotnej do górnej macierzy trójkątnej, w rozkładzie Choleskyego-Banachiewicza macierzy kowariancji. Ta macierz odwrotna ma co najwyżej trzy niezerowe elementy w każdej kolumnie – w pierwszym wierszu, na głównej przekątnej i tuż nad nią, co więcej – elementy te wyrażają się prostymi formułami, dzięki czemu nie muszą być tablicowane. Po pomnożeniu wyjściowego

¹²² K. Mańk, *Speeding Up Minimum Distance Randomness Tests*, Journal of Telecommunications and Information Technology, 2022, vol. 2, s. 103-109

wektora sum przez tę macierz otrzymuje się wektor nieskorelowanych zmiennych losowych, dla których można już wykonać test zgodności rozkładu.

Autorzy DIEHARD, zarówno w opisie testu, jak i w komentarzach zawartych w implementacji, twierdzą, że pomnożenie przez wspomnianą powyżej macierz daje w wyniku ciąg niezależnych zmiennych losowych o rozkładzie normalnym. O ile stwierdzenie o niezależności jest prostą nadinterpretacją nieskorelowania tych zmiennych, o tyle teza o rozkładzie tych nowych zmiennych jest już całkowicie błędna. Autorzy musieli zauważyć, że otrzymywane, na podstawie tylko tak zmodyfikowanego wektora, wyniki nie są poprawne, gdyż w kodzie programu, zamiast prostego podstawienia

$$x[i] = \text{phi}(x[i])$$

w którym

$x[i]$ to i -ty element wektora po wymnożeniu przez macierz odwrotną,

$\text{phi}()$ to funkcja obliczająca $\Phi(x)$ – dystrybuantę rozkładu $N(0,1)$,

znaleźć można 101-elementowy wektor f rosnących liczb rzeczywistych z przedziału $[0,1]$ i cztery następujące linie kodu

```
p = phi(x[i])
h = 100.*p
j = h
x[i] = f[j] + (h-j) * (f[j+1] - f[j]) .
```

Całość zinterpretowaliśmy jako próbę „przeskalowania” niepoprawnych wyników, tak by uzyskać w przybliżeniu równomierny rozkład uzyskiwanych prawdopodobieństw.

Tak, jak w kodzie programu przyjmijmy

$$y_i = (s_i - 50)\sqrt{12}, \quad i = 1..100,$$

jako ciąg częściowo ustandaryzowanych sum, zaś x_i będzie ciągiem odpowiadającym y_i po przemnożeniu przez macierz odwrotną.

Na podstawie centralnego twierdzenia granicznego, zmienne losowe opisujące poszczególne y_i mają w przybliżeniu rozkład normalny o wartości oczekiwanej 0 i wariancji N .

Dla x_1 mamy po prostu $x_1 = y_1/\sqrt{N} = y_1/10$, tak więc można przyjąć, że x_1 ma rozkład $N(0,1)$. Następnie mamy

$$x_2 = -y_1 \frac{N-1}{\sqrt{N(2N-1)}} + y_2 \sqrt{\frac{N}{2N-1}},$$

co przekształcamy, wykorzystując wcześniejsze formuły dla y_i i s_i , otrzymując ostatecznie

$$x_2 = (u_{N+1} - u_1) \sqrt{12 \frac{N}{2N-1}} + \frac{y_1}{\sqrt{N(2N-1)}},$$

co dla dużych N można przybliżyć jako

$$x_2 = (u_{101} - u_1)\sqrt{6} + \frac{s_1 - 50}{100}\sqrt{6}.$$

Pierwszy składnik to prosta różnica dwóch zmiennych losowych o rozkładzie równomiernym na przedziale $[0,1)$, ze współczynnikiem $\sqrt{6}$, przez co wariancja takiego układu wynosi 1 (w rzeczywistości nieco mniej). Drugi składnik to z kolei niepoprawnie ustandaryzowana suma s_1 , w wyniku czego wariancja tego składnika to $1/(2N-1)$, zaś całość ma, wobec niewielkiego skorelowania tych składników, wariancję bliską 1, a więc prawidłową. Wartości oczekiwana całości jest również bardzo bliska 0.

Trudne jest określenie rozkładu zmiennej losowej opisującej x_2 . przewagę będzie miał w nim rozkład pierwszego składnika, a to z dwóch powodów:

- zakres przyjmowanych przez niego wartości to $[-\sqrt{6}; \sqrt{6})$, wobec $[-\sqrt{6}/2; \sqrt{6}/2)$ dla drugiego składnika,
- rozproszenie wartości, mierzone odchyleniem standardowym, jest dla niego ponad 14 razy większe.

Pomimo tych trudności, jedno jest pewne: x_2 nie ma rozkładu, nawet w przybliżeniu, normalnego.

Powyższe były bardzo szczególnymi przypadkami. Dla pozostałych x_i mamy:

$$x_i = y_1 \frac{1}{\sqrt{2a(a-1)}} - y_{i-1} \sqrt{\frac{a-1}{2a}} + y_i \sqrt{\frac{a}{2(a-1)}}, \quad i = 3..100,$$

gdzie $a = 2N + 2 - i$. Podobnie, jak dla x_2 możemy powyższe przekształcić do postaci:

$$x_i = \frac{1}{\sqrt{2a(a-1)}}(y_1 + y_{i-1}) + \sqrt{\frac{6a}{(a-1)}}(u_{N+i-1} - u_{i-1}),$$

co można przybliżyć jako:

$$x_i = (u_{99+i} - u_{i-1})\sqrt{6} + (s_1 + s_{i-1} - 100) \frac{\sqrt{6}}{200-i}.$$

Nadal dominującym będzie pierwszy składnik o rozkładzie takim jak wcześniej. Drugi składnik jest bardziej kłopotliwy. Dla małych i , można przyjąć, że $s_1 + s_{i-1} \approx 2s_1 \approx 2s_{i-1}$, a wówczas powyższa formuła upraszcza się do postaci identycznej, jak dla x_2 . Z kolei dla i bliskich N obie sumy mają już bardzo mało wspólnych składników, więc ich sumę można interpretować jako sumę $2N$ wyrazów badanego ciągu, toteż w obu skrajnych przypadkach można przyjąć, że drugi składnik ma, jak poprzednio, rozkład normalny o wariancji bliskiej $1/100$ dla małych wartości i , zaś rosnącej aż do $1/50$ dla i bliskich N .

Jak widać zaproponowana formuła nie zapewnia nawet stałości podstawowych parametrów rozkładu x_i , stąd należy przypuszczać, że umieszczony w kodzie programu wektor służący do przeskalowywania prawdopodobieństw powstał na podstawie eksperymentalnych uśrednień.

Chcąc uzyskiwać dokładne wartości należałoby sporządzić 99 takich wektorów, gdyż tylko x_1 ma już prawidłowy rozkład.

Test w zaproponowanej przez autorów postaci wydaje się nieprzydatny, gdyż:

- w większym stopniu bada różnicę wyrazów ciągu odległych o N , niż zachowanie się nakładających się sum,
- na skutek przeskalowywania, uzyskiwane prawdopodobieństwa mają jedynie charakter bardzo zgrubnych przybliżeń, o ile nie orientacyjny,
- bez wykonania przeskalowania ma wartość jeszcze mniejszą, ze względu na wykazaną przez nas niezgodność rozkładów.

2.12. *Test serii (Runs Test)*

Jest to test opisany w punkcie II.2.22.2. Choć dołączony opis sugeruje, że zliczane są wspólnie serie zarówno rosnące, jak i malejące, co dawałoby inny wariant testu serii. W implementacji oba typy serii zliczane są co prawda równoległe, ale oddzielnie i wyznaczane są dwie statystyki testowe.

Testowane są paczki po dziesięć ciągów liczbowych o długości 10000 elementów każdy, w ten sposób otrzymywane są dwa zestawy prawdopodobieństw p_{val} , dla których przeprowadzany jest test zgodności Kołomogorowa-Smirnowa.

2.13. *Test oparty na grze w kości (Craps Test)*

W pakiecie zaimplementowany został test dla kości sześciościennych, w którym wykonywane jest $N = 200000$ gier w kości, wg zasad Craps. Opis testu znaleźć można w punkcie II.2.25.

3. FIPS PUB 140

Odnosimy się tu do dwóch wydań FIPS PUB 140-1¹²³ i 140-2¹²⁴ amerykańskiej normy określającej wymagania dla modułów kryptograficznych. Dla generatorów, zarówno losowych, jak i pseudolosowych, nakazuje ona przeprowadzanie każdorazowego testu rozruchowego. Wcześniejsze wersje normy, w ramach tego testu, przewidywały wykonanie kilku podstawowych testów losowości na próbce o objętości 20000 bitów, a więc zostały określone bardzo szczegółowe procedury testowe.

¹²³ Federal Information Processing Standard Publication 140-1, FIPS PUB 140-1, 1994

¹²⁴ Federal Information Processing Standard Publication 140-2, FIPS PUB 140-2, 2001

Pomimo usunięcia z normy w wydaniu 140-2¹²⁵, ze względu na prostotę implementacji, zestaw ten jest wciąż często używany.

Ponieważ każda z procedur zwraca jednobitową wartość, jest to zupełnie inny rodzaj pakietu, dedykowanego do zwartych implementacji (w tym sprzętowych).

3.1. *Test monobitowy*

Tożsamy z testem momentu pierwszego rzędu dla bloku jednobitowego opisanym w punkcie II.2.2.2., przy czym wartości progowe wyznaczone zostały przy użyciu rozkładu dwumianowego.

3.2. *Test pokerowy*

W rzeczywistości jest to test częstości dla nienakładających się bloków czterobitowych – patrz punkt II.2.1.1. Ciekawostką jest wykorzystanie dwustronnego, niesymetrycznego przedziału krytycznego w teście ze statystyką chi-kwadrat.

3.3. *Testy liczby serii*

W pakiecie było to łącznie 12 procedur, polegających na sprawdzeniu czy liczby serii zer oraz jedynek, oddzielnie o długościach 1, 2, 3, 4, 5 oraz 6 i więcej, mieszczą się w wyznaczonych przedziałach. Wartości progów wyznaczono prawdopodobnie empirycznie, gdyż użycie wyników z punktu II.5.3. nie daje zadowalającej dokładności dla tak małej długości ciągu.

3.4. *Test długich serii*

Zaobserwowanie długiej serii (w zależności od wariantu od 26 do 28 bitów) zer albo jedynek uznawane było jako nieprzejsie przez ciąg testu.

4. AIS-31

Jest to fragment niemieckiej normy określającej wymagania dla generatorów losowych, pochodzący z tego samego okresu oraz bazującej na tych samych testach i ideach jak FIPS PUB 140. W pokrywających się przypadkach wartości liczbowe tożsame są z FIPS PUB 140-1¹²⁶.

W publikacji [KS01]¹²⁷ można znaleźć odwołanie się do poniższych testów.

4.1. *Test rozłączności (disjointness test)*

Według opisu, test bada ciąg bloków $B^{d,n} = b_1^d, \dots, b_n^d$ w poszukiwaniu wystąpienia tej samej wartości w dwóch kolejnych blokach.

W [KS01] określono, że $d = 48$ i $n = 2^{16}$, zaś prawdopodobieństwo błędu pierwszego rodzaju wynosi 2^{-17} , co nie może być prawdą. Wartość taką otrzymamy, jeśli w całym ciągu poszukiwać

¹²⁵ Federal Information Processing Standard Publication 140-2, FIPS PUB 140-2, 2001, s. 57-58

¹²⁶ Federal Information Processing Standard Publication 140-1, FIPS PUB 140-1, 1994

¹²⁷ W. Killmann, W. Schindler, *AIS-31 – Functionality classes and evaluation methodology for true (physical) random number generators*, Version 3.1, 2001, s. 33-37

będziemy powtórzenia się wartości każdego z bloków. Jest to o tyle istotne, że oznacza potężną komplikację zwartej implementacji, dla której test ten został przeznaczony.

Dla tej kombinacji parametrów test wymaga 3 145 728 bitów, co nijak nie współgra z pozostałymi testami.

4.2. *Test monobitowy*

Tożsamy z testem monobitowym z FIPS PUB 140-1 (patrz punkty III.3.1. i II.2.2.2.).

4.3. *Test pokerowy*

Tożsamy z testem pokerowym z FIPS PUB 140-1 (patrz punkty III.3.2. i II.2.1.1.).

4.4. *Testy liczby serii*

Tożsamy z testem liczby serii z FIPS PUB 140-1 (patrz punkt III.3.3.).

4.5. *Test długich serii*

Tożsamy z testem długich serii z FIPS PUB 140-1 (patrz punkt III.3.4.).

4.6. *Test autokorelacyjny*

Test opisano w punkcie II.5.1., przy czym:

- przyjęto zakres przesunięć $t = 1..5000$,
- sumy $A_t = \sum_{i=1}^{n-t} s_i \oplus s_{i+t}$ również ograniczone zostały do 5000 początkowych składników,
- testowaniu poddane zostanie jedynie 10 000 bitów, a nie 20 000 jak w pozostałych testach.

4.7. *Test równomierności rozkładu*

Test bada częstości wystąpień poszczególnych wartości d -bitowego bloku w ciągu $B^{d,n}$. Poszczególne częstości oceniane są niezależnie od siebie, poprzez sprawdzenie zachodzenia relacji:

$$\frac{1}{n} \# \{i = 1..n : b_i^d = x\} \in [2^{-d} - a; 2^{-d} + a],$$

dla każdego $x \in \{0;1\}^d$ oraz $2^{-d} - a$ i $2^{-d} + a$ będących odpowiednimi kwantylami rozkładu dwumianowego z liczbą prób n i prawdopodobieństwem sukcesu 2^{-d} .

Autorzy nie odnoszą się w ogóle do skorelowania wyników uzyskiwanych dla poszczególnych wartości bloku. Taki przebieg testu jest tylko minimalnie mniej złożony od klasycznego testu zgodności chi-kwadrat (patrz punkt II.1.1.). Co więcej, w teście pokerowym z tego pakietu liczona jest taka właśnie statystyka (patrz punkt III.4.3.), więc dziwi nas tak ułomne podejście zaproponowane w tym przypadku.

4.8. Test porównawczy dla rozkładów wielowymiarowych

Jest to kolejny wariant testu służącego ocenie równomierności rozkładu częstości dla nienakładających się bloków d -bitowych. Autorzy proponują porównywanie wektorów zliczeń uzyskanych dla różnych ciągów (kolejnych fragmentów badanego ciągu).

Niech h będzie liczbą porównywanych ciągów, każdy złożony z n bloków, zaś v_i^j liczbami obserwacji bloku i w ciągu j . Niech

$$p_i = \frac{1}{hn} \sum_{j=1}^h v_i^j, \text{ dla } i = 0..2^d - 1,$$

będą całkowitymi częstościami występowania bloku i we wszystkich ciągach.

Statystyka testowa

$$S = \sum_{j=1}^h \sum_{i=0}^{2^d-1} \frac{(v_i^j - np_i)^2}{np_i}$$

ma asymptotycznie rozkład chi-kwadrat z $(h-1)(2^d-1)$ stopniami swobody.¹²⁸

4.9. Test entropii

Tożsamy z testem Corona opisanym w punkcie II.2.17.

5. NIST SP 800-22 Statistical Test Suite

Sygnowanie przez NIST i wykorzystanie w trakcie konkursu, którego celem było wyłonienie nowego standardu szyfrowania symetrycznego – AES, spowodowało, że jest to jedna z najpoważniejszych publikacji, dotyczących testowania losowych i pseudolosowych generatorów ciągów binarnych.¹²⁹ Pozycja ta zawiera w pierwszym rozdziale podstawowe informacje z zakresu testowania losowości ciągów binarnych, dwa kolejne to opisy 16 proponowanych testów od strony użytkowej i podbudowy teoretycznej. Rozdział czwarty dotyczy metodologii testowania generatorów, zaś piąty jest podręcznikiem użytkownika oprogramowania stanowiącego implementację omawianych testów.

Zbiór testów zaproponowanych przez NIST stał się jednym z podstawowych narzędzi badawczych, a wyniki uzyskiwane przy jego pomocy są ważnym elementem składowym oceny wszystkich współczesnych algorytmów kryptograficznych. W tym sensie znaczenia tego pakietu nie sposób przecenić.

W 2022 rozpoczęto formalną procedurę w celu wprowadzenia poprawek, jednak do tej pory nie rozpoczęto nawet przygotowywania draftu nowej wersji.

W kolejnych punktach omawiamy testy, które znalazły się w pakiecie.

¹²⁸ G.K. Kanji, *100 Statistical Tests*, 3rd Edition, SAGE Publications, 2006, s. 137

¹²⁹ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001

5.1. Test częstości bitów (*Frequency Monobit Test*)

Test jest tożsamy z testem momentu pierwszego rzędu dla bloku jednobitowego opisanym w punkcie II.2.2.2., przy czym wykorzystuje przybliżenie dystrybucji statystyki testowej przy użyciu tylko pierwszego wyrazu.

5.2. Test częstości w bloku (*Frequency Test within a Block*)

Test jest zgodny z testem wag Hamminga opisanym w punkcie II.5.12.2.

W SP 800-22¹³⁰ zaleca się by długość badanego ciągu wynosiła, co najmniej, 100 bitów, długość pojedynczego bloku – 20 bitów, zaś liczba bloków nie przekraczała 100. Dla przyjętych w pakiecie parametrów ostatni z warunków jest przekroczony 78-krotnie. Z kolei dla domyślnych parametrów test przeprowadzany jest dla 7812 bloków 128-bitowych.

5.3. Test serii (*Runs Test*)

Test został opisany w punkcie II.5.2.

Użycie mniej dokładnych formuł dla wartości oczekiwanej i odchylenia standardowego nie ma praktycznego wpływu na jakość otrzymywanych wyników dla długich ciągów, jednak należy mieć świadomość, że liczba nadmiarowo odrzucanych przez test ciągów spada poniżej 1% dla długości badanego ciągu ok. 370 bitów przy idealnej częstości jedynek i dopiero przy przeszło 24 000 bitów dla skrajnego dopuszczanego przypadku – 0,7. Tak więc, przyjmując ten ostatni, nie powinniśmy w teście badać ciągów krótszych niż 24 000 bitów, jeśli chcemy, by błąd naszej oceny nie przekraczał 1%.

5.4. Test najdłuższej serii jedynek w bloku (*Test for the Longest Run of Ones in a Block*)

Test został opisany w punkcie II.5.11.

Parametry dla przypadku $M = 10000$ zostały błędnie obliczone, poprawione wartości prezentujemy w tabeli III.1.

Tabela III.1. Oryginalne i poprawione wartości parametrów w teście najdłuższej serii jedynek w bloku, dla przypadku $M = 10000$

numer klasy	zakres długości serii	π_i wg [NIST01] ¹³¹	π_i poprawne
0	≤ 10	0,0882	0,08663
1	11	0,2092	0,20820
2	12	0,2483	0,24842
3	13	0,1933	0,19391
4	14	0,1208	0,12146
5	15	0,0675	0,06801
6	≥ 16	0,0727	0,07337

¹³⁰ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, Rev. 1a, 2010, s. 2-5

¹³¹ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001, s. 69

Nieprawidłowe wydają się być również wartości dla M wynoszących 512 oraz 1000 przedstawione w cytowanej pracy¹³², które nie zostały zawarte w dostępnej implementacji testu.

Poważne wątpliwości budzą sugerowane długości badanych ciągów. Wartość $M = 8$ bitów, oznacza podział ciągu na 16 bloków, co daje następujące wartości oczekiwane dla kolejnych klas: 3,4375, 5,8750, 3,6875 oraz 3,0. Dla długości bloku $M = 128$ i minimalnej długości ciągu najmniejszą wartością oczekiwaną jest 5,0324, zaś dla $M = 10000$ dostajemy odpowiednio 5,0625 lub 5,1008 w zależności od tego, który z podanych powyżej zestawów wartości π_i przyjmiemy. Choć w literaturze spotyka się różne zalecane minimalne wartości oczekiwane dla klasy, to nie spotkaliśmy się, jak dotychczas, z wartością mniejszą niż 5, a źródła podają niekiedy nawet 20. Ponieważ, tylko dla najkrótszego bloku, pojawiają się wartości mniejsze od 5, więc uznać należy to za błąd wynikający z swoistej „ładności” wartości 2^7 , podczas gdy powinna pojawić się wartość 216.

Kolejną wątpliwość budzi prowadzenie testu jedynie w odniesieniu do jedynek. W punkcie 2.4.8.¹³³ podany jest przykład, z którego najistotniejsze elementy prezentujemy poniżej.

Tabela III.2. Przebieg testów najdłuższej serii jedynek oraz najdłuższej serii zer w bloku dla przykładu z SP 800-22

blok	najdłuższa seria 1	najdłuższa seria 0
11001100	2	2
01101100	2	2
11100000	3	5
01001101	2	2
00010011	2	3
10000000	1	7
11001100	2	2
11011000	2	3
00010101	1	3
01001100	2	2
00000010	1	6
01010001	1	3
11010110	2	1
11010111	3	1
11100110	3	2
10110010	2	2

W zależności od tego, czy przeprowadzimy procedurę testową ze względu na 1 czy 0 otrzymamy dwie bardzo różne wartości statystyki testowej. W pierwszym przypadku będzie to około 4,8825, zaś w drugim 0,84304, co prawda w obu przypadkach ciąg przechodzi test, lecz różnica jest bardzo znacząca, gdyż odpowiadające wartości p_{val} to odpowiednio 0,18061 i 0,83915.

Wystarczy jednak tylko nieco zmodyfikować badany ciąg, co uczyniliśmy to tak, aby nie zmieniły się wartości otrzymywane dla 1, by zamiast bardzo dobrego wyniku dla 0 otrzymać rezul-

¹³² A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001, s. 68-69

¹³³ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001, s. 23-24

tat fatalny. W tabeli III.3 kolorem czerwonym oznaczyliśmy wszystkie bity zmodyfikowane w stosunku do oryginalnego przykładu.

Tabela III.3. Przebieg testów najdłuższej serii jedynek oraz najdłuższej serii zer w bloku dla przykładu zmodyfikowanego

blok	najdłuższa seria 1	najdłuższa seria 0
11001000	2	3
01101000	2	3
11100010	3	3
10001101	2	3
00010011	2	3
10001000	1	3
11001000	2	3
11011000	2	3
00010101	1	3
01011000	2	3
00010010	1	3
01010001	1	3
11000110	2	3
11000111	3	3
11100010	3	3
10100011	2	3

Tym razem wartość statystyki testowej, odpowiadającej 0, to 53,42373, co oznacza $p_{val} \approx 1,48903 \cdot 10^{-11}$. Z pewnością istnieje pewna korelacja pomiędzy wynikami testów przeprowadzonych w odniesieniu do 1 i 0, tym niemniej stwierdzenie zawarte w punkcie 2.4.1 publikacji, mówiące o implikowaniu nieregularności dla 0 przez zaobserwowane dla 1 jest, jak widać, całkowicie nieuzasadnione.

5.5. Test rzędu macierzy binarnej (Binary Matrix Rank Test)

Test opisano w punkcie II.5.13. Wykonywany jest dla M i Q równych 32.

Statystyka testowa ma postać:

$$S_{obs} = \frac{(F_{32} - 0,2888N)^2}{0,2888N} + \frac{(F_{31} - 0,5776N)^2}{0,5776N} + \frac{(N - F_{32} - F_{31} - 0,1336N)^2}{0,1336N},$$

gdzie F_{32} oraz F_{31} to zaobserwowane liczby macierzy o rzędzie 32 oraz 31. S_{obs} ma asymptotycznie rozkład chi-kwadrat z 2 stopniami swobody.

Minimalna, zalecana liczba rozpatrywanych macierzy to 38.

Do wyznaczenia parametrów liczbowych dla statystyki testowej posłużono się przybliżeniem, w którym zamiast „krótkiego”, jak w [M95]¹³⁴, występuje nieskończony iloczyn. Uzyskane w ten sposób wartości nie odbiegają jednak od poprawnych o więcej niż 10^{-9} .

¹³⁴ G. Marsaglia, *DIEHARD Battery of Tests of Randomness*, <http://stat.fsu.edu/pub/diehard/>, 1995, s. 9

5.6. Test spektralny – dyskretna transformata Fouriera (*Discrete Fourier Transform Test*)

Test spektralny został przedstawiony w punkcie II.5.9.1, przy czym w dostępnych implementacjach pakietu STS nie uwzględniono opublikowanych dla niego poprawek.

5.7. Test nienakładających się wzorców (*Non-overlapping Template Matching Test*)

Test został opisany w punkcie II.5.14. W pakiecie STS przyjęto następujące wartości parametrów:

- liczba powtórzeń: $N = 8$,
- długość podciągu: $M = 131072$,
- długości bloków $d = 9$ i 10 .

5.8. Test nakładających się wzorców (*Overlapping Template Matching Test*)

Test został opisany w punkcie II.5.15. W pakiecie STS przyjęto następujące wartości parametrów $M = 1032$, $d = 9$ lub 10 , $K = 5$ oraz $n \geq 10^6$.

5.9. Uniwersalny test Maurera (*Maurer's Universal Statistical Test*)

Test został opisany w punkcie II.2.16.

W pakiecie STS wykorzystano najstarsze, opublikowane przez Maurera w pierwotnej pracy¹³⁵, heurystyczne przybliżenie współczynnika c , reprezentującego wpływ kowariancji sumowanych zmiennych losowych:

$$c^* = 0,7 - \frac{0,8}{d} + \left(4 + \frac{32}{d}\right) \frac{K^{-3/d}}{15}.$$

Jednocześnie autorzy nieprawidłowo powołują się na publikację [CN98]¹³⁶, w której wyznaczono dokładną postać tego współczynnika, gdyż:

- zignorowali zupełnie fakt podania tamże dokładnych parametrów dla testu,
- odrzucili aproksymację:

$$c^{**} = 0,7 - \frac{0,8}{d} + \left(1,6 + \frac{12,8}{d}\right) K^{-4/d},^{137,138}$$

- uzasadnili to cytatem stwierdzającym, że powyższa aproksymacja jest tak niedokładna, że odsetek ciągów odrzucanych przez test może okazać się 2,67 razy większy od poziomu istotności, nie przywiązując wagi do tego, że wynik ten dotyczy poziomu istotności 0,001, nie zaś 0,01 przyjętego w pakiecie.

¹³⁵ U.M. Maurer, *A Universal Statistical Test for Random Bit Generators*, Journal of Cryptology, 5, 1992, Springer-Verlag, s. 101

¹³⁶ J.S. Coron, D. Naccache, *An Accurate Evaluation of Maurer's Universal Test*, Proceedings of SAC '98 (Lecture Notes in Computer Science), Springer-Verlag, 1998

¹³⁷ J.S. Coron, D. Naccache, *An Accurate Evaluation of Maurer's Universal Test*, Proceedings of SAC '98 (Lecture Notes in Computer Science), Springer-Verlag, 1998, s. 1

¹³⁸ A. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, s. 184

Niestety poważnej wpadki nie ustrzeżono się również w [CN98]. Z tabeli 2.¹³⁹ wybrano wartość z ostatniego wiersza – wspomniane wyżej ok. 2,67, zamiast zwrócić uwagę na wiersz pierwszy, odpowiadający $d = 3$, w którym podano wartość 0,1541921, oznaczającą, że test odrzuca niemal 6,5-krotnie zbyt rzadko. Poniżej powtórzyliśmy te obliczenia, przyjmując jednak model bardziej przystający do naszego przypadku, czyli $K = 1000 \cdot 2^d$ oraz poziom istotności 0,01. Do obliczenia poprawnej wartości c wykorzystaliśmy wyniki przedstawione w [Ma08]¹⁴⁰.

Tabela III.4. Wpływ postaci formuły określającej współczynnik kowariancji w teście Maurera na jakość wyników testu

L	c^*	c^{**}	c	błąd % dla c^*	błąd % dla c^{**}
2	0,30001	0,30000	0,48670	+ 1023,378	+ 1023,436
3	0,43346	0,43337	0,52322	+ 228,504	+ 228,849
4	0,50056	0,50030	0,55223	+ 95,514	+ 96,155
5	0,54137	0,54104	0,57452	+ 52,149	+ 52,787
6	0,56913	0,56900	0,59108	+ 31,318	+ 31,521
7	0,58941	0,58985	0,60296	+ 18,039	+ 17,414
8	0,60500	0,60632	0,61122	+ 7,845	+ 6,129
9	0,61741	0,61988	0,61679	– 0,742	– 3,667
10	0,62755	0,63136	0,62044	– 8,218	– 12,363
11	0,63602	0,64128	0,62278	– 14,766	– 20,067
12	0,64321	0,65000	0,62424	– 20,480	– 26,840
13	0,64940	0,65775	0,62514	– 25,444	– 32,752
14	0,65478	0,66469	0,62568	– 29,741	– 37,890
15	0,65951	0,67097	0,62600	– 33,460	– 42,349
16	0,66369	0,67667	0,62619	– 36,683	– 46,222

Z powyższej tabeli odczytać można dwie informacje. Po pierwsze, z wyszarzonego obszaru, odpowiadającego zestawowi danych przyjętych w pakiecie STS widać, że popełniany błąd wartości p_{val} waha się od niemal –37% do przeszło +31% i tylko dla jednego przypadku jest mniejszy od 1%. Drugie spostrzeżenie dotyczy przedstawionej w [MOV96]¹⁴¹ heurystycznej formuły dla c , która daje widocznie gorsze wyniki od pierwotnej wersji, co więcej, jest to efekt globalny, występujący również dla innych wartości poziomu istotności.

5.10. Test złożoności liniowej (Linear Complexity Test)

Test opisano w punkcie II.5.7.3.

W [NIST01]¹⁴² zaproponowano podział na 7 kategorii: ≤ -3 , -2 , -1 , 0 , 1 , 2 , ≥ 3 , dla których, nie do końca poprawnie, podano prawdopodobieństwa – w rozdziałach 2 i 3 znajdują się różne zestawy tych wartości, niemal poprawny jest ten z rozdziału 3, przy czym powinno być $\pi_0 = 0,010417$, niestety niepoprawny jest również zestaw zaimplementowany w kodzie pakietu.

¹³⁹ J.S. Coron, D. Naccache, dz. cyt., s. 11

¹⁴⁰ K. Mańk, *Wyznaczenie dokładnych wartości parametrów dla testu Maurera*, Biuletyn WAT, LVII, 4, Warszawa, 2008, s. 117-149

¹⁴¹ A. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, s. 184

¹⁴² A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001

5.11. Test częstości nakładających się bloków (*Serial Test*)

Test jest złożeniem dwóch procedur testowych: testów nakładających się wektorów t i $t-1$ elementów jednobitowych ($d=1$), przy czym, w odróżnieniu od testu opisanego w punkcie II.4.1.4., zastosowano wybieg opisany m.in. w [MOV96]¹⁴³ i [Go53]¹⁴⁴.

Badany ciąg rozszerza się, dodając na jego końcu $t-1$ początkowych bitów.

Niech v_{i_t} , $i = 0..2^t - 1$ będą zaobserwowanymi liczbami wystąpień wszystkich możliwych t -bitowych bloków, $v_{i_{t-1}}$, $i = 0..2^{t-1} - 1$ – liczbami wystąpień dla bloków o długości $t-1$ oraz $v_{i_{t-2}}$, $i = 0..2^{t-2} - 1$ – liczbami wystąpień dla bloków o długości $t-2$ bity. Najpierw obliczamy:

$$\begin{aligned}\psi_t^2 &= \frac{2^t}{n} \sum_{i=0}^{2^t-1} v_{i_t}^2 - n, \\ \psi_{t-1}^2 &= \frac{2^{t-1}}{n} \sum_{i=0}^{2^{t-1}-1} v_{i_{t-1}}^2 - n, \\ \psi_{t-2}^2 &= \frac{2^{t-2}}{n} \sum_{i=0}^{2^{t-2}-1} v_{i_{t-2}}^2 - n,\end{aligned}$$

a na ich podstawie wyznaczamy dwie statystyki testowe:

$$\nabla \psi_t^2 = \psi_t^2 - \psi_{t-1}^2 \text{ oraz } \nabla^2 \psi_t^2 = \psi_t^2 - 2\psi_{t-1}^2 + \psi_{t-2}^2,$$

które mają rozkład chi-kwadrat z, odpowiednio, 2^{t-1} i 2^{t-2} stopniami swobody.

Zaleca się, by $t \leq \lg n - 2$.

5.12. Test entropii (*Approximate Entropy Test*)

Test został opisany w punkcie II.4.1.5.

5.13. Test sum (*Cumulative Sums Test*)

Test bada ciąg w sposób identyczny z testem maksymalnego odchylenia ścieżki, opisanym w II.5.5, różni się jedynie formułą służącą do wyznaczania p_{val} :

$$P_{val} = 1 - \sum_{k=\left(\frac{-n}{z}\right)^{1/4}}^{\left(\frac{n}{z}\right)^{1/4}} \left[\Phi\left(\frac{4k+1}{\sqrt{n}}z\right) - \Phi\left(\frac{4k-1}{\sqrt{n}}z\right) \right] + \sum_{k=\left(\frac{-n}{z}\right)^{1/4}}^{\left(\frac{n}{z}\right)^{1/4}} \left[\Phi\left(\frac{4k+3}{\sqrt{n}}z\right) - \Phi\left(\frac{4k+1}{\sqrt{n}}z\right) \right].^{145}$$

oraz przeprowadzaniem badania również dla ciągu przeglądanego od końca.

¹⁴³ A. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, s. 181

¹⁴⁴ I.J. Good, *The serial test for sampling number and other tests for randomness*, Proceedings Cambridge Philosophical Society, 49, 1953, s. 276-284

¹⁴⁵ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001, s. 53



W cytowanej pracy, granice określone są od minus do plus nieskończoności,¹⁴⁶ co z oczywistych względów nie mogło znaleźć się w implementacji, w której znaleźć można następujące dwie linie:

```
start = (-n/(int)z+1)/4;
finish = (n/(int)z-1)/4;
```

ważne jest tu, iż zmienne `start`, `finish` oraz `n` są typu `int`, co oznacza, że w rzeczywistości realizowana jest następująca postać omawianej formuły:

$$P_{val} = 1 - \sum_{k=\left\lceil\left(\left\lceil\frac{-n}{z}\right\rceil+1\right)/4\right\rceil}^{\left\lfloor\left(\left\lfloor\frac{n}{z}\right\rfloor-1\right)/4\right\rfloor} \left[\Phi\left(\frac{4k+1}{\sqrt{n}}z\right) - \Phi\left(\frac{4k-1}{\sqrt{n}}z\right) \right] + \sum_{k=\left\lceil\left(\left\lceil\frac{-n}{z}\right\rceil-3\right)/4\right\rceil}^{\left\lfloor\left(\left\lfloor\frac{n}{z}\right\rfloor-1\right)/4\right\rfloor} \left[\Phi\left(\frac{4k+3}{\sqrt{n}}z\right) - \Phi\left(\frac{4k+1}{\sqrt{n}}z\right) \right].$$

Nie jest do końca jasne, czy było to intencją autorów. Na szczęście nawet dla minimalnej zalecanej wartości $n = 100$ zmiana orientacji zaokrągleń nie wpływa na wynik końcowy, gdyż dla wszystkich możliwych wartości z argumenty są tak duże, że funkcja $\Phi(x)$ przyjmuje wartość 0 albo 1 z bardzo dużą dokładnością.

5.14. Test losowych wycieczek (*Random Excursions Test*)

Pomimo zbieżności nazw, w teście wyznaczane są inne statystyki, aniżeli opisane w punkcie II.5.5. Na podstawie badanego ciągu tworzymy ciąg sum częściowych

$$S_k = \sum_{i=1}^k (2s_i - 1), \quad k = 1..n,$$

który następnie poszerzamy o dwa wyrazy $S_0 = S_{n+1} = 0$.

Niech $J = \#\{k = 1..n+1 : S_k = 0\}$ będzie liczbą tzw. cykli, czyli odcinków ciągu S_k rozpoczynających i kończących się 0 i nie posiadających wewnątrz 0, albo używając wycieczkowej terminologii – liczbą przejść przez 0 na ścieżce odpowiadającej ciągowi S_k . Jeśli zachodzi $J < 0,005\sqrt{n}$, to przerywamy procedurę testową odrzucając hipotezę o losowości badanego ciągu.

Następnie dla każdego $x \in \{-4, -3, -2, -1, 1, 2, 3, 4\}$ wyznaczana jest liczba cykli, w których liczba x występuje dokładnie 0, 1, 2, 3, 4 oraz co najmniej 5 razy, co oznaczymy jako: $v_k(x)$, $k = 0..5$.

Dla każdego x obliczamy następnie statystykę testową:

$$S_{obs}(x) = \sum_{k=0}^5 \frac{(v_k(x) - J\pi_k(x))^2}{J\pi_k(x)},$$

¹⁴⁶ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001, s. 93

gdzie $\pi_k(x)$ jest prawdopodobieństwem, że wartość x wystąpi w cyklu dokładnie k razy:

$$\pi_0(x) = 1 - \frac{1}{2|x|},$$

$$\pi_k(x) = \frac{1}{4x^2} \left(1 - \frac{1}{2|x|}\right)^{k-1}, \quad k = 1..4,$$

$$\pi_5(x) = \frac{1}{2|x|} \left(1 - \frac{1}{2|x|}\right)^4.$$

$S_{obs}(x)$ ma w przybliżeniu rozkład chi-kwadrat z 5 stopniami swobody.

Ponieważ prawdopodobieństwa uzyskiwane dla $k = 4$ są bardzo małe, więc, aby zapewnić warunki dla testu chi-kwadrat, zaleca się stosowanie ciągów o długości nie mniejszej niż 10^6 lub dodatkowego warunku $J \geq 5 / \pi_4(4)$.

W omawianej pracy znaleźć można następujące uzasadnienie dla wyboru $J < 0,005\sqrt{n}$, jako dolnego progu liczby cykli w ciągu:

$$\Pr(J < J_{obs}) \approx \sqrt{\frac{2}{\pi}} \int_0^{J_{obs}/\sqrt{n}} e^{-u^2/2} du = igf\left(\frac{1}{2}, \frac{J_{obs}^2}{2n}\right),^{147}$$

gdzie igf oznacza niekompletną funkcję gamma. Łatwo zauważyć, że mamy tu do czynienia z całką z gęstości rozkładu $N(0;1)$, przez co prawą stronę można zapisać również jako:

$$\sqrt{\frac{2}{\pi}} \int_0^{J_{obs}/\sqrt{n}} e^{-u^2/2} du = 2 \cdot \Phi\left(\frac{J_{obs}}{\sqrt{n}}\right) - 1,$$

dzięki czemu, bez problemu, można z tablic rozkładu $N(0;1)$ odczytać, że poszukiwane wartości progowe to

$$0,00627\sqrt{n} \leq J < 2,80703\sqrt{n}$$

i odpowiadają one łącznie prawdopodobieństwu 0,01 w powyższej formule.

Dla porównania wartość $0,005\sqrt{n}$ daje prawdopodobieństwo równe w przybliżeniu 0,00399, co więcej zalecaną długością ciągu jest 10^6 , więc dostajemy $\sqrt{n} = 1000$, a więc w oryginalnej formule odrzucane będą ciągi z liczbą cykli co najwyżej 4, w wyniku czego prawdopodobieństwo to spada do 0,00319.

O wiele gorsze konsekwencje ma jednak drugie ograniczenie dla J , które na podstawie dokumentacji powinno mieć postać $J \geq 477$, lecz w kodzie programu zapisano na stałe $J \geq 500$. Dramatem nie jest jednak to potknięcie, a warunek sam w sobie, gdyż oznacza on, ni mniej ni więcej, że testu, z przyczyn czysto technicznych, nie stosuje się przeciętnie do przeszło 38% ciągów.

¹⁴⁷ A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001, s. 93

Wobec tego warunku pierwszy, w oryginalnej postaci, zacznie być istotny dopiero dla ciągów o długości 10^{10} bitów, czyli ok. 1,16 GiB. Sytuację tę można znacząco poprawić wprowadzając możliwość grupowania klas w teście zgodności. W ten sposób możliwe jest osiągnięcie odsetka nieco ponad 7,3% ciągów, dla których test nie będzie się stosował.

Ostatnim poważnym uchybieniem jest nierozróżnianie przez implementację testu obu powyższych powodów przerwania procedury testowej – nawet gdy $J < 0,005\sqrt{n}$, a więc wykryto usterkę losowości, stwierdza się, że test nie może być zastosowany ze względu na niewystarczającą liczbę cykli.

5.15. Alternatywny test losowych wycieczek (*Random Excursions Variant Test*)

Tak jak w poprzednim teście, oznaczmy ciąg sum częściowych rozszerzony o dwa wyrazy:

$$S_k = \sum_{i=1}^k (2s_i - 1), \quad k = 1..n,$$

$$S_0 = S_{n+1} = 0,$$

oraz liczbę cykli zawartych w tym ciągu:

$$J = \#\{k = 1..n+1 : S_k = 0\}.$$

Ponadto niech $v(x)$ dla $x \in \{-9, \dots, -1, 1, \dots, 9\}$ oznaczają liczbę wystąpień stanu x w ciągu S_k . Rozpatrujemy 18 statystyk testowych postaci

$$S_{obs}(x) = \frac{v(x) - J}{2\sqrt{J(2|x| - 1)}},$$

które mają asymptotycznie standardowy rozkład normalny.

Należy zwrócić uwagę na identyczne, jak w poprzednim teście, ograniczenie dotyczące J , powodujące identyczne ograniczenia w stosowalności testu.

5.16. Test drugiego poziomu

W pracy [Z16] autorzy odnoszą się do metody przeprowadzania testu drugiego poziomu w pakiecie NIST. W teście pierwszego poziomu wyznaczana jest wartość p_{val} , odpowiadająca prawdopodobieństwu, że idealny generator wytworzyłby ciąg posiadających większe lub równe obserwowanemu odchylenie. Z jednej strony, takie prawdopodobieństwo posiada oczywisty walor informacyjny – możemy je bezpośrednio porównywać z poziomem istotności. Z drugiej jednak, w części przypadków, wprowadza znaczące zakłócenie, przekłamujące dane będące podstawą do przeprowadzenia testu drugiego poziomu. Efekt ten pojawia się w przypadku testów, które wykorzystują statystykę przybliżaną rozkładem normalnym. Stosuje się wówczas obustronny przedział krytyczny, odpowiadający dużym odchyleniom od wartości oczekiwanej.

Samo prawdopodobieństwo wyznaczone jest jako: $p_{val} = 1 - |1 - 2\Phi(S)|$, gdzie S jest odpowiednio znormalizowaną statystyką testową.

Zamiast powyższego autorzy proponują stosowanie wartości nazwanej przez nich q -value, czyli: $q_{val} = 1 - \Phi(S)$. Widoczne dopełnienie dla prawdopodobieństwa najprawdopodobniej ma na celu ujednolicenie postaci prawdopodobieństwa wyznaczanego w testach wykorzystujących rozkłady normalny i chi-kwadrat. Naszym zdaniem, w sytuacji, gdy tak wyznaczone cząstkowe prawdopodobieństwa nie są wprost interpretowane, a jedynie stanowią podstawę do przeprowadzenia testu drugiego poziomu, należałoby w obu przypadkach zrezygnować z odejmowania, gdyż nic ono nie wnosi.

Autorzy wykazali, iż, przy założeniu zgodności S ze standardowym rozkładem normalnym, oba prawdopodobieństwa p_{val} i q_{val} mają rozkład równomierny na przedziale $(0;1)$.

6. TestU01

TestU01 jest zbiorem różnorodnych testów statystycznych zaimplementowanych w ANSI C na Uniwersytecie Montrealu. Celem twórców było stworzenie biblioteki posiadającej o wiele większe możliwości i znacznie elastyczniejszej, niż ma to miejsce w przypadku bibliotek DIEHARD, czy też opracowanej w NIST. W skład omawianego zbioru weszła duża część testów omówionych w rozdziale II.

Zawartość pakietu, jak również wyniki analityczne dotyczące części testów omówione zostały w dwóch publikacjach Pierrea L'Ecuyera i Richarda Simarda^{148,149}, które dołączone są do biblioteki. Pakiet podzielony został na kilka klas testów, omówionych w kolejnych punktach.

6.1. Klasa scomp

Klasa scomp zawiera testy związane ze złożonością liniową ciągu oraz złożonością Lempela-Ziva.

6.1.1. Testy złożoności liniowej

Zaimplementowane testy opisano w punkcie II.5.7.

6.1.2. Test złożoności Lempela-Ziva

Test opisano w punkcie II.5.8.

Choć w dokumentacji pakietu¹⁵⁰ znaleźć można informację o zastosowaniu późniejszej wersji algorytmu, to jednak analiza implementacji wykonana przez Piotra Szabuńko¹⁵¹ wskazała, że korzystano z oryginalnej, opisanej w [ZL78]¹⁵².

¹⁴⁸ P. L'Ecuyer, R. Simard, *TestU01: A C Library for Empirical Testing of Random Number Generators*, 2002

¹⁴⁹ P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, Université de Montreal, 2002, 2007, <http://www.iro.umontreal.ca/~lecuyer/>

¹⁵⁰ P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, Université de Montreal, 2002, 2007, s. 142

Ponieważ, podane przez nas i użyte przez autorów pakietu, przybliżenie okazuje się być bardzo niedokładne dla praktycznych długości ciągu, na potrzeby implementacji wyznaczony został eksperymentalnie zestaw średnich i odchyłeń standardowych dla $k = 3..28$.¹⁵³

Na podstawie wartości oczekiwanych dla $k = 19$ i 20 aproksymowaliśmy wartość odpowiadającą $n = 10^6$, otrzymując 69591,02 i odnieśliśmy ją do dokładniejszych wyników zawartych w pakiecie STS, otrzymując $p_{val} \approx 0,73$. Taka, dość niska wartość, uzasadnia uwagę autorów, by N , czyli liczba powtórzeń testu na potrzeby testu drugiego poziomu, nie była zbyt duża – w przeciwnym razie wykryta zostanie nie tyle nielosowość ciągu, co niepoprawność przyjętego rozkładu statystyki testowej.

6.2. Klasa *smultin*

Klasa *smultin* zawiera zbiór testów wykorzystujących rozkłady wielowymiarowe, zarówno dla wektorów nakładających się, jak i nie.

6.2.1. Wektory nienakładające się

Testy zaimplementowane w tej grupie opisano w punkcie II.3.1.

6.2.2. Wektory nakładające się

Testy zaimplementowane w tej grupie opisano w punkcie II.4.1.

6.3. Klasa *snpair*

Klasa *snpair* zawiera zbiór testów badających odległości między punktami rozmieszczonymi w t -wymiarowym torusie, powstałym poprzez utożsamienie przeciwległych boków hipersześcianu.

6.3.1. Test m najbliższych par (m Nearest Pairs Test)

Test opisano w punkcie II.3.2.

6.3.2. Test zgodności bitowej (Close Pairs Bit Match)

Test opisano w punkcie II.3.3.

6.3.3. Test Bickela-Breimana

Test opisano w punkcie II.3.4.

6.4. Klasa *sspacings*

Klasa *sspacings* zawiera zbiór testów badających rozkład odstępów w posortowanym niemalejąco ciągu liczb z przedziału.

6.4.1. Test sumy logarytmów odstępów

Test opisano w punkcie II.2.12.

¹⁵¹ P. Szabuńko, *Wykonanie implementacji i analiza teoretyczna testu złożoności sekwencji wykorzystującego algorytm kompresji Lempel-Ziv*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2018

¹⁵² J. Ziv, A. Lempel, *Compression of Individual Sequences via Variable-Rate Coding*, IEEE on Information Theory, 24, 5, 1978, s. 533

¹⁵³ P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, Universite de Montreal, 2002, 2007

6.4.2. Test sumy kwadratów odstępów

Test opisano w punkcie II.2.13.

6.4.3. Test największej liczby elementów w przedziale

Test opisano w punkcie II.2.15.

6.5. Klasa *sspectral*

Klasa *sspectral* zawiera trzy testy oparte na dyskretnej transformacie Fouriera.

6.5.1. Test spektralny

Test opisano w punkcie II.5.9.1.

6.5.2. Test spektralny sumy kwadratów modułów – wariant 1

Test opisano w punkcie II.5.9.2.

6.5.3. Test spektralny sumy kwadratów modułów – wariant 2

Test opisano w punkcie II.5.9.3.

6.6. Klasa *sstring*

Klasa *sstring* zawiera testy przeznaczone dla sekwencji bitów.

6.6.1. Test okresowości w ciągu (Periods in Strings Test)

Test opisano w punkcie II.5.10.

6.6.2. Test najdłuższej serii jedynek w bloku

Test opisano w punkcie II.5.11.

6.6.3. Test częstości wag Hamminga

Test opisano w punkcie II.5.12.

6.6.4. Test wag Hamminga

Test opisano w punkcie II.5.12.2.

6.6.5. Test korelacji wag Hamminga

Test opisano w punkcie II.5.12.3.

6.6.6. Testy niezależności wag Hamminga

Test opisano w punkcie II.5.12.4.

6.6.7. Testy do liczby serii N

Test opisano w punkcie II.5.4.

6.6.8. Test autokorelacyjny

Test opisano w punkcie II.5.1.

6.7. Klasa *svaria*

Klasa *svaria* zawiera zbiór podstawowych testów, które nie znalazły miejsca w innych klasach.

6.7.1. Test średniej (Sample Mean)

Test opisano w punkcie II.2.3.

6.7.2. Test korelacji (Sample Correlation)

Test opisano w punkcie II.2.5.

6.7.3. Test iloczynów (Sample Product)

Test opisano w punkcie II.2.6.

6.7.4. Test sumy logarytmów (Sum of Logarithms)

Test opisano w punkcie II.2.7.

6.7.5. Test częstości dla podprzedziału (Weight Distribution)

Test opisano w punkcie II.2.8.

6.7.6. Test kolizji indeksów maksymalnych elementów (Collision Argument Maximal)

Test opisano w punkcie II.2.9.

6.7.7. Test sum do g (Sum Collector)

Test opisano w punkcie II.2.10.

6.7.8. Test odległości (Appearance Spacings)

Jest to uniwersalny test Maurera, przy czym wykorzystano wyniki zawarte w [CN98]¹⁵⁴. Test opisano w punkcie II.2.16.

6.8. Klasa *swalk*

Klasa `swalk` zawiera implementacje testów opartych na losowym błędzeniu na zbiorze liczb całkowitych.

6.8.1. Testy losowych wycieczek (Random Walk)

Test opisano w punkcie II.5.5.

Autorzy proponują prowadzenie testów dla wszystkich parzystych wartości L z przedziału $[L_0, L_1]$. Jednak nie określono jak należy taki zbiór dalej rozpatrywać, a jest oczywiste, że dla kolejnych L wartości wszystkich wyznaczanych statystyk są bardzo mocno skorelowane, stąd poważnym błędem byłoby poddanie ich zbioru wprost testowi drugiego poziomu. Wobec powyższego nie dziwi, że w definicjach baterii testów zawsze przyjmowane jest $L_0 = L_1$.

6.8.2. Testy losowych wycieczek z liniowym wyznaczaniem kierunku

Test opisano w punkcie II.5.6.

Załączona implementacja dopuszcza rozpiętości kombinacji liniowych do $t \leq 31$.

6.8.3. Test drogi do (Random Walk)

Procedura testowa polega na N -krotnym wyznaczeniu r – liczby kolejnych elementów ciągu U^n , które należy przejrzeć, aby znaleźć pierwszą wartość nie mniejszą niż ustalone g . Następnie dla otrzymanego zbioru przeprowadza się test zgodności z rozkładem teoretycznym.

Test ten jest szczególnym przypadkiem testu odstępów, opisanego w punkcie II.2.11.

¹⁵⁴ J.S. Coron, D. Naccache, *An Accurate Evaluation of Maurer's Universal Test*, Proceedings of SAC '98 (Lecture Notes in Computer Science), Springer-Verlag, 1998

7. Dieharder

Dieharder jest opensourceowym projektem zainicjowanym na Duke University przez Roberta G. Browna. Wedle słów autora ma rozwiązać problemy związane z niejednoznacznościami wokół licencji, na których rozpowszechniane są kody bibliotek DIEHARD oraz NIST Statistical Test Suite, jak również dzięki reimplementacji w C testów z DIEHARD dostarczyć czytelnego kodu w języku C.¹⁵⁵

Jako test drugiego poziomu, zastosowany został test Kuipera, nie zaś testu Andersona-Darlinga, jak to miało miejsce w pakiecie DIEHARD.

W kolejnych wersjach pakiet był poprawiany i uzupełniany. Obecnie pakiet zawiera kilkanaście, bardzo zróżnicowanych, testów. Poniżej prezentujemy testy odpowiadające domyślnemu badaniu w ostatniej, dostępnej na początku 2024 roku, wersji – 3.31.1.¹⁵⁶

7.1. Testy zaczerpnięte z pakietu DIEHARD

7.1.1. Test odstępów dni urodzin (`diehard_birthdays`)

Test opisano w punkcie II.2.14, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.1.

7.1.2. Test nakładających się permutacji piątek (`diehard_operm5`)

Test opisano w punkcie II.4.2, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.2.

Przeglądając implementację testu zauważyliśmy adnotację o tym, że raz na kilkadziesiąt uruchomień występował błąd w postaci ujemnej wartości statystyki testowej, z którym poradzono sobie biorąc wartość bezwzględną tejże. Wydaje się, że znaleźliśmy uzasadnienie tego zjawiska – błędne założenie o tym, że wszystkie liczby w ramach piątki są parami różne, w połączeniu ze specyficznym sposobem określania, z którym przypadkiem permutacji mamy do czynienia. Prawdopodobieństwo tego, że w ramach jednej piątki, co najmniej, dwie liczby będą takie same, równe jest $1 - (1 - 1/2^{32}) \cdot (1 - 2/2^{32}) \cdot (1 - 3/2^{32}) \cdot (1 - 4/2^{32}) \approx 0,23283 \cdot 10^{-8}$, co oznacza, że z taką piątką spotkamy się przeciętnie raz na 429,5 badanych ciągów.

W związku z powyższym należałoby bądź to zabezpieczyć implementację przed takim przypadkiem, bądź uwzględnić go w rozważaniach teoretycznych, tak jak uczyniliśmy to w tej pracy.

7.1.3. Test rzędu macierzy binarnej (`diehard_rank_32x32` i `diehard_rank_6x8`)

Test opisano w punkcie II.5.13, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.3.

7.1.4. Test nakładających się słów 20-bitowych (`diehard_bitstream`)

Test opisano w punkcie II.4.3, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.4.1.

7.1.5. Test nakładających się par (`diehard_opso`)

Test opisano w punkcie II.4.3, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.4.2.

¹⁵⁵ <https://webhome.phy.duke.edu/~rgb/General/dieharder.php>

¹⁵⁶ webhome.phy.duke.edu/~rgb/General/dieharder.php

7.1.6. Test nakładających się czwórek (diehard_oqso)

Test opisano w punkcie II.4.3, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.4.3.

7.1.7. Test DNA (diehard_dna)

Test opisano w punkcie II.4.3, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.4.4.

7.1.8. Test „policz jedyne w ciągu bajtów” (diehard_count_1s_str)

Test opisano w punkcie II.4.4, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.5.

7.1.9. Test „policz jedyne w bajtach” (diehard_count_1s_byt)

Test opisano w punkcie II.4.4, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.6.

7.1.10. Test parkingowy (diehard_parking_lot)

Test opisano w punkcie II.3.7, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.7.

7.1.11. Test minimalnej odległości (diehard_2dsphere)

Test opisano w punkcie II.3.2, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.8.

7.1.12. Test kul (diehard_3dsphere)

Test opisano w punkcie II.3.2, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.9.

7.1.13. Test ściskania (diehard_squeeze)

Jest to szczególny przypadku test opisanego w punkcie II.2.24., a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.10.

7.1.14. Test nakładających się sum (diehard_sums)

Test opisano w punkcie III.2.11.

7.1.15. Test serii wg DIEHARD (diehard_runs)

Test opisano w punkcie II.2.22.2, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.12.

7.1.16. Test oparty na grze w kości (diehard_craps)

Test opisano w punkcie II.2.25, a jego wystąpienie w pakiecie DIEHARD w punkcie III.2.13.

7.2. Testy zaczerpnięte z pakietu STS

7.2.1. Test częstości bitów (sts_monobit)

Test jest tożsamy z testem momentu pierwszego rzędu dla bloku jednobitowego, opisanym w punkcie II.2.2.2, jego wystąpienie w pakiecie STS omówiono w punkcie III.5.1.

7.2.2. Test serii wg NIST (sts_runs)

Test opisano w punkcie II.5.2., a jego wystąpienie w pakiecie STS w punkcie III.5.3.

7.2.3. Test częstości nakładających się bloków (sts_serial)

Test opisano w punkcie III.5.11.

Wykonywanych jest 16 procedur testowych, dla $m = 1..16$, przy czym:

- dla $m = 1$ wykorzystywana jest statystyka ψ_1^2 , odpowiadająca testowi monobitowemu,
- dla $m = 2$ wykorzystywana jest statystyka $\nabla \psi_2^2$,

- dla $m > 2$ wykorzystywana jest statystyka $\nabla^2 \psi_m^2$.

7.3. Test największego wspólnego dzielnika (*marsaglia_tsang_gcd*)

Test opisano w punkcie II.2.26.

Do wyznaczenia empirycznych tablic koniecznych do testu liczby kroków Brown posłużył się czterema generatorami: gfsr4, mt19937 1999, rndlxs2 i taus2.

Po 100-krotnym powtórzeniu procedury pierwszego poziomu, uzyskane p_{val} są oceniane przy użyciu testu Kołmogorowa-Smirnowa.

7.4. Test częstości bloków bitów (*rgb_bitdist*)

Test bada rozkład częstości obserwacji bloków bitów, nie jest to jednak klasyczne podejście wykorzystujące test częstości Pearsona.

Na podstawie badanego ciągu binarnego tworzony jest ciąg nienakładających się d -bitowych bloków, a ten z kolei, na podciągi o długości M bloków. Dla każdego z N takich podciągów obliczana jest liczba wystąpień wybranego d -bitowego bloku, a następnie uzyskany rozkład porównywany jest z rozkładem teoretycznym przy użyciu testu zgodności Pearsona.

Autor zwraca uwagę na dwa krytyczne aspekty opisanej procedury. Pierwszą z nich jest zakres stosowalności testu zgodności Pearsona, który wymusza grupowanie kategorii lub ignorowanie części z nich – poziom tzw. odcięcia ustawiony został na 20. Drugą jest zależność statystyk uzyskiwanych dla różnych wartości bloku, stąd decyzja o losowym wyborze tylko jednej z nich, pomimo, że obliczenia prowadzone są dla wszystkich 2^d wartości bloku. Jest to istotny mankament, gdyż powodować może przeoczenie usterki losowości objawiającej się jedynie dla części możliwych wartości bloków.

Powyższa procedura jest nowa i nie była wcześniej stosowana w żadnym pakiecie. Wydaje się być ciekawym rozszerzeniem badania przy użyciu testu częstości dla bloków d -bitowych.

W przypadku implementowania testu rozkładu częstości bloków, należy przeanalizować korelację statystyk testowych uzyskiwanych dla różnych wartości bloku, gdyż pomimo oczywistej zależności odpowiadających im zmiennych losowych, w naszym odczuciu, siła korelacji jest na tyle mała, że będzie mogła być zignorowana.

7.5. Uogólniony test minimalnej odległości (*rgb_minimum_distance*)

Test jest uogólnieniem testów minimalnej odległości. Autor posiłkował się wynikami uzyskanymi przez M. Fischler'a. Pokazał on, że zmienna losowa opisująca najmniejszą odległość pomiędzy N punktami rozmieszczonymi w t -wymiarowej hiperkostce jednostkowej, ma w przybliżeniu rozkład opisany formułą:

$$P(S_{obs}^2 < x^2) = 1 - e^{-\frac{N(N-1)}{2}V_t(x)} \left(1 + \frac{2+Q}{6} N^3 V_t^2(x) \right),^{157}$$

gdzie

$$V_t(x) = \begin{cases} \frac{\pi^{t/2} x^t}{(t/2)!}, & 2 \mid t, \\ \frac{2(2\pi)^{(t-1)/2} x^t}{t!!}, & 2 \nmid t, \end{cases}$$

jest objętością t -wymiarowej hiperkuli o promieniu x , zaś Q jest poprawką związaną z nakładaniem się hiperkul, wynoszącą:

$$V_t(x) = \begin{cases} \frac{2\sqrt{3}}{4\pi}, & t = 2, \\ \frac{17}{32}, & t = 3, \\ \frac{9\sqrt{3}}{8\pi}, & t = 4, \\ \frac{353}{256}, & t = 5. \end{cases}$$

W przytoczonej pracy znajduje się opis procedury mogącej znacząco przyspieszyć wyszukiwanie minimalnej odległości, polegającej na podziale przestrzeni na kostki, których rozmiar może zostać wybrany w sposób suboptymalny – Fischler dobrał rozmiar kostki wychodząc od górnego oszacowania maksymalnej wartości minimalnej odległości, przez co otrzymana wielkość jest stosunkowo duża:

$$\frac{\sqrt{t}}{\sqrt[t]{N}}.$$



W pracy [Ma22] pokazujemy metodę znaczącego przyspieszenia testu dzięki wykorzystaniu podejścia probabilistycznego z arbitralnie małym prawdopodobieństwem błędu.

7.6. Test opóźnionych sum (*rgb_lagged_sums*)

Test ten jest kompilacją trzech elementów – dwóch testów i procedury formowania badanego ciągu. Przeznaczony jest do badania ciągu zmiennych losowych o rozkładzie równomiernym na przedziale $[0;1]$, przy czym właściwe badanie odbywa się dla ciągu powstałego poprzez odrzucanie

¹⁵⁷ M. Fischler, *Distribution of Minimum Distance Among N Random Points in d Dimensions*, DOI: 10.2172/794005, FERMILAB-TM-2170, 2002, s. 11-12

pewnej stałej liczby elementów przed wzięciem kolejnego. Taki przerzedzony ciąg jest następnie badany przy użyciu klasycznego testu momentu pierwszego rzędu, a seria takich badań jest weryfikowana przy pomocy testu Kołmogorowa.

Standardowo badanie wykonywane jest dla ciągu nieprzerzedzonego oraz dla odrzucania od 1 do 32 elementów.

Test został wprowadzony w następstwie obserwacji, iż procedury takiej nie przechodzi generator Mersenne Twister ze stanem 19937-bitowym, dający zbyt równomierny rozkład.

Minimalna usterka w omawianej procedurze wiąże się z wyznaczaniem prawdopodobieństwa pierwszego poziomu, gdzie formuła:

$$\Pr(\bar{X}_n < x) = \Phi\left(\left(x - \frac{1}{2}\right)\sqrt{12n}\right) - \frac{1}{20n}\varphi^{(3)}\left(\left(x - \frac{1}{2}\right)\sqrt{12n}\right),^{158}$$

opisana w punkcie II.2.2.1., ograniczona została do pierwszego składnika.

Autor miesza, budowanie testu losowości z przygotowywaniem ciągu do badania. Zupełnie niezrozumiałe jest, dlaczego nie wykorzystał on procedury formowania badanego ciągu na podstawie wyjścia generatora jako elementu mogącego poprzedzać dowolny test losowości, przez co zyskałby ogromną swobodę badania i możliwość uogólnienia proponowanych procedur.

7.7. Test *lmn* (*rgb_lmn*)

Zaimplementowany w tym zbiorze test `rgb_lmn_test` jest swoistym uogólnieniem opisanego w poprzednim punkcie. Różnica polega na sposobie formowania badanego, ponownie przy użyciu testu momentu pierwszego rzędu, ciągu. W tym przypadku ze źródłowego ciągu pobierane jest l bitów, następnie m bitów jest pomijane i pobierane jest n bitów, przy czym nie jest jasne, czy dotyczy to każdego 32-bitowego słowa oddzielnie, czy też, jak sugeruje opis czynione jest na ciągłej sekwencji.

Ponownie zauważamy, że autor nie wykorzystuje potencjału proponowanej procedury, ograniczając się do jednego tylko testu, co więcej wyraża on przekonanie, że taka procedura, powtarzana dla różnych kombinacji l , m i n powinna być w stanie zastąpić wszystkie / wiele innych testów.

Analiza przeprowadzona, w ramach realizacji pracy dyplomowej, przez mgr inż. Piotra Szabuńko¹⁵⁹ pokazała, że test ten jest bardzo czasochłonny ze względu na rozmiar przestrzeni jego parametrów, zaś przy ograniczeniu się do rozsądnych zakresów nie jest w stanie wykryć nawet tak prostych generatorów jak 127-bitowy LFSR.

7.8. Test nakładających się permutacji (*rgb_operm*)

Test opisany jest w punkcie II.4.2. W przeciwieństwie do oryginału pochodzącego z DIEHARD'a, implementacja Dieharder'owa nie ogranicza się do permutacji pięcioelementowych. Kod umożliwia badanie rozkładu nawet dziesięcioelementowych permutacji, przy czym granica ta wy-

¹⁵⁸ R. Wieczorkowski, R. Zieliński, *Komputerowe generatory liczb losowych*, WNT, Warszawa, 1997, s. 118

¹⁵⁹ P. Szabuńko, *Analiza własności testu losowości rgb_lmn*, promotor dr L. Kowalski, Warszawa, 2019

daje się mocno zawyżona ze względu na rozmiar danych w postaci budowanych macierzy kowariancji.

Zamiast formy kwadratowej, wykorzystującej słabą macierz odwrotną Moore’a–Penrose’a, autor sugeruje w komentarzach statystykę testową w postaci iloczynu macierzy odwrotnej macierzy kowariancji oraz empirycznej macierzy kowariancji. Taki iloczyn dążyć ma asymptotycznie do macierzy jednostkowej, zaś suma elementów na głównej przekątnej iloczynu powinna mieć rozkład chi-kwadrat. Autor bardzo enigmatycznie odwołuje się, najprawdopodobniej do pracy [AKM04]¹⁶⁰, choć nie udało mi się odnaleźć w niej odniesienia do zaproponowanego sposobu liczenia statystyki testowej.

Mocno niepokojące jest, z praktycznego punktu widzenia, iż wymagana w teście macierz kowariancji i jej słaba macierz odwrotna obliczane są każdorazowo w trakcie wykonywania testu. Czasochłonność związanych z tym obliczeń wskazuje, że droga obrona w omawianej implementacji jest niepraktyczna.

Wydaje się, że, choć zapiski w komentarzach wskazują, że ignoruje on możliwość pojawiania się układów niepermutacyjnych, to przyjęta przezeń metoda wyznaczania macierzy kowariancji uwzględnia ich wpływ na ostateczną postać macierzy.

7.9. Test permutacji (*rgb_permutations*)

Jest to implementacja klasycznego testu częstości nienakładających się permutacji, w którym badana jest równomierność częstości występowania nienakładających się permutacji.

Chcielibyśmy zwrócić uwagę na pomijany zazwyczaj aspekt efektywności testu, tzn. rzeczywistego wpływu poszczególnych bitów badanego ciągu na wartość wyznaczanej statystyki testowej.

W omawianej implementacji dane pobierane są w porcjach 32-bitowych i interpretowane jako binarne rozwinięcie liczby z przedziału $[0;1)$. Implementacja dopuszcza różne długości permutacji, co dalej oznaczać będziemy jako t , zaś długość reprezentacji elementu (liczbę bitów nań przypadającą) jako d .

Aby ustalić, z którym układem permutacyjnym mamy do czynienia, należy dokonać porównania w każdej z $\binom{t}{2}$ par – po $t - 1$ razy dla każdego z elementów.

Prawdopodobieństwo, że porównywanie osiągnie k -ty najbardziej znaczący bit i na nim się zakończy wynosi:

$$p(k) = \begin{cases} 2^{-k}, & \text{dla } k = 1..d-1, \\ 2^{1-d}, & \text{dla } k = d. \end{cases}$$

¹⁶⁰ A. Alhakim, J. Kawczak, S. Molchanov, *On the class of nilpotent Markov chains, I: The spectrum of covariance operator*, Markov Processes and Related Fields 10, 2004, s. 645-46

Po wykonaniu $t - 1$ porównań, prawdopodobieństwo, że co najmniej jedno z porównań osiągnie k -ty najbardziej znaczący bit i żadne nie osiągnie dalszego wynosi:

$$p_t(k) = \begin{cases} \left(1 - 2^{-k}\right)^{t-1} - \left(1 - 2^{1-k}\right)^{t-1}, & \text{dla } k = 1..d-1, \\ 1 - \left(1 - 2^{1-d}\right)^{t-1}, & \text{dla } k = d. \end{cases}$$

Na podstawie powyższej formuły możemy numerycznie łatwo wyznaczyć wartości oczekiwane liczby bitów, które zostaną wykorzystane przez, co najmniej, jedno porównanie, dla różnych kombinacji t i d . Zestawiliśmy je w poniższej tabeli.

Tabela III.5. Wartości oczekiwane liczby bitów wykorzystanych w porównaniach dla różnych kombinacji t i d

$d \setminus t$	2	3	4	5	6	8	9	10	100	1000
4	1,875	2,422	2,783	3,035	3,219	3,466	3,552	3,622	4,000	4,000
8	1,992	2,651	3,119	3,474	3,755	4,187	4,359	4,512	7,285	8,000
10	1,998	2,663	3,137	3,497	3,784	4,227	4,405	4,564	7,782	9,838
12	2,000	2,666	3,141	3,503	3,792	4,237	4,417	4,577	7,921	10,847
16	2,000	2,667	3,143	3,505	3,794	4,241	4,421	4,581	7,966	11,267
24	2,000	2,667	3,143	3,505	3,794	4,241	4,421	4,581	7,969	11,298
32	2,000	2,667	3,143	3,505	3,794	4,241	4,421	4,581	7,969	11,298

Dla znanej również z DIEHARD'a kombinacji $t = 5$ i $d = 32$ otrzymaliśmy 3,505. Oznacza to, że z każdego 32-bitowego słowa przeciętnie nieco ponad 3,5 bita jest sprawdzanych w trakcie ustalania układu permutacyjnego. Pozostałe 89% nie ma **żadnego znaczenia** i nie wpływa na wartość statystyki testowej.

Co ciekawe, w teście z poprzedniego punktu, w związku z nakładaniem się permutacji, każdy z elementów porównywany jest z dwa razy większą liczbą sąsiadów, w związku z czym średnia liczba „zaangażowanych” bitów wzrasta do 4,421, co i tak sprawia, że 86% jest pomijanych.

W obu przypadkach należy rozważyć bądź to zmniejszenie długości słowa opisującego element permutacji, bądź powtarzanie testu, po uprzednim obrocie słowa o kilka bitów. Drugi z wariantów jest prostszy w realizacji, gdyż nie wymaga wyznaczania nowych zestawów stałych, lecz prowadzi do skorelowania tak uzyskiwanych wyników.

7.10. Test niezmienności bitu (*rgb_persist*)

Test ma za zadanie wykrywać niezmiennosc bitu zajmującego określoną pozycję w słowie zadanej długości (domyślnie 32 bity). Dla kolejnych fragmentów, obejmujących 256 słów, wykrywane jest zajście zdarzenia polegającego na stałości wartości bitu dla którejkolwiek pozycji w słowie.

Elementarnie możemy obliczyć prawdopodobieństwo błędu pierwszego rodzaju w tym teście. Niech d będzie długością słowa, t – liczbą słów rozpatrywanych w pojedynczym sprawdzeniu, zaś n długością ciągu wyrażoną w bitach. Zakładamy, że $d \cdot t \mid n$.

Prawdopodobieństwo, że na określonej pozycji w słowie, w pojedynczym sprawdzeniu, zaobserwujemy przypadkowo stałą wartość bitu to 2^{1-t} . Stąd prawdopodobieństwo, że na dowolnej pozycji w słowie, w pojedynczym sprawdzeniu, zaobserwujemy przypadkowo stałą wartość bitu, wyraża się jako:

$$1 - (1 - 2^{1-t})^d \approx d \cdot 2^{1-t}.$$

W końcu prawdopodobieństwo, że dla n bitowego ciągu zaobserwowany zostanie przynajmniej jeden taki przypadek wyniesie:

$$1 - \left((1 - 2^{1-t})^d \right)^{\frac{n}{d}} = 1 - (1 - 2^{1-t})^{\frac{n}{t}} \approx \frac{n}{t} \cdot 2^{1-t}.$$

Niezależność uzyskanego prawdopodobieństwa od parametru d wynika wprost z przebiegu testu – mamy t -elementowe wektory słów d -bitowych, których odpowiadające sobie współrzędne konkatenujemy, otrzymując wektor słów n/t -bitowych, w którym poszukujemy niezmienności bitu.

Przyjmując $t = 256$ i $n = 2^{33}$ (1 GiB) uzyskamy absurdalnie niską wartość prawdopodobieństwa błędu pierwszego rodzaju: 2^{-230} , podczas gdy w pakiet posługuje się dwoma progami: 10^{-2} i 10^{-6} .

Na podstawie powyższej formuły można łatwo wyznaczyć wartości t , dające, dla danej długości ciągu, najbliższą założonej wartości prawdopodobieństwa przypadkowej obserwacji. Przykładowe wartości, dla kilku klasycznych kombinacji, znaleźć można w poniższej tabeli.

Tabela III.6. Wartości liczby słów, dające najbliższe założonym prawdopodobieństwa błędu pierwszego rodzaju, w teście niezmienności bitu

dług. ciągu	założone prawdopodobieństwo			
	10^{-1}	10^{-2}	10^{-3}	10^{-6}
2^{13}	14	17	20	29
2^{20}	20	23	26	36
$2^{10.5^7}$	26	29	32	42
2^{30}	30	33	36	45
2^{33}	32	36	39	48

Jedynie dla dwóch kombinacji parametrów udało nam się uzyskać t będące dzielnikiem długości ciągu, w dwóch innych uzyskiwane prawdopodobieństwo jest naprawdę bliskie założonej wartości. Oznacza to, że dobór parametrów jest w tym teście bardzo kłopotliwy. Co więcej test zwraca jedynie binarny wynik, przez co bardziej odpowiada pakietom typu FIPS PUB 140, aniżeli budowanemu przez nas. Nie widzimy praktycznego sposobu przekształcenia go do pożądanej przez nas wersji.

7.11. Test częstości bajtów (*dab_bytedistrib*)

Podstawą jest klasyczny test częstości dla bajtów (nienakładających się bloków 8-bitowych – patrz punkt II.2.1.1.), wykorzystujący statystykę Pearsona dla 256 klas. Badanie przeprowadzane

jest oddzielnie dla każdego ze skrajnych bajtów oraz jednego z wewnętrznych bajtów w słowie generatora.

7.12. *Test dyskretnej transformaty kosinusowej (dab_dct)*

Badany ciąg dzielony jest na fragmenty stanowiące ciąg słów stałej długości, będącej potęgą 2, które poddawane są dyskretnej transformacie kosinusowej typu II. Test bada rozkład pozycji maksymalnych, co do modułu, wartości w transformacie, przy użyciu klasycznego testu zgodności Pearsona.

7.13. *Test wypełniania drzewa binarnego na podstawie ciągu liczb (dab_filltree)*

Opis testu znaleźć można w punkcie II.2.27.

Implementacja w pakiecie Dieharder przygotowana została dla drzew o wysokości 4, bada ciąg słów 32-bitowych, wykonując cztery przebiegi poprzez rotowanie ich o wielokrotność 8 bitów. Analiza wykazała, że wbrew opisowi procedura przerywana jest po wypełnieniu pierwszego liścia.

Implementacja zawiera błąd związany z odrzucaniem klas o zbyt małej wartości oczekiwanej liczby obserwacji, co skutkuje tym, że sumy liczby obserwacji i wartości oczekiwanych liczby obserwacji są różne.¹⁶¹

7.14. *Test wypełniania drzewa binarnego na podstawie ciągu bitów (dab_filltree2)*

Opis testu znaleźć można w punkcie II.5.16.

Implementacja w pakiecie Dieharder przygotowana została dla drzew o wysokości 7. Zawiera ona ten sam błąd jak dla poprzedniego testu.¹⁶²

7.15. *Test wag Hamminga bloków (dab_monobit2)*

Test bada liczbę jedynek w podciągach ciągu o długościach $k = 0..m$ słów, gdzie m jest maksymalną możliwą wartością dla danego ciągu. Każdy z wyników przekształcany jest na p_{val} , poprzez dystrybucję rozkładu dwumianowego. Następnie, spośród tych p_{val} , wybierany jest najbardziej skrajny przypadek, a finalne p_{val} wyznaczane jest zgodnie z rozkładem max-z- $(m + 1)$.

8. Crypt-X

Crypt-X jest komercyjnym pakietem opracowanym w australijskim Institute for Future Environments. Oprócz badania generatorów losowych i pseudolosowych zawiera on procedury przeznaczone dla algorytmów blokowych.

Zaimplementowane zostało siedem testów losowości.

¹⁶¹ K. Mańk, *Upgrading Frequency Test for Overlapping Vectors and Fill Tree Tests*, IJET, 2/2025

¹⁶² K. Mańk, *Upgrading Frequency Test for Overlapping Vectors and Fill Tree Tests*, IJET, 2/2025

8.1. *Test częstości*

Realizowany jest test monobitowy, będący szczególny przypadkiem testu opisanego w punkcie II.2.2.2.

8.2. *Test autokorelacji*

Opis testu znaleźć można w punkcie II.2.4.

8.3. *Test częstości dla różnych długości bloków nienakładających się*

Opis testu znaleźć można w punkcie II.2.1.1.

8.4. *Test serii*

Opis testu znaleźć można w punkcie II.5.3.

8.5. *Test złożoności sekwencji*

Sprawdza występowanie odpowiedniej liczby wzorców, uzyskiwanych w algorytmie kompresji Lempela-Ziva. Opis testu znaleźć można w punkcie II.5.8.

8.6. *Test złożoności liniowej*

Opis testu znaleźć można w punkcie II.5.7.

8.7. *Test entropii w bloku*

Opis testu znaleźć można w punkcie II.5.12.

Rozdział IV

Trójpzozimowa procedura testowa

Choć celem niniejszej pracy jest przygotowanie zestawu testów losowości przeznaczonych do badania mocy algorytmów kryptograficznych – narzędzia służącego, między innymi, do oceny jakości generatorów losowych i pseudolosowych, to aby cel taki osiągnąć, w tym miejscu odwróciliśmy to zagadnienie. Tak, jak zostało to omówione w rozdziale I, posłużyliśmy się trójpzozimową procedurą testowania wybranych generatorów do zweryfikowania jakości poszczególnych procedur testowych. Kluczowym jej elementem była para idealnego generatora i testu zgodności.

1. Generator

W wykonywanym badaniu konieczne były ogromne objętości danych pochodzących z idealnego generatora losowego, których nie byliśmy w stanie zapewnić z dwóch powodów:

- niewystarczającej szybkości generatorów opartych na problemach trudnych obliczeniowo,
- braku formalnej weryfikacji charakterystyk entropii posiadanych źródeł fizycznych.

Z powyższych powodów oraz ze względu na wygodę stosowania, zdecydowaliśmy się *zasymulować* idealne źródło przy użyciu zbioru generatorów pseudolosowych. Nasz wybór padł na:

- generator strumieniowy w trybie licznikowym z algorytmem AES256^{163,164} w roli funkcji wyjściowej generatora oraz LFSRem z wielomianem charakterystycznym $x^{127} + x^{64} + 1$ w roli stanu wewnętrznego,
- Trivium,¹⁶⁵
- RC4,¹⁶⁶
- Rabbit.¹⁶⁷

Algorytmy te mają bardzo zróżnicowaną budowę i z bardzo dużym prawdopodobieństwem można przyjąć, że nie mogą posiadać identycznie objawiających się usterek w losowości wytwarzanych za ich pomocą ciągów. W związku z tym, jeśli dana procedura testowa nie pozwala na odróżnienie wytwarzanych przez nie ciągów, to uzasadnione jest przyjęcie, że są one również nieodróżnialne od dobrego źródła losowego.

¹⁶³ J. Daemen, V. Rijmen, *AES Proposal: Rijndael*, 1999

¹⁶⁴ *Advanced Encryption Standard (AES)*, Federal Information Processing Standards, 2001

¹⁶⁵ C. De Cannière, B. Preneel, *Trivium - A Stream Cipher Construction Inspired by Block Cipher Design Principles*, eSTREAM submitted papers, 2006

¹⁶⁶ <https://cypherpunks.venona.com/archive/1994/09/msg00304.html>

¹⁶⁷ M. Boesgaard, M. Vesterager, T. Pedersen, J. Christiansen, O. Scavenius, *Rabbit: A High-Performance Stream Cipher*, Proceedings FSE 2003, Springer LNCS 2887, s. 307-329

2. Test

Naturalnym wyborem, do roli drugiego składnika idealnej pary, jest któryś z testów wykorzystujących statystykę Kołmogorowa albo Andersona-Darlinga, opisanych w punktach II.2 i II.3.

Na samym wstępie wyeliminowaliśmy test Kołmogorowa-Smirnowa, gdyż ma on sens jedynie przy wykorzystaniu obu statystyk: K_n^+ i K_n^- , dla których z kolei, wobec silnej korelacji, nie istnieje formalnie poprawny sposób na sprowadzenie do pojedynczego p_{val} . Test Kołmogorowa-Smirnowa wykorzystaliśmy jedynie do weryfikacji nieodróżnialności generatorów, oczywiście pozostałe cztery testy: Kołmogorowa, Kuipera, Cramera-von Misesa oraz Andersona-Darlinga, również zostały użyte w tej roli.

3. Weryfikacja generatorów

Wszystkie generatory inicjalizowane były pseudolosowymi ciągami uzyskanymi przy pomocy standardowej funkcji `rand`, a uzyskiwane ciągi dzielone były na nienakładające się 64-bitowe bloki traktowane jako binarne rozwinięcia liczb z przedziału $[0;1]$. Dany ciąg badany był przy użyciu wszystkich pięciu testów – sześciu statystyk testowych. Posłużyliśmy się czterema długościami ciągów: 2^3 , 2^7 , 2^{10} i 2^{20} . Dla każdej z 16 kombinacji generator + długość ciągu badanie powtórzone zostało 10 000 razy, a uzyskane p_{val} były podstawą do poniższej analizy.

Użycie aż 64-bitowej precyzji oznaczało wykorzystanie w tym badaniu odsetka mniejszego niż 2^{-28} spośród możliwych wartości bloku, co pozwala założyć, że wykorzystane dane, nawet w postaci pojedynczego ciągu, nie powinny być odróżnialne od realizacji ciągu zmiennych losowych o rozkładzie równomiernym na przedziale $[0;1]$, na skutek samej tylko ich dyskretyzacji.

W celu zweryfikowania odróżnialności któregoś z generatorów przy użyciu któregoś z rozważanych testów zgodności, uzyskane przy ich użyciu p_{val} pierwszego poziomu, porównywaliśmy przy użyciu omówionych poniżej pięciu testów, otrzymując prawdopodobieństwa drugiego poziomu.

- dla wartości oczekiwanych

Ponieważ liczności prób są bardzo duże, więc uznajemy, że statystyka testowa będąca różnicą uzyskanych średnich

$$\bar{X}_1 - \bar{X}_2$$

ma rozkład normalny z wartością oczekiwaną 0 i odchyleniem standardowym

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}},$$

gdzie S_1^2 i S_2^2 są wariancjami z próby otrzymanymi dla obu próbek.¹⁶⁸

¹⁶⁸ J. Zacharski, *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*, Warszawa, 1974, s. 212

- dla wariancji

Wykorzystaliśmy własność, iż iloraz wariancji z próby

$$\frac{\hat{S}_1^2}{\hat{S}_2^2}$$

ma rozkład F Fischera-Snedekora o $(n_1 - 1, n_2 - 1)$ stopniach swobody.¹⁶⁹

- zgodności Pearsona dla dwóch próbek

Wyznaczana jest statystyka

$$P_2 = \sum_{i=1}^k \frac{(n_i^A - n_i^B)^2}{n_i^A},$$

gdzie przedział $[0;1]$ dzielony był na $k = 500$ rozłącznych odcinków, a zaobserwowane licznosci dla poszczególnych klas wynosiły odpowiednio n_i^A i n_i^B .

Dla dużych n statystyka P_2 ma rozkład chi-kwadrat z $2(k-1)$ stopniami swobody.

- ANOVA

Wyznaczana jest statystyka

$$A_{k,n} = \frac{nk(n-1) \sum_{i=1}^k (\bar{X}_n^i - \bar{X})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^n (x_j^i - \bar{X}_n^i)^2},$$

gdzie

k – jest liczbą generatorów,

n – jest liczbą p_{val} dla pojedynczego generatora,

x_j^i – jest j -tym p_{val} dla i -tego generatora,

$\bar{X}_n^i$ – jest średnią p_{val} dla i -tego generatora,

$\bar{X}$ – jest średnią średnich $\bar{X}_n^i$,

która ma rozkład F Fischera-Snedekora z $(k-1, k(n-1))$ stopniami swobody, a p_{val} wyznaczone jest jako wartość prawostronnej dystrybucji $F_{k-1, k(n-1)}$ w punkcie $A_{k,n}$.¹⁷⁰

- test chi-kwadrat Bartletta

Ponieważ test przeznaczony jest dla rozkładu normalnego, przed jego wykonaniem w miejsce poszczególnych p_{val} podstawione zostały odpowiadające im kwantyle standardowego rozkładu normalnego $\Phi^{-1}(p_{val})$.

¹⁶⁹ J. Zacharski, *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*, Warszawa, 1974, s. 207

¹⁷⁰ L. Kowalski, *Statystyka*, BEL Studio, 2001, wydanie I, s. 190-191

Wyznaczana jest statystyka

$$B_{k,n} = \frac{(n-1) \left(k \ln \left(\frac{1}{k} \sum_{i=1}^k \hat{S}_i^2 \right) - \sum_{i=1}^k \ln \hat{S}_i^2 \right)}{1 + \frac{k+1}{3k(n-1)}},$$

gdzie

k – jest liczbą generatorów,

n – jest liczbą p_{val} dla pojedynczego generatora,

$\hat{S}_i^2$ – jest wariancją z próby dla próbki i -tego generatora,

która ma rozkład chi-kwadrat z $k-1$ stopniami swobody.^{171,172}

W tabelach IV.1 do IV.5 zebraliśmy uzyskane prawdopodobieństwa drugiego poziomu, w każdym przypadku podajemy 6 liczb, w kolejności, dla statystyk: Kołmogorowa-Smirnowa K^+ , Kołmogorowa-Smirnowa K^- , Kołmogorowa, Andersona-Darlinga, Cramera-von Misesa, Kuipera. Pomarańczowym kolorem zaznaczyliśmy wartości mniejsze od 0,1, a czerwonym – mniejsze od 0,01.

Tabela IV.1. P_{val} drugiego poziomu uzyskane w teście dla wartości oczekiwanej dla ciągów różnej długości

	AES i Trivium	AES i RC4	AES i Rabbit	Trivium i RC4	Trivium i Rabbit	RC4 i Rabbit
2^3	0,47087 0,40875 0,70571 0,86018 0,95327 0,62662	0,96778 0,68494 0,69094 0,71169 0,90441 0,53142	0,95431 0,52880 0,63502 0,15209 0,25015 0,40980	0,49677 0,21708 0,98511 0,84812 0,85861 0,88903	0,50643 0,14369 0,39555 0,21198 0,22864 0,73428	0,98661 0,82358 0,38374 0,28839 0,30370 0,84128
2^7	0,31052 0,71713 0,94289 0,96225 0,77467 0,64523	0,01608 0,39897 0,12062 0,36569 0,47215 0,10624	0,84324 0,89628 0,75743 0,77397 0,53233 0,84561	0,16277 0,62947 0,13865 0,34045 0,31432 0,25042	0,22609 0,62224 0,81248 0,81016 0,73477 0,78924	0,00930 0,33036 0,21340 0,23280 0,17896 0,15447
2^{10}	0,35080 0,53355 0,60412 0,14892 0,19688 0,68691	0,11089 0,78355 0,72055 0,78990 0,82407 0,45031	0,37433 0,55934 0,59829 0,86295 0,81726 0,85644	0,50207 0,72670 0,37941 0,08671 0,12941 0,24694	0,96704 0,97122 0,29586 0,10737 0,12934 0,55985	0,47788 0,75517 0,86349 0,92642 0,99182 0,56709

¹⁷¹ L. Kowalski, *Statystyka*, BEL Studio, 2001, wydanie I, s. 188-189

¹⁷² M.S. Bartlett, *Properties of Sufficiency and Statistical Tests*, Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences, vol. 160, no. 901, 1937, s. 268-282

2^{20}	0,97425	0,40315	0,35586	0,38368	0,33765	0,93218
	0,17329	0,61779	0,78938	0,06308	0,27161	0,44281
	0,56230	0,43137	0,73571	0,83383	0,80811	0,65127
	0,65794	0,29619	0,70520	0,54585	0,94746	0,50189
	0,68443	0,30091	0,62482	0,52816	0,93465	0,58208
	0,06726	0,29000	0,11711	0,44185	0,79188	0,61285

Do oceny uzyskanych wyników wykorzystaliśmy rozkład dwumianowy. Obliczane były prawdopodobieństwa uzyskania co najmniej zaobserwowanej liczby przekroczeń progów. Wobec braku niezależności zamieszczonych w tabelach IV.1 do IV.5 prawdopodobieństw uzyskane w ten sposób p_{val} trzeciego poziomu należy traktować jako oszacowania dolne, które były dla nas wystarczające.

Wyniki uzyskane w teście wartości oczekiwanej uznać można za wzorcowe dla testowania źródeł losowych. W pięciu przypadkach zaobserwowaliśmy p_{val} niższe od 0,1, co daje bardzo wysokie p_{val} trzeciego poziomu, równe 0,99912. Tylko w jednym przypadku uzyskaliśmy wartość niższą od 0,01 – co daje p_{val} trzeciego poziomu równe 0,76478.

Tabela IV.2. P_{val} drugiego poziomu uzyskane w teście dla wariancji dla ciągów różnej długości

	AES i Trivium	AES i RC4	AES i Rabbit	Trivium i RC4	Trivium i Rabbit	RC4 i Rabbit
2^3	0,62826	0,85758	0,95908	0,76058	0,50643	0,98661
	0,83218	0,84447	0,55416	0,98745	0,14369	0,82358
	0,68124	0,92759	0,84355	0,74906	0,39555	0,38374
	0,59953	0,98697	0,84693	0,61092	0,21198	0,28839
	0,68569	0,89897	0,92475	0,78121	0,22864	0,30370
	0,99361	0,97819	0,74201	0,97180	0,73428	0,84128
2^7	0,98093	0,73276	0,89472	0,75082	0,91366	0,83435
	0,85793	0,95832	0,95265	0,89914	0,81159	0,91111
	0,98205	0,92509	0,93737	0,94298	0,95528	0,98767
	0,88396	0,92694	0,89984	0,95673	0,98397	0,97274
	0,91835	0,94275	0,99107	0,97551	0,92724	0,95166
	0,71052	0,93063	0,84434	0,77634	0,57038	0,77687
2^{10}	0,45607	0,84737	0,72525	0,58036	0,69367	0,87368
	0,77594	0,89333	0,48504	0,67542	0,67917	0,40524
	0,95293	0,62228	0,95303	0,66459	0,90612	0,58128
	0,82396	0,79172	0,70320	0,62659	0,87402	0,51887
	0,82001	0,74581	0,73692	0,58116	0,91368	0,50919
	0,94877	0,93274	0,84366	0,88184	0,89422	0,77824
2^{20}	0,72963	0,92756	0,87310	0,66245	0,85253	0,80209
	0,86619	0,83357	0,70098	0,70497	0,82937	0,55243
	0,92732	0,77451	0,82577	0,70566	0,89742	0,61242
	0,90042	0,85590	0,66578	0,75905	0,75898	0,53951
	0,80609	0,95821	0,65451	0,76580	0,83988	0,61713
	0,43440	0,76488	0,47110	0,62937	0,95136	0,67330

Wyniki testów dla wariancji są bardzo dobre, średnia prawdopodobieństw wynosi 0,77, a najniższe z nich to 0,14369. Nie zaobserwowaliśmy żadnych przekroczeń przyjętych progów.

Tabela IV.3. P_{val} drugiego poziomu uzyskane w teście zgodności Pearsona dla ciągów różnej długości

	AES i Trivium	AES i RC4	AES i Rabbit	Trivium i RC4	Trivium i Rabbit	RC4 i Rabbit
2^3	0,76651 0,19769 0,97285 0,63106 0,14661 0,47737	0,65561 0,57488 0,42153 0,07829 0,35191 0,11978	0,57993 0,77767 0,03735 0,74237 0,02853 0,01795	0,81626 0,17673 0,77743 0,13108 0,98645 0,53328	0,99392 0,17352 0,82658 0,92099 0,93592 0,63190	0,97632 0,83218 0,48458 0,87129 0,66730 0,20360
2^7	0,90470 0,54413 0,76937 0,57403 0,41481 0,99546	0,79457 0,99752 0,96379 0,89578 0,91234 0,57320	0,28892 0,01299 0,38452 0,54098 0,18507 0,99929	0,35533 0,62021 0,05388 0,41119 0,81614 0,28302	0,73687 0,05298 0,37361 0,89447 0,03440 0,54609	0,50963 0,30333 0,80281 0,17045 0,89671 0,94645
2^{10}	0,65593 0,49260 0,79684 0,36834 0,73454 0,31422	0,00013 0,88297 0,50144 0,22603 0,70216 0,94103	0,00244 0,99142 0,21875 0,07992 0,37664 0,98346	0,38432 0,98059 0,86149 0,84251 0,17525 0,70634	0,06570 0,64161 0,83691 0,82706 0,18530 0,94051	0,05546 0,99337 0,20912 0,01538 0,01189 0,86922
2^{20}	0,02220 0,29224 0,48940 0,96445 0,19161 0,24142	0,51301 0,14075 0,00726 0,80541 0,01028 0,29965	0,18086 0,04157 0,47492 0,85214 0,37904 0,42037	0,00481 0,28859 0,21193 0,52557 0,29646 0,15358	0,92006 0,04575 0,98832 0,91039 0,94704 0,86594	0,36760 0,64159 0,37756 0,69909 0,64616 0,53299

W teście zgodności Pearsona dla dwóch próbek uzyskaliśmy najslabsze wyniki – w dwóch przypadkach – porównania AES’a i Rabbit’a oraz RC4 i Rabbita po trzy p_{val} są niższe od 0,1. Takiemu przypadkowi odpowiada p_{val} trzeciego poziomu równe 0,01585, co nie jest krytycznie małą wartością, co więcej, łącznie takich przekroczeń jest 21, co daje, z kolei, p_{val} równe 0,05067. Cztery przekroczenia progu 0,01 dają p_{val} trzeciego poziomu równe 0,05741.

Dla poszczególnych generatorów dostajemy:

- AES – 10 przekroczeń – $p_{val} = 0,17992$,
- Trivium – 8 przekroczeń – $p_{val} = 0,43253$,
- RC4 – 10 przekroczeń – $p_{val} = 0,17992$,
- Rabbit – 14 przekroczeń – $p_{val} = 0,01122$.

W teście ANOVA uzyskaliśmy bardzo dobre wyniki, wskazujące na brak istotnych odchyleń w wartościach oczekiwanych sekwencji p_{val} dla poszczególnych generatorów. Zostały one zebrane w tabeli IV.4.

Tabela IV.4. P_{val} drugiego poziomu uzyskane w teście ANOVA dla ciągów różnej długości

	Kołmogorowa -Smirnowa +	Kołmogorowa -Smirnowa –	Kołmogorowa	Andersona -Darlinga	Cramera -von Misesa	Kuipera
2^3	0,86912	0,47996	0,80025	0,48235	0,58528	0,86340
2^7	0,03610	0,76696	0,36877	0,64895	0,58057	0,37082
2^{10}	0,45912	0,91623	0,73247	0,27636	0,36311	0,70911
2^{20}	0,65524	0,29372	0,87568	0,76992	0,77826	0,26870

Również wyniki testu chi-kwadrat Bartletta są bardzo dobre. Wskazują na brak istotnych odchyleń w wariancjach sekwencji p_{val} dla poszczególnych generatorów. Zostały one zebrane w tabeli IV.5.

Tabela IV.5. P_{val} drugiego poziomu uzyskane w teście chi-kwadrat Bartletta dla ciągów różnej długości

	Kołmogorowa -Smirnowa +	Kołmogorowa -Smirnowa –	Kołmogorowa	Andersona -Darlinga	Cramera -von Misesa	Kuipera
2^3	0,29476	0,42538	0,47369	0,16492	0,28665	0,23764
2^7	0,66734	0,49255	0,99630	0,80927	0,64562	0,57403
2^{10}	0,16905	0,34370	0,83248	0,72874	0,53299	0,98582
2^{20}	0,94181	0,25137	0,66672	0,47014	0,84598	0,23929

Na podstawie zaprezentowanych powyżej wyników stwierdzamy, że żaden z testów nie przejawia zdolności do rozróżnienia zastosowanych generatorów. W związku z tym oraz mając świadomość możliwości popełnienia błędu drugiego rodzaju, w dalszej części badań wykorzystywać będziemy, ale bardzo ostrożnie, wszystkie te generatory.

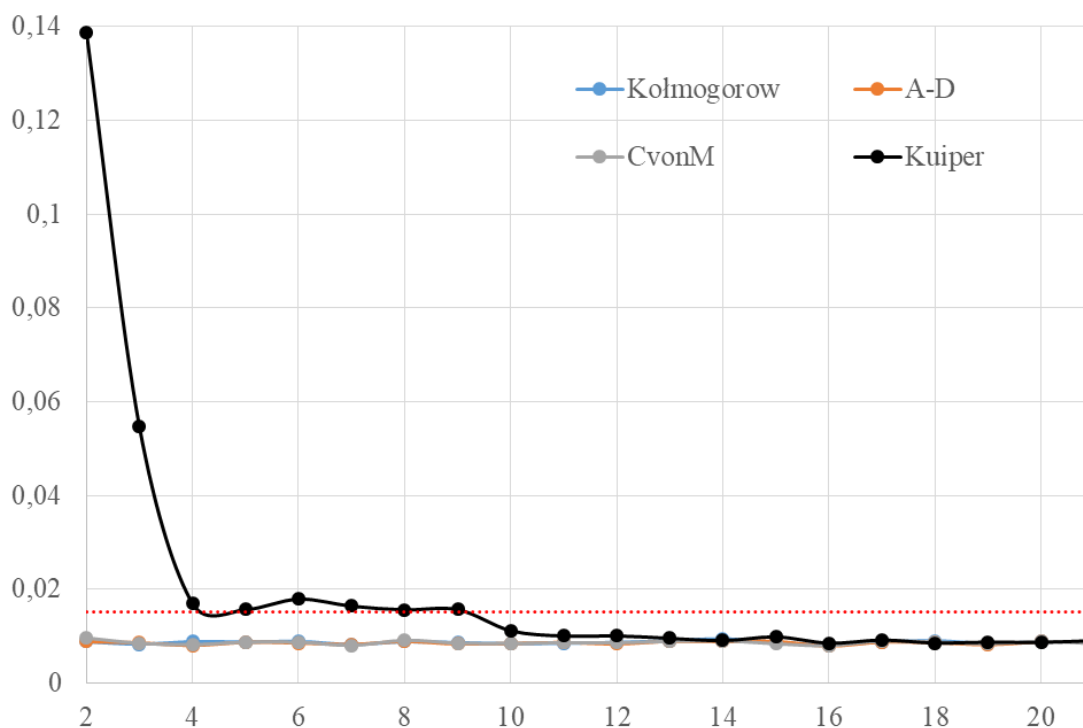
4. Weryfikacja testów

Definicja idealnej pary, zaproponowana w punkcie I.3.4., posiada pewną słabość – przy niskiej zdolności detekcyjnej testu, nie będzie on w stanie wykryć swojej własnej słabości, co umożliwi jego dowolne iterowanie. Potrzebujemy więc zaufanego testu, aby wyłonić zaufany test. Spośród pięciu testów wymienionych na początku rozdziału, do takiej roli najlepszy wydaje się test Kołmogorowa-Smirnowa, któremu, dla poprawy wiarygodności, poddana zostanie próbka o bardzo dużej objętości.

Dla każdej z długości badanego ciągu, od 2^2 do 2^{21} , wygenerowaliśmy 128 powtórzeń po 10 000 ciągów liczb rzeczywistych w reprezentacji 64-bitowej, które poddaliśmy badaniu przy użyciu rozpatrywanych testów. W ten sposób wytworzyliśmy dla 20 przypadków po 128 ciągów 10 000 liczb. Dla każdego z tych ciągów wyznaczona została statystyka Kołmogorowa $D_n = \max\{D_n^+, D_n^-\}$, a dla 128 powtórzeń zebraliśmy, dla każdego z testów i każdej z długości badanego ciągu z osobna:

- maksymalną wartość odchylenia D_n ,
- średnią wartość odchylenia D_n ,
- kwantyl z próby rzędu 0,9 zbioru odchyleń D_n ,
- kwantyl z próby rzędu 0,1 zbioru odchyleń D_n .

Na rysunku IV.1 prezentujemy zależność średniego odchylenia D_n od długości badanego ciągu n . Czerwona, przerywana linia, oznacza poziom odchylenia dopuszczalny dla statystyk K^+ i K^- na poziomie istotności 0,99. Oś odciętych ma skalę logarytmiczną – liczby w opisie należy traktować jako wykładniki potęgi 2.



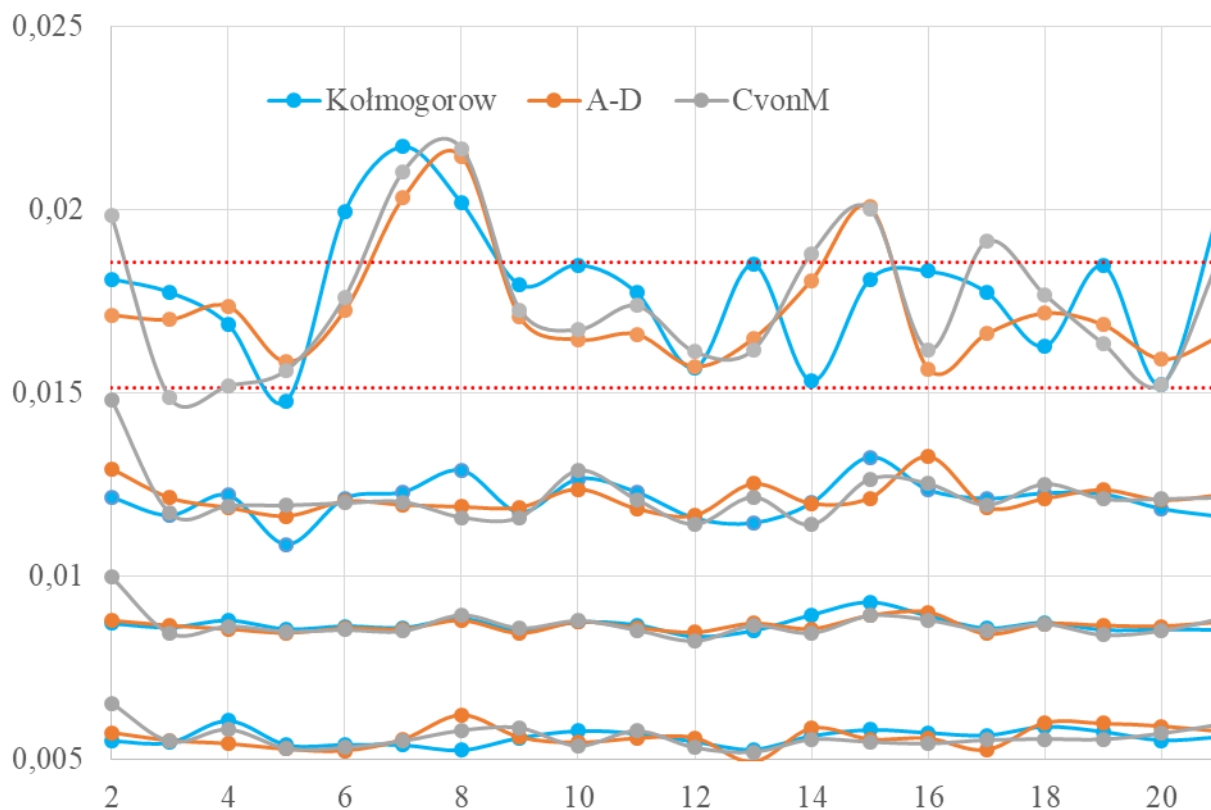
Rysunek IV.1. Średnie odchylenie dystrybuanty empirycznej w funkcji długości badanego ciągu

Z rozważanej czwórki testów, zdecydowanie negatywnie, wyróżnia się test Kuipera, który dopiero dla ciągów długości 2^{10} daje przeciętne p_{val} mieszczące się w założonym przedziale. Uznajemy test Kuipera za zbyt niedokładny dla małych długości ciągu.

Dla pozostałych trzech testów sporządziliśmy dodatkowe wykresy, prezentowane na rysunku IV.2, zestawiającym wszystkie cztery wyznaczone miary – w kolejności od góry:

- maksymalną wartość odchylenia,
- kwantyl z próby rzędu 0,9,
- średnią wartość odchylenia,
- kwantyl z próby rzędu 0,1.

Dodatkowo umieściliśmy jeszcze dwie linie oznaczające poziomy odchylenia dopuszczalne dla statystyk K^+ i K^- na poziomach istotności 0,999 i 0,99.



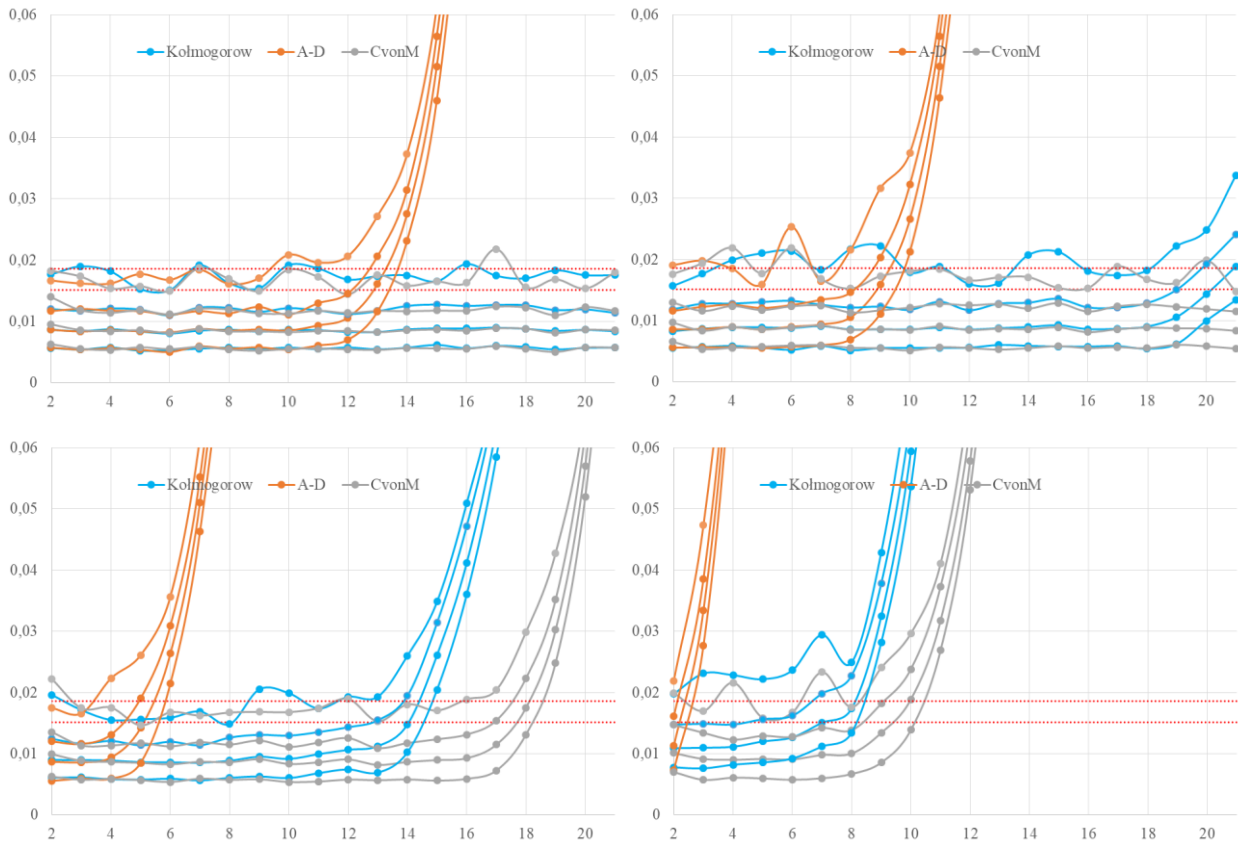
Rysunek IV.2. Średnie i maksymalne odchylenia dystrybucyj empirycznej oraz kwantyle empiryczne w funkcji długości badanego ciągu

Z powyższych wykresów wynika, że:

- badane testy zachowują się niemal identycznie, bez możliwości wskazania wyraźnego lidera,
- wartości średnie oraz kwantyle rzędu 0,9 mieszczą się poniżej progu dla poziomu istotności 0,99,
- zaobserwowane wartości maksymalne przekraczają zwykle próg dla poziomu istotności 0,99 (czego można oczekiwać po 128 próbach) i w większości nie przekraczają progu dla poziomu istotności 0,999.

Opisane powyżej badanie powtórzyliśmy jeszcze kilkakrotnie, przy czym zmianie ulegała precyzja, z jaką losowane były liczby rzeczywiste tworzące badane ciągi – rozważyliśmy przypadki 32, 20, 16, 12 i 8-bitowe. Spośród nich jedynie przypadek 32-bitowy nie wykazywał istotnych różnic w porównaniu do 64-bitowego – nie był odróżniany przez żaden z testów nawet dla ciągów o długości 2^{21} elementów.

Na rysunku IV.3 prezentujemy przebiegi analogiczne do rysunku IV.2 dla pozostałych przypadków.



Rysunek IV.3. Średnie i maksymalne odchylenia dystrybucyj empirycznych oraz kwantyle empiryczne w funkcji długości badanego ciągu dla precyzji 20, 16, 12 i 8 bitowej

Na podstawie powyższych wykresów stwierdzamy, że:

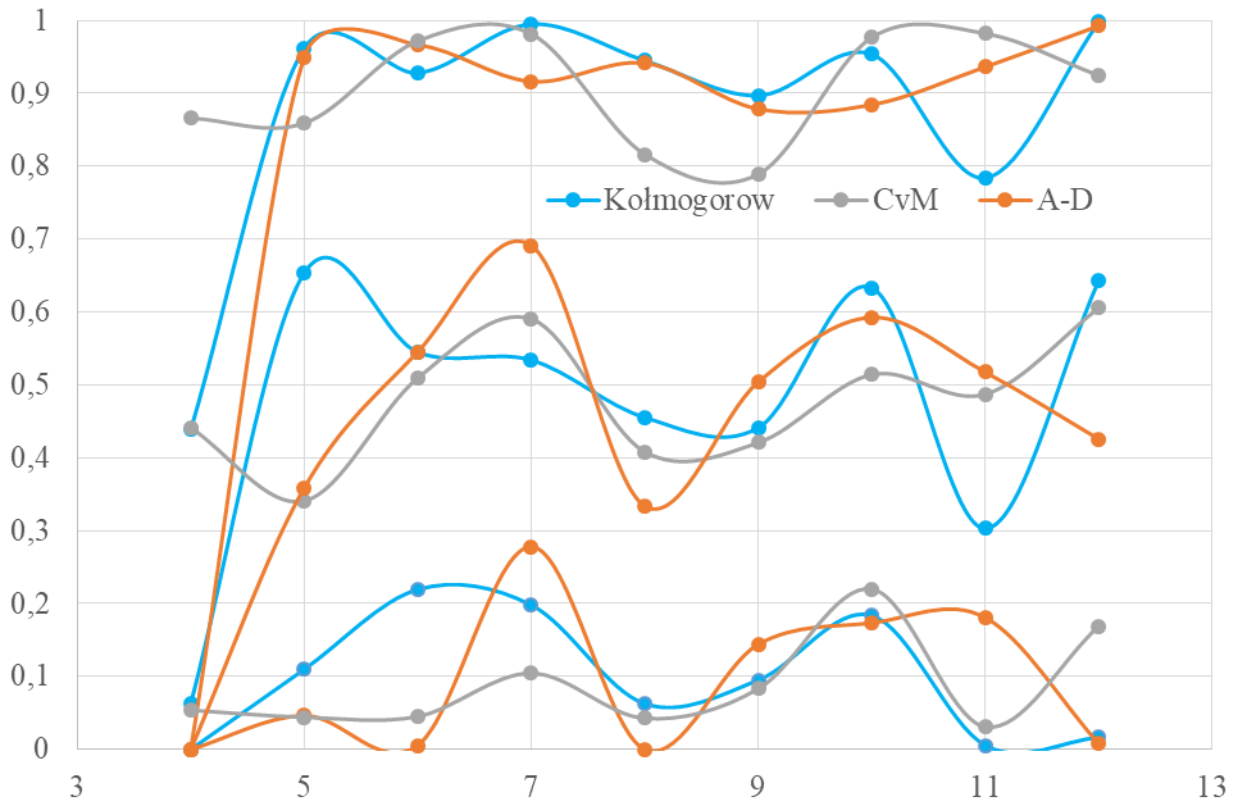
- każdy z testów posiada zdolność detekcji niedostatecznej precyzji reprezentacji liczb rzeczywistych tworzących badany ciąg,
- najlepszym pod tym względem wydaje się być test Andersona-Darlinga,
- najslabszym zaś test Cramera von Misesa.

Wybrany w tej części test posłużył nam jako test drugiego poziomu (a nierzadko trzeciego i czwartego) w budowanym pakiecie, stąd wymóg „idealności” testu. W celu weryfikacji tej cechy wykonaliśmy pięciopoziomową procedurę, w której:

- na poziomie pierwszym losowane było, z precyzją 64-bitową, 2^d dla $d = 4..12$ liczb z przedziału $[0;1]$ i ciąg taki oceniany był przy użyciu rozpatrywanych testów, dostarczając p_{val} pierwszego poziomu,
- na poziomie drugim, po 1024-krotnym powtórzeniu eksperymentu pierwszego poziomu, rozkład uzyskanych p_{val} pierwszego poziomu oceniany był przy użyciu rozpatrywanych testów (każdy odpowiednio oceniał tylko „swoje” wyniki), dostarczając p_{val} drugiego poziomu,
- na poziomie trzecim, po 1024-krotnym powtórzeniu eksperymentu drugiego poziomu, rozkład uzyskanych p_{val} drugiego poziomu oceniany był przy użyciu rozpatrywanych testów (każdy odpowiednio oceniał tylko „swoje” wyniki), dostarczając p_{val} trzeciego poziomu,

- na poziomie czwartym, po 1024-krotnym powtórzeniu eksperymentu trzeciego poziomu, rozkład uzyskanych p_{val} trzeciego oceniany był przy użyciu rozpatrywanych testów (każdy odpowiednio oceniał tylko „swoje” wyniki), dostarczając p_{val} czwartego poziomu,
- na poziomie piątym, po dziesięciokrotnym powtórzeniu eksperymentu czwartego poziomu, wykonaliśmy analizę przedstawioną poniżej.

Na rysunku IV.4 prezentujemy przebiegi trzech wyznaczanych miar – dla każdego z testów, dla każdej z długości bloku, zaznaczono maksymalne, średnie i minimalne uzyskane p_{val} .



Rysunek IV.4. Minimalne, średnie i maksymalne wartości p_{val} uzyskane po dziesięciokrotnym powtórzeniu czteropoziomowej procedury dla testów Kołmogorowa, Andersona-Darlinga i Cramera-von Misesa

Etap ten pokazał, że nawet przy miliardkrotnym wykorzystaniu testów Kołmogorowa, Andersona-Darlinga, czy też Cramera-von Misesa, nie uwidaczniają się żadne słabości i wszystkie trzy testy mogą być z powodzeniem stosowane w roli testu drugiego-trzeciego-czwartego-... poziomu. Z uwagi na lepsze własności detekcyjne, wykazane we wcześniejszej części, zdecydowaliśmy się, w zasadniczej części tej pracy, korzystać z testu Andersona-Darlinga.

W ramach kolejnej części rozpatrywać będziemy ponownie wszystkie powyższe testy, tym razem w roli testów losowości pierwszego poziomu i przy tej okazji zbadamy skorelowanie uzyskiwanych przy ich użyciu p_{val} .

Rozdział V

Określenie przydatności procedur testowych do badania prymitywów kryptograficznych

Rozdział ten poświęciliśmy omówieniu wyników praktycznego badania procedur testowych możliwych do zrealizowania dla ciągów binarnych o długościach 1 kiB i 1 Mb. Jedyne pierwsza z podanych długości wydaje się wymagać uzasadnienia – jest ona kompromisem pomiędzy możliwie małą długością ciągu, a możliwością uzyskania jakichkolwiek rozsądnych wyników.

W kolejnych punktach wskazujemy, które z zaimplementowanych procedur testowych mogą znaleźć zastosowanie podczas wykonywania badania dla różnych liczb powtórzeń. W kolejnym rozdziale analizujemy jedynie te procedury, które przeszły najgęstsze sito eliminacji.

1. Procedura

Dla obu długości badanych ciągów wykonywaliśmy 4 kroki:

1. Weryfikacja rozróżnialności generatorów.
2. Weryfikacja przydatności procedury testowej do badania pojedynczych ciągów.
3. Weryfikacja przydatności procedury testowej do badania generatorów przy małej liczbie powtórzeń.
4. Weryfikacja przydatności procedury testowej do badania generatorów przy dużej liczbie powtórzeń.

1.1. Weryfikacja rozróżnialności generatorów

Krok pierwszy warunkował możliwość przeprowadzenia trzech pozostałych, ze względu na ogromną liczbę procedur testowych został przez nas zautomatyzowany i ograniczony jedynie do testów ANOVA, chi-kwadrat Bartletta, omówionych już w punkcie IV.3, oraz testu Browna-Forsythe’a, ze względu na jego odporność na niezgodność badanych próbek z rozkładem normalnym.

W teście Browna-Forsythe’a wyznaczana jest statystyka:

$$BF_{k,n} = \frac{kn(n-1) \sum_{i=1}^k (\bar{z}^i - \bar{z})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^n (z_j^i - \bar{z}^i)^2},$$

gdzie

k – jest liczbą generatorów,

n – jest liczbą p_{val} dla pojedynczego generatora,

x_j^i – jest j -tym p_{val} dla i -tego generatora,

$\tilde{x}^i$ – jest medianą p_{val} dla i -tego generatora,

$$z_j^i = |x_j^i - \tilde{x}^i|,$$

$\bar{z}^i$ – jest średnią z_j^i dla i -tego generatora,

$\bar{z}$ – jest średnią z_j^i dla wszystkich generatorów.

$BF_{k,n}$ ma rozkład Fischera-Snedekora z $(k-1, k(n-1))$ stopniami swobody.^{173,174}

Dla każdego z czterech generatorów oraz pobitowego xor'a ich wyjść wygenerowane zostały po 1024 ciągi, a uzyskane wektory p_{val} poddane zostały ocenie zgodności rozkładów. Testom poddawane były tylko te zestawy, w których próbki dla wszystkich generatorów wykazywały wariancje wyników większe od 10^{-10} . W pierwszej kolejności brane pod uwagę były testy ANOVA i Browna-Forsythe'a, a w przypadku negatywnych wyników tego drugiego poszukiwano ich potwierdzenia w teście Bartletta.

Prezentowane dalej wyniki bazują na pięciokrotnym powtórzeniu opisanej procedury. Przyjęto poziom istotności równy 0,999.

1.2. Weryfikacja przydatności procedury testowej do badania pojedynczych ciągów

Przyjęliśmy, że procedurę testową będziemy kwalifikować jako przydatną do badania pojedynczych ciągów, jeśli trójpoziomowa procedura sprawdzająca, w której:

- poziom pierwszy polegał na wykonaniu badanej procedury dla **128** ciągów będących pobitowymi xor'ami ciągów wygenerowanych przy użyciu wykorzystanych generatorów,
- poziom drugi polegał na 1024-krotnym wykonaniu testu zgodności z rozkładem równomiernym na przedziale $[0;1]$ p_{val} uzyskanych dla badanej procedury pierwszego poziomu,
- poziom trzeci polegał na wykonaniu testu zgodności z rozkładem równomiernym na przedziale $[0;1]$ p_{val} uzyskanych dla procedury drugiego poziomu,

kończyła się pozytywnie na poziomie istotności 0,999999. Dla przypadków wątpliwych stosowaliśmy dopuszczenie do kroku 3.

1.3. Weryfikacja przydatności procedury testowej do badania generatorów przy małej liczbie powtórzeń

W tym kroku rozpatrywane były tylko procedury zakwalifikowane jako przydatne do badania pojedynczych ciągów.

Przyjęliśmy, że procedurę testową będziemy kwalifikować jako przydatną do badania generatorów przy małej liczbie powtórzeń, jeśli trójpoziomowa procedura sprawdzająca, w której:

- poziom pierwszy polegał na wykonaniu badanej procedury dla **1024** ciągów będących pobitowymi xor'ami ciągów wygenerowanych przy użyciu wykorzystanych generatorów,

¹⁷³ M.B. Brown, A.B. Forsythe, *Robust tests for the equality of variances*, Journal of the American Statistical Association, 69 (346), 1974, s. 364–367

¹⁷⁴ www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm

- poziom drugi polegał na 1024-krotnym wykonaniu testu zgodności z rozkładem równomiernym na przedziale $[0;1]$ p_{val} uzyskanych dla badanej procedury pierwszego poziomu,
- poziom trzeci polegał na wykonaniu testu zgodności z rozkładem równomiernym na przedziale $[0;1]$ p_{val} uzyskanych dla procedury drugiego poziomu,

kończyła się pozytywnie na poziomie istotności 0,999999. Dla przypadków wątpliwych stosowaliśmy dopuszczenie do kroku 4.

1.4. Weryfikacja przydatności procedury testowej do badania generatorów przy dużej liczbie powtórzeń

W tym kroku rozpatrywane były tylko procedury zakwalifikowane jako przydatne do badania generatorów przy małej liczbie powtórzeń.

Przyjęliśmy, że procedurę testową będziemy kwalifikować jako przydatną do badania generatorów przy dużej liczbie powtórzeń, jeśli trójpoziomowa procedura sprawdzająca, w której:

- poziom pierwszy polegał na wykonaniu badanej procedury dla **8192** ciągów będących pobitowymi xor'ami ciągów wygenerowanych przy użyciu wykorzystanych generatorów,
- poziom drugi polegał na 1024-krotnym wykonaniu testu zgodności z rozkładem równomiernym na przedziale $[0;1]$ p_{val} uzyskanych dla badanej procedury pierwszego poziomu,
- poziom trzeci polegał na wykonaniu testu zgodności z rozkładem równomiernym na przedziale $[0;1]$ p_{val} uzyskanych dla procedury drugiego poziomu,

kończyła się pozytywnie na poziomie istotności 0,999.

Dla przypadków wątpliwych wykonywaliśmy dodatkowe cztery powtórzenia tego kroku, przyjmując ostateczny werdykt na zasadzie większości wskazań. Taka procedura zapewnia prawdopodobieństwo przypadkowego odrzucenia badanej procedury testowej mniejsze niż 10^{-8} .

2. Procedury dla ciągów długości 1 kiB

2.1. Zestaw początkowy

Dla ciągów o długości 1 kiB, przygotowane przez nas oprogramowanie składało się z wywołania 36 584 procedur wywiedzionych z 77 testów. Włączone zostały również procedury, w których przydatność wątpiliśmy od samego początku, woleliśmy jednak wyeliminować możliwość wystąpienia swoistego błędu drugiego rodzaju.

Pełne zestawienie wszystkich procedur testowych rozważanych dla tej długości ciągu zawarliśmy w załączniku A.

2.2. Weryfikacja rozróżnialności generatorów

Przeprowadzone, zgodnie z procedurą z punktu V.1.1., badanie nie wskazało na możliwość odróżnienia, przez którąkolwiek z procedur któregoś ze źródeł.

13 889 procedur zwróciło jednorodne lub niemal jednorodne wektory p_{val} , co oznacza, że zostały nadmiarowo wywołane dla tej długości ciągu – z pewnością okażą się nieprzydatne nawet do badania pojedynczych ciągów.

Spośród 22 881 pozostałych:

- w teście ANOVA:
 - 138 uzyskało wynik negatywny w jednym z pięciu powtórzeń procedury,
- w teście Bartletta:
 - 228 uzyskało wynik negatywny w jednym z pięciu powtórzeń procedury,
 - 125 uzyskało wynik negatywny w dwóch z pięciu powtórzeń procedury,
 - 196 uzyskało wynik negatywny w trzech z pięciu powtórzeń procedury,
 - 80 uzyskało wynik negatywny w czterech z pięciu powtórzeń procedury,
 - 316 uzyskało wynik negatywny we wszystkich pięciu powtórzeniach procedury,
- w teście Browna-Forsythe’a:
 - 83 uzyskało wynik negatywny w jednym z pięciu powtórzeń procedury.

Wyniki testu Bartletta wynikają z braku równomierności uzyskiwanych dla tych procedur p_{val} , do czego wrócimy w kolejnym punkcie.

Testy ANOVA i Browna-Forsythe’a dały dużo większe, od spodziewanych, liczby odrzuceń, na co wpływ miały małe wariancje. Jedynie dla jednej procedury (drugiej próby testu losowych wycieczek, dla podciągów długości 194 bitów – momentu pierwszego osiągnięcia wartości $y = 8$) oba testy dały wynik negatywny.

Wobec powyższego stwierdzamy, że żadna z rozpatrywanych procedur nie posiada zdolności rozróżniania użytych generatorów, a wymieniony powyżej przypadek potraktowaliśmy jako realizację błędu pierwszego rodzaju.

2.3. Weryfikacja przydatności do badania pojedynczych ciągów długości 1 kiB

Badanie przeprowadzono zgodnie z procedurą opisaną w punkcie V.1.2. Do badania pojedynczych ciągów zakwalifikowały się 20 133 procedury, wywiedzione z 64 testów. Stanowi to odpowiednio 55% i 83% początkowego wolumenu.

W tej fazie odrzucone zostały wszystkie procedury pochodzące z następujących testów:

- częstości serii 0 i 1 osobno zliczanych,
- dwupoziomowego Bickela-Breimana,
- kolizji,
- korelacji wag Hamminga dla bloku M -bitowego,
- liczby komórek z co najmniej b punktami dla wektorów,
- liczby skoków w profilu liniowym dla ciągu dzielonego na fragmenty m -bitowe,
- logarytm prawdopodobieństwa dla nakładających się wektorów,

- losowych wycieczek dla podciągów – momentu pierwszego osiągnięcia wartości y ,
- odstępów dni urodzin,
- rozkładu skoków w profilu liniowych,
- serii malejących z macierzą kowariancji,
- serii rosnących z macierzą kowariancji,
- zgodności bitowej.

Wszystkie pierwotnie rozważane procedury zachowały 32 testy.

Pełne zestawienie wszystkich procedur testowych dla ciągów długości 1 kiB zakwalifikowanych do badania pojedynczych ciągów zawarliśmy w punkcie 2 załącznika A.

2.4. Weryfikacja przydatności do badania generatorów przy małej liczbie powtórzeń

Badanie przeprowadzono zgodnie z procedurą opisaną w punkcie V.1.3. Do badania generatorów przy małej liczbie powtórzeń zakwalifikowały się 18 622 procedury, wywiedzione z 48 testów. Stanowi to odpowiednio 51% i 62% początkowego wolumenu.

W tej fazie do zbioru testów, których wszystkie procedury zostały odrzucone, dołączyły testy:

- całkowitej liczby serii,
- częstości serii 0 i 1 łącznie,
- iloczynów,
- Kuipera,
- liczby komórek z dokładnie b punktami dla wektorów,
- liczby skoków w profilu liniowym,
- losowych wycieczek dla podciągów długości M bitów – liczby przejść przez 0 ze zmianą strony,
- pochodnej ciągu bloków,
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty m -bitowe,
- sumy kwadratów odstępów m -elementowych bloków d -bitowych, z zapętlaniem,
- sumy logarytmów odstępów m -elementowych bloków d -bitowych, bez zapętlenia,
- sumy logarytmów odstępów m -elementowych bloków d -bitowych, z zapętlaniem.

Wszystkie procedury zachowało już jedynie 13 testów.

Pełne zestawienie wszystkich procedur testowych dla ciągów długości 1 kiB zakwalifikowanych do badania generatorów przy małej liczbie powtórzeń zawarliśmy w punkcie 3 załącznika A.

2.5. Weryfikacja przydatności do badania generatorów przy dużej liczbie powtórzeń

Badanie przeprowadzone zgodnie z procedurą opisaną w punkcie V.1.4. Do badania generatorów przy dużej liczbie powtórzeń zakwalifikowało się 13 417 procedur, wywiedzionych z 36 testów. Stanowi to odpowiednio 37% i 47% początkowego wolumenu.

W tej fazie do zbioru testów, których wszystkie procedury zostały odrzucone, dołączyły:

- autokorelacji dla ciągu binarnego,
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- Corona,
- częstości dla nakładających się wektorów,
- kolizji indeksów maksymalnych elementów,
- losowych wycieczek dla podciągów długości M bitów,
 - liczby powrotów do 0,
 - łącznego „czasu” odchylenia w prawo,
 - maksymalnego odchylenia ścieżki,
- nakładających się t -bitowych wzorców 0, 1 w bloku M -bitowym,
- niezależności wag Hamminga – wariant 2,
- sumy kwadratów modułów DFT dla podciągów 2^d -bitowych,
- sumy kwadratów odstępów m -elementowych bloków d -bitowych, bez zapętlania.

Wszystkie procedury zachowało już jedynie 5 testów.

Pełne zestawienie wszystkich procedur testowych dla ciągów długości 1 kiB zakwalifikowanych do badania generatorów przy dużej liczbie powtórzeń zawarliśmy w punkcie 4 załącznika A.

2.6. Podsumowanie eliminacji procedur dla ciągów 1 kiB

W punkcie V.2.1 wymieniliśmy 77 testów i 36 584 procedury z nich wywiedzione, przy czym 45% spośród nich stanowiły procedury testów autokorelacyjnego dla ciągu bloków i ciągu bitów. Zgodnie z przyjętą metodologią odrzuciliśmy ponad 63% procedur i 53% testów, przy czym obraz ten znacząco poprawia test autokorelacyjny dla ciągu bloków. Gdyby obydwa testy autokorelacyjne wykluczyć z powyższych statystyk, to okazałoby się, że ponad 95% pierwotnie rozważanych procedur testowych przeznaczonych dla ciągów o długości 1 kiB nie powinno być używane do badania generatorów przy dużej liczbie powtórzeń.

3. Procedury dla ciągów długości 1 Mb

3.1. Zestaw początkowy

Dla ciągów o długości 1 Mb, przygotowane przez nas oprogramowanie składało się z wywołania 604 869 procedur testowych wywiedzionych z 76 testów.

Pełne zestawienie wszystkich procedur testowych rozważanych dla tej długości ciągu zawarliśmy w załączniku B.

3.2. Weryfikacja rozróżnialności generatorów

Przeprowadzone, zgodnie z procedurą z punktu V.1.1., badanie nie wskazało na możliwość odróżnienia, przez którąkolwiek z procedur któregoś ze źródeł.

4817 procedur zwróciło jednorodne lub niemal jednorodne wektory p_{val} , co oznacza, że zostały nadmiarowo wywołane dla tej długości ciągu – z pewnością okażą się nieprzydatne nawet do badania pojedynczych ciągów.

Spośród 600 052 pozostałych procedur:

- w teście ANOVA:
 - 2 uzyskało wynik negatywny w dwóch z pięciu powtórzeń procedury,
 - 2763 uzyskało wynik negatywny w jednym z pięciu powtórzeń procedury,
- w teście Bartletta:
 - 29606 uzyskało wynik negatywny w jednym z pięciu powtórzeń procedury,
 - 3511 uzyskało wynik negatywny w dwóch z pięciu powtórzeń procedury,
 - 460 uzyskało wynik negatywny w trzech z pięciu powtórzeń procedury,
 - 8732 uzyskało wynik negatywny w czterech z pięciu powtórzeń procedury,
 - 279 uzyskało wynik negatywny we wszystkich pięciu powtórzeniach procedury,
- w teście Browna-Forsythe’a:
 - 4 uzyskało wynik negatywny w dwóch z pięciu powtórzeń procedury,
 - 2546 uzyskało wynik negatywny w jednym z pięciu powtórzeń procedury.

Również dla ciągu 1 Mb testy ANOVA i Browna-Forsythe’a dały większe, od spodziewanych, liczby odrzuceń. Dzięki dużo większej próbie (liczbie procedur testowych) niż dla ciągów o długości 1 kiB, mogliśmy ocenić te wyniki również poprzez ich wewnętrzną spójność.

W przypadku testu ANOVA zaobserwowaliśmy łącznie 2767 sygnałów, co daje częstość obserwacji wynoszącą 0,000922, dla testu Browna-Forsythe’a mamy odpowiednio – 2554 sygnały i częstość wynoszącą 0,000851. Przy założeniu, że poszczególne obserwacje są realizacjami niezależnych eksperymentów, wartości oczekiwane liczby procedur, dla których w dwóch powtórzeniach otrzymamy sygnał wynoszą: 5,09 dla testu ANOVA i 4,34 dla testu Browna-Forsythe’a, w rzeczywistości zaobserwowaliśmy, odpowiednio, 2 i 4 takie przypadki. Co więcej, wartość oczekiwana liczby procedur, dla których w co najmniej jednym powtórzeniu oba testy sygnalizować będą błąd wynosi 11,74, zaobserwowaliśmy zaś 9 takich przypadków.

Powyższe pozwala przypuszczać, że obserwujemy jedynie błędy pierwszego rodzaju, zaś za podwyższoną liczbę sygnalizacji odpowiada niedokładne określenie rozkładu statystyk obu tych testów, podstawowy problem, z którym borykamy się w tej pracy.

Stwierdzamy, że żadna z rozpatrywanych procedur nie posiada zdolności rozróżniania użytych generatorów.

3.3. Weryfikacja przydatności do badania pojedynczych ciągów długości 1 Mb

Badanie przeprowadzono zgodnie z procedurą opisaną w punkcie V.1.2. Do badania pojedynczych ciągów zakwalifikowały się 561 853 procedury, wywiedzione z 68 testów. Stanowi to odpowiednio 93% i 89% początkowej ich liczby.

W tej fazie odrzucone zostały wszystkie procedury pochodzące z następujących testów:

- dwupoziomowego Bickela-Breimana,
- korelacji wag Hamminga dla bloku M -bitowego,
- liczby skoków w profilu liniowym dla ciągu dzielonego na fragmenty m -bitowe,
- logarytmu prawdopodobieństwa dla nakładających się wektorów,
- losowych wycieczek dla podciągów długości M bitów – momentu pierwszego osiągnięcia wartości y ,
- rozkładu skoków w profilu liniowym,
- serii rosnących / malejących z macierzą kowariancji,
- zgodności bitowej.

Wszystkie procedury zachowało 25 testów.

Pełne zestawienie wszystkich procedur testowych dla ciągów długości 1 Mb zakwalifikowanych do badania pojedynczych ciągów zawarliśmy w punkcie 2 załącznika B.

3.4. Weryfikacja przydatności do badania generatorów przy małej liczbie powtórzeń

Badanie przeprowadzone zgodnie z procedurą opisaną w punkcie V.1.3. Do badania generatorów przy małej liczbie powtórzeń zakwalifikowały się 530 762 procedury, wywiedzione z 68 testów. Stanowi to odpowiednio 88% i 89% początkowego wolumenu.

Żaden test nie został odrzucony w całości w tej fazie, natomiast wszystkie procedury zachowały już jedynie 22 testy.

Pełne zestawienie wszystkich procedur testowych dla ciągów długości 1 Mb zakwalifikowanych do badania generatorów przy małej liczbie powtórzeń zawarliśmy w punkcie 3 załącznika B.

3.5. Weryfikacja przydatności do badania generatorów przy dużej liczbie powtórzeń

Badanie przeprowadzono zgodnie z procedurą opisaną w punkcie V.1.4. Do badania generatorów przy dużej liczbie powtórzeń zakwalifikowały się 528 763 procedury, wywiedzione z 62 testów. Stanowi to odpowiednio 87% i 82% początkowego wolumenu.

W tej fazie do zbioru testów, których wszystkie procedury zostały odrzucone, dołączyły testy:

- częstości dla nakładających się wektorów,
- kolizji,
- liczby komórek z co najmniej b punktami dla wektorów,
- liczby komórek z dokładnie b punktami dla wektorów,
- największej liczby elementów w przedziale – podciągi 1 kiB,
- odstępów dni urodzin.

Wszystkie procedury zachowało już jedynie 18 testów.

Pełne zestawienie wszystkich procedur testowych dla ciągów długości 1 Mb zakwalifikowanych do badania generatorów przy dużej liczbie powtórzeń zawarliśmy w punkcie 4 załącznika B.

3.6. Podsumowanie eliminacji procedur dla ciągów 1 Mb

W punkcie V.3.1 wymieniliśmy 80 testów i 604 869 procedur z nich wywiedzionych. Ze względu na identyczność rozkładów pominęliśmy 4 263 733 procedury testu nienakładających się wzorców i 2 010 835 procedur testów autokorelacyjnych. Uzyskane wyniki wskazują, że odrzucone zostałyby wszystkie pominięte procedury testu nakładających się wzorców i przyjęte wszystkie procedury testów autokorelacyjnych. Oznacza to, że odrzucone zostało 63% procedur dla ciągów 1 Mb.

3.7. Weryfikacja procedur testowych pakietu *Statistical Test Suite*

W rozdziale III opisaliśmy dość szczegółowo zawartość pakietu *Statistical Test Suite* i poczyniliśmy szereg krytycznych uwag dotyczących jakości określenia rozkładów statystyk testów w nim zaimplementowanych. Dysponując wynikami zawartymi w punkcie V.3.5 możemy pokusić się o określenie na ile poszczególne jego procedury mogą mieć zastosowanie do wykonywania badań generatorów przy dużej liczbie powtórzeń.

Przyjęliśmy, zgodnie z domyślnymi parametrami proponowanymi dla pakietu, że badaniu poddawane są ciągi o długości 10^6 bitów, a wynikiem każdego takiego badania jest zestaw 188 prawdopodobieństw, przy czym odpowiadające im procedury testowe oparte są o 15 testów losowości opisanych w punkcie III.5. W tabeli V.1 zebraliśmy wyniki tej analizy.

Aż 5 z 15 testów oceniamy jako nieprzydatne w badaniu dużych serii ciągów, przy czym należy zaznaczyć, że ocena bazuje na wynikach otrzymanych dla wykonanych w ramach tej pracy implementacji, nie zaś zawartości pakietu STS. W kilku przypadkach, gdy pakiet ten był przez nas używany, zmuszeni byliśmy do wykonywania badania porównawczego, w przeciwnym razie większość procedur zwracała wynik negatywny. Zauważone przez nas błędy opisaliśmy już w punkcie III.5.

Tabela V.1. Wyniki analizy przydatności procedur pakietu *Statistical Test Suite* do wykonywania badań generatorów przy dużej liczbie powtórzeń

test	procedury testowe	wynik analizy
częstości bitów	$d = 1$	pozytywny
częstości w bloku	$M = 128$	pozytywny
serii	–	pozytywny
najdłuższej serii jedynek w bloku	$M = 10000$	pozytywny
rzędu macierzy binarnej	$M = Q = 32$	brak danych
spektralny - DTF	–	negatywny
nienakładających się wzorców	$M = 10000, d = 9$	tylko dla pojedynczych ciągów
nakładających się wzorców	$M = 1032, d = 9$ 148 wzorców	negatywny

uniwersalny Maurera	$L = 7, Q = 1280, K = 141577$	pozytywny
złożoności liniowej	$M = 500$	pozytywny
częstości nakładających się bloków	$d = 16$	pozytywny
entropii	$m = 10$	pozytywny
sum	w przód, wstak	pozytywny
losowych wycieczek	$x \in \{-4, -3, -2, -1, 1, 2, 3, 4\}$	negatywny
alternatywny losowych wycieczek	$x \in \{-9, \dots, -1, 1, \dots, 9\}$	negatywny

4. Podsumowanie wyników eliminacji

Zastosowana przez nas procedura weryfikacji przydatności procedur testowych do stosowania dla różnych wolumenów badanych ciągów unaoczniała przyczyny problemów, z którymi stykaliśmy się wykorzystując popularne pakiety testów. Odrzucenie dla 53% testów wszystkich procedur dla ciągów o długości 1 kiB oraz 18% – dla ciągów 1 Mb, oznacza, że nawet przy gruntownym krytycznym przeglądzie literatury, poszerzonym o własne wyniki, istnieje bardzo duża szansa posłużenia się testem, który dla danej długości ciągu nie powinien być stosowany.

Dokumentacja dołączona do pakietu STS wskazuje, że dość sumienie wykonano część analityczną, jednakże brak analogicznego do wykonanego przez nas badania spowodował, że większość procedur testowych nie powinna była się w pakiecie znaleźć.

Rozdział VI

Skorelowanie procedur testowych

Jeśli badanie opisane w poprzednim rozdziale można określić jako poszukiwanie procedur testowych, których **można użyć** do badania mocy algorytmów kryptograficznych bez obawy o jakość wykonanego badania, to w tym rozdziale postaramy się pośród nich znaleźć te, które **można wyłączyć** z badania bez obawy o jego jakość. Poszukiwać będziemy przypadków gdy wartość prawdopodobieństwa testowego p_{val} dla danej procedury można wystarczająco dokładnie wyznaczyć na podstawie p_{val} innych procedur.

W kolejnych punktach omawiamy wyniki uzyskane w kolejnych krokach. Rozpoczynamy od weryfikacji istnienia skorelowania pomiędzy parami procedur, następnie wykorzystujemy obliczeniowo najprostszą korelację liniową jako narzędzie dalszej eliminacji nadmiarowych procedur. Następnie uogólniliśmy ją do testowania 11 modeli nieliniowych. W punkcie 3 przeanalizowaliśmy klasy, na które zbiór procedur dzieliła relacja skorelowania, co dało asumpt do wykonania, opisanego w punkcie 4, wielowymiarowej regresji liniowej.

1. Weryfikacja istotności korelacji

Pierwszym krokiem naszej analizy było wykorzystanie współczynnika korelacji Pearsona z próby:

$$\rho_{i,j} = \frac{\sum_{k=1}^n (u_k^i - \bar{u}^i)(u_k^j - \bar{u}^j)}{\sqrt{\sum_{k=1}^n (u_k^i - \bar{u}^i)^2} \sqrt{\sum_{k=1}^n (u_k^j - \bar{u}^j)^2}},$$

gdzie $\bar{u}^i = \frac{1}{n} \sum_{k=1}^n u_k^i$ jest średnią i -tej próbki, a n licznoscią próbki.

Niezerowa wartość współczynnika $\rho_{i,j}$ wskazywać będzie na zależność pomiędzy p_{val} otrzymywanymi dla różnych procedur testowych. Statystyczną istotność otrzymanej wartości oraz dokładniejszy charakter zależności rozpatrujemy w kolejnych punktach.

1.1. Procedura weryfikacji istotności korelacji

Ponieważ, nawet dla bardzo dużych próbek, wyznaczany współczynnik z próby $\rho_{i,j}$ niemal zawsze będzie różnić się od rzeczywistej jego wartości, potrzebna jest metoda oceny, czy obserwowana wartość istotnie odbiega od 0. W tym celu wykorzystaliśmy dwa testy:

- istotności współczynnika korelacji

Ponieważ test przeznaczony jest dla rozkładu normalnego, przed jego wykonaniem w miejsce poszczególnych p_{val} podstawione zostały odpowiadające im kwantyle standardowego rozkładu normalnego $\Phi^{-1}(p_{val})$.

Wyznaczana jest statystyka

$$RI(\rho) = \rho \sqrt{\frac{n-2}{1-\rho^2}},$$

gdzie ρ jest współczynnikiem otrzymanym dla pary n -elementowych próbek.

Zmienna losowa RI ma rozkład T-Studenta z $n-2$ stopniami swobody.^{175,176}

- test rangowy Spearman'a

Niech r_k^i i r_k^j będą numerami k -tej pary przy sortowaniu względem, odpowiednio, składnika u_k^i i u_k^j . Statystyka

$$RS = \frac{6 \sum_{k=1}^n (r_k^i - r_k^j)^2 - n(n^2 - 1)}{n(n+1)\sqrt{n-1}}$$

ma, dla dużych n , standardowy rozkład normalny.¹⁷⁷

Poniżej rozpatrywane były jedynie procedury, które przeszły pomyślnie wszystkie etapy weryfikacji równomierności rozkładu uzyskiwanych przy ich pomocy prawdopodobieństw p_{val} , stąd można przyjąć, że rozkład liczb $\Phi^{-1}(p_{val})$ dla danej procedury testowej odpowiadał realizacjom niezależnych zmiennych losowych o standardowym rozkładzie normalnym i spełnione były założenia testu istotności współczynnika korelacji.

1.2. Wyniki weryfikacji istotności korelacji dla procedur dla ciągów o długości 1 kiB

Badanie polegało na wyznaczeniu statystyk RI i RS dla wszystkich par procedur testowych wymienionych w punkcie V.2.5. W badaniu wykorzystane zostały wyniki zebrane dla 13 432 procedur i 1 000 000 ciągów.

Wyznaczając kwantyle statystyki RI :

$$P\left(\left|RI(\rho_\alpha)\right| > T_{n-2}^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = \alpha,$$

możemy uzmysłwić sobie jak bliski zera jest, dzięki dużej liczności próbki, nasz poziom detekcji.

Biorąc $\alpha = 0,01$ otrzymujemy $\rho_{0,01} = 0,0025758$. Dla 1 709 528 par, stanowiących niecałe 1,9% wszystkich par, wartość bezwzględna współczynnika korelacji Pearsona była większa od $\rho_{0,01}$. Odsetek ten 1,9 raza przekracza założony poziom błędu pierwszego rodzaju.

¹⁷⁵ L. Kowalski, *Statystyka*, BEL Studio, 2001, wydanie I, s. 178

¹⁷⁶ G.K. Kanji, *100 Statistical Tests*, 3rd Edition, SAGE Publications, 2006, s. 37

¹⁷⁷ G.K. Kanji, *100 Statistical Tests*, 3rd Edition, SAGE Publications, 2006, s. 109

Przy założeniu braku korelacji pomiędzy procedurami, przy $\alpha = 0,01$, dla każdej procedury spodziewamy się przeciętnie 134,3 procedur, dla których uzyskalibyśmy fałszywie pozytywny wynik. Na poziomie istotności 0,999 liczba ta nie powinna przekraczać 171. W rzeczywistości przeciętna liczba wynosi 253,3, zaś jedynie 18,4% procedur korelowało z nie więcej niż 171 innymi procedurami.

Najmniejszą liczbę korelujących procedur posiada test 6 najbliższych par dla wektorów 2 elementów 23-bitowych, metryka euklidesowa – jest to 112 procedur. Największą – test najdłuższej serii 1 w bloku 16-bitowym – 919 korelujących procedur.

Biorąc $\alpha = 0,001$ mamy $\rho_{0,001} = 0,0032905$. Dla 628 514 par, stanowiących niemal 7‰ wszystkich rozważonych, wartość bezwzględna współczynnika korelacji Pearsona była większa od $\rho_{0,001}$. Odsetek ten siedmiokrotnie przekracza założony poziom błędu pierwszego rodzaju.

Ponownie, zakładając brak korelacji pomiędzy procedurami, przy $\alpha = 0,001$, dla każdej procedury spodziewamy się przeciętnie 13,4 procedury, dla których uzyskalibyśmy fałszywie pozytywny wynik, na poziomie istotności 0,999 liczba ta nie powinna przekraczać 26. W rzeczywistości przeciętna liczba wynosi 88,5, zaś jedynie 5,1% procedur korelowało z nie więcej niż 26 innymi procedurami.

Na poziomie istotności 0,999 najmniejszą liczbę skorelowanych procedur – 12 – posiadają testy 5 najbliższych par dla wektorów 3 elementów 21-bitowych, metryka nieskończoność, 6 najbliższych par dla wektorów 2 elementów 23-bitowych, metryka euklidesowa i autokorelacji dla ciągu bloków 2-bitowych – przesunięcie o 393 bloki, największą – test najdłuższej serii 1 w bloku 16-bitowym – 762.

Analogicznie, wyznaczając kwantyle statystyki RI :

$$P\left(|RS| > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = \alpha,$$

możemy ocenić powszechność zjawiska korelowania procedur przy użyciu testu rangowego Spearman’a.

Biorąc $\alpha = 0,01$ otrzymujemy $|RS| > 2,5758$. Dla 1 590 257 par, stanowiących niecałe 1,8% wszystkich par, wartość bezwzględna statystyki Spearman’a była większa od 2,5758.

Jak poprzednio, przy założeniu braku korelacji pomiędzy procedurami, przy $\alpha = 0,01$, dla każdej procedury spodziewamy się przeciętnie 134,3 procedur, dla których uzyskalibyśmy fałszywie pozytywny wynik, na poziomie istotności 0,999 liczba ta nie powinna przekraczać 171. W rzeczywistości przeciętna liczba wynosi 235,2, zaś jedynie 24,1% procedur korelowała z nie więcej niż 171 innymi procedurami.

Na poziomie istotności 0,99 najmniejszą liczbę skorelowanych procedur posiada test autokorelacji dla ciągu bloków 2-bitowych – przesunięcie o 393 bloki – 118, największą – testy najdłuższej serii 1 w bloku 16 i 32-bitowym – 885.

Biorąc $\alpha = 0,001$ mamy $|RS| > 3,2905$. Dla 552 433 par, stanowiących ponad 6‰ wszystkich, wartość bezwzględna statystyki Spearman’a była większa od 3,2905.

Zakładając brak korelacji pomiędzy procedurami, przy $\alpha = 0,001$, dla każdej procedury spodziewamy się przeciętnie 13,4 procedur, dla których uzyskalibyśmy fałszywie pozytywny wynik, na poziomie istotności 0,999 liczba ta nie powinna przekraczać 26. W rzeczywistości przeciętna liczba wynosi 77, zaś jedynie 7,3% procedur korelowało z nie więcej niż 26 innymi procedurami.

Na poziomie istotności 0,999 najmniejszą liczbę skorelowanych procedur – 12 – posiadają testy 5 najbliższych par dla wektorów 3 elementów 21-bitowych, metryka nieskończoność i pokero-wy dla nienakładających się wektorów 14 elementów 5-bitowych, największą – test najdłuższej se-rii 1 w bloku 16-bitowym – 759.

Powyższe wyniki jednoznacznie wskazują na skorelowanie większości statystyk testowych, przy czym wartości progowe współczynnika korelacji są tak małe, że nie dają szansy skutecznej eliminacji procedur. Oczywiście pośród nich występują również wartości większe, a nawet zaobserwowane zostały jedynki, co zostanie omówione w dalszej części.

Kilka stron dalej, po zapoznaniu się z punktem VI.2.2 może zrodzić się pytanie: na ile świadomie pozostawione przez nas przypadki procedur równoważnych zaważyły na powyższych wyni-kach. Obliczyliśmy, że spowodowały one wzrost liczby skorelowanych par:

- w teście istotności współczynnika korelacji dla $\alpha = 0,01$ o 3,6%,
- w teście istotności współczynnika korelacji dla $\alpha = 0,001$ o 10,5%,
- w teście rangowym Spearman’a dla $\alpha = 0,01$ o 3,9%,
- w teście rangowym Spearman’a dla $\alpha = 0,001$ o 12,1%.

Wartości te nie wydają się istotne wobec wskazanych powyżej przekroczeń, sięgających 2930%.

1.3. Wyniki weryfikacji istotności korelacji dla procedur dla ciągów o długości 1 Mb

Badanie przebiegało analogicznie do przypadku ciągów długości 1 kiB, z tą różnicą, że już na wstępie wyeliminowaliśmy nadmiarowe procedury równoważne, co pozwoliło zredukować liczbę rozpatrywanych procedur do 528 519.

Dla 2 786 632 741 par, stanowiących niemal dokładnie 2% wszystkich par, wartość bezwzględna współczynnika korelacji Pearsona była większa od $\rho_{0,01}$, co jest wynikiem dwukrotnie wyższym od założonej wartości błędu pierwszego rodzaju.

Przy założeniu braku korelacji pomiędzy procedurami, przy $\alpha = 0,01$, dla każdej procedury spodziewamy się przeciętnie 5285,2 procedur, dla których uzyskalibyśmy fałszywie pozytywny

wynik. Na poziomie istotności 0,999 liczba ta nie powinna przekraczać 5510. W rzeczywistości przeciętna liczba wynosi 10 545 i wszystkie procedury zdecydowanie przekraczają podany próg.

Na poziomie istotności 0,99 najmniejszą liczbę skorelowanych procedur posiada test autokorelacji dla ciągu bloków 32-bitowych – przesunięcie o 11 450 bloków – 8511, największą – test zgodności Pearsona dla wektorów 2 elementów 1-bitowych – 16 366.

Dla 285 425 360 par, stanowiących nieco ponad 2% wszystkich, wartość bezwzględna współczynnika korelacji Pearsona była większa od $\rho_{0,001}$. Odsetek ten dwukrotnie przekracza założony poziom błędu pierwszego rodzaju.

Ponownie, zakładając brak korelacji pomiędzy procedurami, przy $\alpha = 0,001$, dla każdej procedury spodziewamy się przeciętnie 528,5 procedur, dla których uzyskalibyśmy fałszywie pozytywny wynik. Na poziomie istotności 0,999 liczba ta nie powinna przekraczać 601. W rzeczywistości przeciętna liczba wynosi 1080 i zaledwie 16 procedur korelowało z mniej niż 601 innymi.

Na poziomie istotności 0,999 najmniejszą liczbę skorelowanych procedur – 550 – posiada test autokorelacji dla ciągu bloków 6-bitowych – przesunięcie o 78 563 bloki, największą – test zgodności Pearsona dla wektorów 2 elementów 1-bitowych – 5062.

Analogicznie dla statystyki Spearman’a:

Dla 2 786 063 114 par, stanowiących niemal 2% wszystkich par, wartość bezwzględna statystyki Spearman’a była większa od 2,5758.

Jak poprzednio, przy założeniu braku korelacji pomiędzy procedurami, przy $\alpha = 0,01$, dla każdej procedury spodziewamy się przeciętnie 5285,2 procedur, dla których uzyskalibyśmy fałszywie pozytywny wynik, na poziomie istotności 0,999 liczba ta nie powinna przekraczać 5510. W rzeczywistości przeciętna liczba wynosi 10 543 i dla wszystkich procedur otrzymaliśmy przekroczenie progu.

Na poziomie istotności 0,99 najmniejszą liczbę skorelowanych procedur posiada test autokorelacji dla ciągu bloków 4-bitowych – przesunięcie o 118 508 bloków – 8581, największą – test logarytmu prawdopodobieństwa dla wektorów 1-elementowych 2-bitowych – 15 642.

Dla 285 083 886 par, stanowiących nieco ponad 2% wszystkich, wartość bezwzględna statystyki Spearman’a była większa od 3,2905.

Zakładając brak korelacji pomiędzy procedurami, przy $\alpha = 0,001$, dla każdej procedury spodziewamy się przeciętnie 528,5 procedur, dla których uzyskalibyśmy fałszywie pozytywny wynik. Na poziomie istotności 0,999 liczba ta nie powinna przekraczać 601, w rzeczywistości przeciętna liczba wynosi 1079. Dla wszystkich procedur jest większa od wartości oczekiwanej, zaś jedynie dla 9 nie przekracza progu.

Na poziomie istotności 0,999 najmniejszą liczbę skorelowanych procedur – 550 – posiada test autokorelacji dla ciągu bloków 11-bitowych – przesunięcie o 4783 bloki, największą – test zgodności Pearsona dla wektorów 2 elementów 1-bitowych – 4964.

Powyższe wyniki jednoznacznie wskazują na skorelowanie niemal wszystkich rozpatrywanych statystyk testowych.

W punkcie VI.2.4 wskazujemy 71 grup procedur, dla których współczynnik korelacji wynosi 1. Ich wpływ na obserwowaną liczbę sygnalizacji korelowania jest zdecydowanie mniej znaczący niż dla przypadku ciągów o długości 1 kiB, gdyż odpowiadają za zaledwie 141 *dodatkowych* par skorelowanych procedur.

2. Dwuwymiarowa regresja liniowa i nieliniowa dla prawdopodobieństw testowych

Na podstawie wyników uzyskanych w poprzednim punkcie oceniliśmy, że poszukiwanie korelacji dla wektorów zmiennych losowych o rozkładzie normalnym nie jest właściwym podejściem. Wynika to z kombinacji kilku czynników:

- wartości p_{val} dostarczane przez procedury testowe są interpretowane w skali liniowej,
- transformacja do rozkładu normalnego odbywa się przez silnie nieliniową funkcję $\Phi^{-1}(p_{val})$,
- pojawiające się, naturalnie lub z powodu skończoności arytmetyki komputera, wartości 0 i 1 p_{val} nie są transformowalne, przez co wymuszają stosowanie arbitralnie przyjętych dla nich przeciwobrazów,
- stosowanym przez nas narzędziem będzie głównie regresja liniowa, w której punkty takie, jak przytoczone powyżej, mogą wprowadzać istotne zniekształcenia.

Stąd, w dalszej części, operować będziemy na wektorach p_{val} , dla których rozpoczniemy od identyfikacji par procedur skorelowanych, a następnie, poszukując bardzo dobrych aproksymacji liniowych i nieliniowych, jedno- i wielowymiarowych będziemy starać się wskazywać te procedury, których wykonywanie nie wnosi żadnej nowej informacji lub też może zostać zastąpione prostą operacją arytmetyczną.

2.1. Regresja dwuwymiarowa

W tej części pracy wykorzystaliśmy powszechnie znane narzędzie statystyczne jakim jest regresja dwuwymiarowa. Dla każdej pary procedur testowych, na podstawie odpowiadających im wektorów p_{val} , zbadane zostało istnienie zależności następującej postaci:

- $U^j = aU^i + b$,
- $U^j = e^b (U^i)^a$,
- $U^j = b \cdot a^{U^i}$,
- $U^j = a \ln U^i + b$,
- $U^j = a(U^i)^3 + b(U^i)^2 + c(U^i) + d$,
- $\ln U^j = a \ln^3 U^i + b \ln^2 U^i + c \ln U^i + d$,

- $\ln U^j = a(U^i)^3 + b(U^i)^2 + c(U^i) + d$,
- $U^j = a \ln^3 U^i + b \ln^2 U^i + c \ln U^i + d$,
- $U^i = a(U^j)^3 + b(U^j)^2 + c(U^j) + d$,
- $\ln U^i = a \ln^3 U^j + b \ln^2 U^j + c \ln U^j + d$,
- $\ln U^i = a(U^j)^3 + b(U^j)^2 + c(U^j) + d$,
- $U^i = a \ln^3 U^j + b \ln^2 U^j + c \ln U^j + d$.

Rozważenie wielomianów stopnia 3 uznaliśmy za rozsądny kompromis pomiędzy wszechstronnością badania, a nakładem obliczeniowym.

Dla każdego z przypadków analizowana była jedynie jedna wartość – współczynnik korelacji Pearsona z próby. W pierwszym przypadku jest on wyborem oczywistym, w pozostałych mogliśmy posłużyć się na przykład współczynnikiem korelacji krzywoliniowej:

$$\tilde{\rho}_{i,j} = \sqrt{1 - \frac{\sum_{k=1}^n (u_k^j - \hat{u}_k^j)^2}{\sum_{k=1}^n (u_k^j - \bar{u}^j)^2}},^{178}$$

gdzie zakładamy, że U^j jest objaśniania przy użyciu U^i i:

- $\hat{u}_k^j = e^b (u_k^i)^a$ w przypadku drugiej formuły, analogicznie trzeciej i czwartej,
- $\hat{u}_k^j = a(u_k^i)^3 + b(u_k^i)^2 + c(u_k^i) + d$ w przypadku piątej formuły i analogicznie kolejnych,
- $\bar{u}^j = \frac{1}{n} \sum_{k=1}^n u_k^j$.

Jednakże biorąc pod uwagę przyświecający cel, uznaliśmy za właściwszy wskaźnik korelację liniową pomiędzy zmienną wyjaśnianą, a wektorami otrzymanymi według poszczególnych formuł. Posłużenie się tym właśnie współczynnikiem uzasadnia dodatkowo potrzeba jego użycia w wykorzystanej dalej regresji wielowymiarowej.

Dodatkowo, w wybranych przypadkach, na wektorze procedury wyjaśniającej stosowane było przekształcenie $U_{||}^i = 1 - 2|U^i - \frac{1}{2}|$, a wektor $U_{||}^i$ zastępował U^i w formułach wymienionych na początku punktu. Uzasadnienie dla tego kroku znaleźć można w punkcie VI.2.3.2., na przykładzie wykresów z rysunków VI.12 i VI.13.

Dla procedur dla ciągów 1 kiB zebrane zostały wszystkie przypadki, w których wartość bezwzględna współczynnika korelacji dla którejkolwiek z formuł przekraczała 0,01.

Ze względu na dużą liczbę procedur ($N = 528\,519$) dla ciągów 1 Mb, które znalazły się w tym etapie, zmuszeni byliśmy do opracowania metody znaczącej redukcji obliczeń. Podstawowym zało-

¹⁷⁸ L. Kowalski, *Statystyka*, BEL Studio, 2001, wydanie I, s. 69

żeniem jest wykonanie sprawdzenia dla każdej pary, obieralnym parametrem zaś dopuszczalny nakład obliczeń, który dla własnej wygody określiliśmy jako krotność obliczeń wykonanych dla ciągów 1 kiB: $K \binom{13\,432}{2}$ regresji dla par wektorów $n = 10^6$ elementowych.

Dla każdej pary procedur wykonane zostanie badanie *pilotażowe*, wykorzystujące początkowe n_p elementów, gwarantujące wykrycie, z prawdopodobieństwem co najmniej $1 - 1/(0,37N)$, każdego przypadku, gdy wartość bezwzględna współczynnika korelacji przekracza pewne ρ . Dopiero dla tych przypadków wyznaczane będą korelacje dla pełnych wektorów. Formuła określająca powyższe prawdopodobieństwo przyjęta została na podstawie wyników przedstawionych w punkcie VI.1.3, gdzie maksymalna zaobserwowana liczba przypadków, gdy przynajmniej jeden z testów sygnalizował skorelowanie na poziomie istotności 0,99, wyniosła 19 753, a więc około $0,037N$. Powinno to zagwarantować oczekiwaną liczbę niewykrytych skorelowań, dla każdej procedury z osobna, mniejszą od 0,1.

Oczywiście przyjęta formuła jest podwójnie nieprawidłowa, tzn.:

- poszukujemy wartości ρ znacznie przekraczających, użyte w punkcie VI.1.3, $\rho_{0,01}$,
- poszukujemy również korelacji nieliniowych, których część nie mogła być wykrywana w badaniu opisanym w punkcie VI.1,

lecz obserwacja wyników badania dla ciągów 1 kiB wskazała, że nie prowadzi to do uzyskania nieprawdziwego, czy też zbyt zgrubnego oszacowania dla liczby skorelowanych procedur.

Celem przyjętej strategii było znalezienie możliwie małej wartości ρ_0 . W tym celu posłużyliśmy się złożeniem dwóch testów, omówionego już w punkcie VI.1.1 testu istotności współczynnika korelacji oraz testu dla wartości współczynnika korelacji, opisanego poniżej.

Zakładamy, że poszukiwane korelacje będą dodatnie i weryfikujemy czy na poziomie istotności $1 - \beta$ uzyskana wartość współczynnika korelacji ρ z próby n_0 -elementowej może wskazywać, że rzeczywista korelacja jest mniejsza od ρ_0 . Statystyka testowa:

$$RI_{\rho_0}(\rho) = \left(\frac{1}{2} \ln \frac{(1+\rho)(1-\rho_0)}{(1-\rho)(1+\rho_0)} - \frac{\rho_0}{2(n_0-1)} \right) \sqrt{n_0-3}$$

ma standardowy rozkład normalny, toteż dla przyjętej hipotezy alternatywnej przedział krytyczny ma postać: $(-\infty; -\Phi^{-1}(1-\beta)]$.¹⁷⁹

Z nierówności $RI_{\rho_0}(\rho) \leq -\Phi^{-1}(1-\beta)$ wyznaczyć możemy graniczną wartość współczynnika korelacji ρ z próby:

¹⁷⁹ L. Kowalski, *Statystyka*, BEL Studio, 2001, wydanie I, s. 178

$$\rho_g \leq \frac{\frac{1+\rho_0}{1-\rho_0} \exp\left(\frac{\rho_0}{n_0-1} - \frac{2\Phi^{-1}(1-\beta)}{\sqrt{n_0-3}}\right) - 1}{\frac{1+\rho_0}{1-\rho_0} \exp\left(\frac{\rho_0}{n_0-1} - \frac{2\Phi^{-1}(1-\beta)}{\sqrt{n_0-3}}\right) + 1}.$$

Dla tak wyznaczonego ρ_g możemy obliczyć odpowiadające mu prawdopodobieństwo błędu pierwszego rodzaju w teście istotności współczynnika korelacji, dla nas oznaczające prawdopodobieństwo fałszywie pozytywnego wyniku badania sondażowego, skutkującego wykonaniem obliczeń dla pełnych wektorów. W ten sposób uzyskujemy formułę określającą nakład obliczeniowy:

$$\left(2 \left(1 - T_{n_0-2} \left(\rho_g \sqrt{\frac{n_0-2}{1-\rho_g^2}} \right)\right) + \frac{n_0}{n} \right) \binom{N}{2}.$$

Dla uzyskanej z poprzedniego rozdziału liczby procedur $N = 528\,519$, otrzymaliśmy $\beta = 5,06 \cdot 10^{-6}$ oraz przyjęliśmy $K = 19$. Dla kolejnych ρ_0 wykonywaliśmy minimalizację powyższej formuły względem n_0 i w ten sposób znaleźliśmy minimalną wartość ρ_0 , dla której nakład obliczeń nie przekraczał założonego.

W powyższym sensie optymalną parą jest: $\rho_0 = 0,1$ i $n_0 = 6000$. Ostatecznie nakład obliczeń jest osiemdziesięciokrotnie niższy od wariantu z obliczeniami na pełnych wektorach.

W wybranych przypadkach wykonaliśmy również regresję i tzw. regresję geometryczną¹⁸⁰ dla dopasowania do okręgu. W pierwszym przypadku posługiwaliśmy się powszechnie znanym przekształceniem równania okręgu o środku $(x_0; y_0)$ i promieniu r do postaci:

$$x^2 + y^2 = ax + by + c,$$

gdzie $a = 2x_0$, $b = 2y_0$, $c = r^2 - x_0^2 - y_0^2$.

Z uwagi na to, że standardowa regresja do okręgu często nie daje zbyt dobrych rezultatów posilkowaliśmy się również wynikami zawartymi w [GGS96]¹⁸¹, z użyciem metody Gaussa-Newtona¹⁸² do znalezienia x_0 , y_0 i r dających minimum funkcji:

$$\sum_{k=1}^n \left(\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2} - r \right)^2.$$

W otoczeniu przybliżonego rozwiązania (środku okręgu) budowane jest liniowe przybliżenie nieliniowej funkcji, jaką jest odległość euklidesowska, przy użyciu pochodnych cząstkowych, dla którego łatwo jest już wykonać regresję metodą najmniejszych kwadratów. Uzyskane rozwiązanie wykorzystywane jest jako punkt startowy do kolejnej iteracji. Jako punkt startowy można wykorzy-

¹⁸⁰ W. Gander, G.H. Golub, R. Strebel, *Least-Squares Fitting of Circles and Ellipses*, Bulletin of the Belgian Mathematical Society, Supplement to Volume 3, 1996, s. 63

¹⁸¹ W. Gander, G.H. Golub, R. Strebel, *Least-Squares Fitting of Circles and Ellipses*, Bulletin of the Belgian Mathematical Society, Supplement to Volume 3, 1996

¹⁸² P.E. Gill, W. Murray, M.H. Wright, *Practical Optimization*, Academic Press Limited, wyd. 11, 1997, s. 133-136

stać środek okręgu uzyskany z pierwszej metody, warunkiem końca jest uzyskanie dostatecznie małej poprawki dla środka okręgu.

Dla punktu startowego $(x_0; y_0)$ zastępujemy minimalizowaną funkcję następującą:

$$\sum_{k=1}^n \left(\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2} + a \frac{x_0 - u_k^i}{\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2}} + b \frac{y_0 - u_k^j}{\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2}} - r \right)^2,$$

a rozwiązanie znajdujemy z układu równań liniowych postaci:

$$\begin{bmatrix} \sum_{k=1}^n \frac{(x_0 - u_k^i)^2}{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2} & \sum_{k=1}^n \frac{(x_0 - u_k^i)(y_0 - u_k^j)}{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2} & \sum_{k=1}^n \frac{-(x_0 - u_k^i)}{\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2}} \\ \sum_{k=1}^n \frac{(x_0 - u_k^i)(y_0 - u_k^j)}{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2} & \sum_{k=1}^n \frac{(y_0 - u_k^j)^2}{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2} & \sum_{k=1}^n \frac{-(y_0 - u_k^j)}{\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2}} \\ \sum_{k=1}^n \frac{(x_0 - u_k^i)}{\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2}} & \sum_{k=1}^n \frac{(y_0 - u_k^j)}{\sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2}} & -n \end{bmatrix} \times \begin{bmatrix} a \\ b \\ r \end{bmatrix} = \begin{bmatrix} -\sum_{k=1}^n (x_0 - u_k^i) \\ -\sum_{k=1}^n (y_0 - u_k^j) \\ -\sum_{k=1}^n \sqrt{(x_0 - u_k^i)^2 + (y_0 - u_k^j)^2} \end{bmatrix}.$$

Punktem startowym w kolejnej iteracji jest $(x_0 + a; y_0 + b)$.

2.2. Korelacja wystarczająco bliska 1

We wstępie do tego rozdziału zasygnalizowaliśmy, że poszukiwać będziemy procedur, których p_{val} można wystarczająco dokładnie wyznaczyć na podstawie p_{val} innych procedur. Ten punkt poświęcamy określeniu co oznacza owo „wystarczająco”.

Oczywistym jest, że współczynnik korelacji równy 1 albo -1 odpowiada możliwości wystarczająco dokładnego wyznaczenia, zaś równy 0 – nie, a gdzieś pomiędzy musimy postawić arbitralny próg.

Budowany pakiet testów służyć ma badaniu generatorów poprzez ocenę równomierności rozkładu p_{val} uzyskiwanych dla poszczególnych procedur testowych, stąd też, z punktu widzenia przyszłego użytkownika, nie tyle jest istotna niemal idealna zgodność p_{val} obu procedur, a fakt, czy przejście przez generator badania przy użyciu jednej procedury implikuje również przejście dla drugiej. Dzięki takiemu przeformułowaniu naszych oczekiwań, tytułowe „wystarczająco” będziemy mogli wyrazić w języku prawdopodobieństw.

Dla kolejnych wartości współczynnika korelacji $\rho = \{0,9750..0,9989; 0,999000..0,999999\}$ rozpatrywaliśmy realizacje zmiennych losowych postaci:

$$Z_\rho = X + bX(1 - X)Y,$$

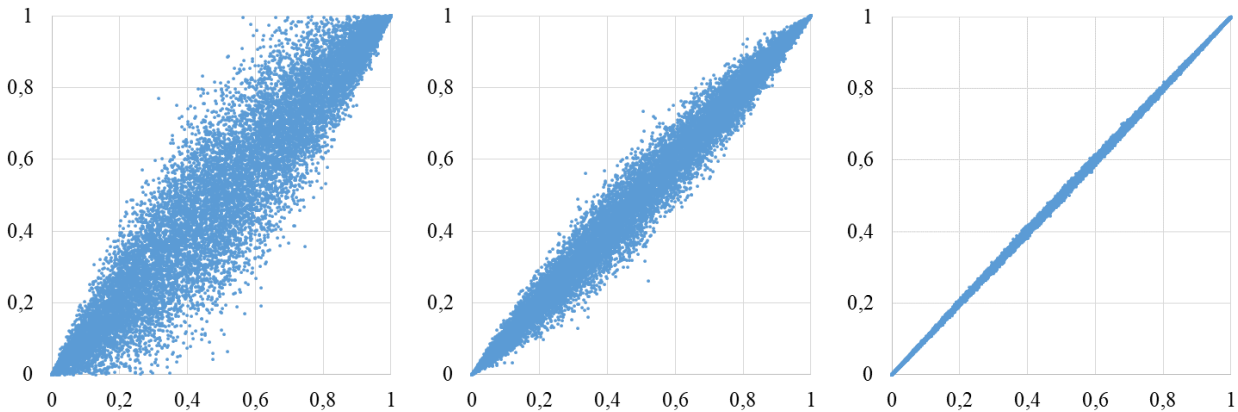
gdzie

– X ma rozkład równomierny na przedziale $(0;1)$,

– Y ma standardowy rozkład normalny,

$$b = \sqrt{\frac{5}{2} \left(\frac{1}{\rho^2} - 1 \right)}.$$

Postać współczynnika b powoduje, że współczynnik korelacji dla pary zmiennych X i Z_ρ wynosi ρ . Na rysunku VI.1 prezentujemy trzy przykładowe wykresy, na które naniesiono punkty $(x_i; z_i^\rho)$, dla różnych wartości ρ . Analogiczne wykresy, bazujące już na rzeczywistych danych, prezentowane są w dalszej części rozdziału, podobieństwo jest na tyle duże, że powinno legitymować przyjęty przez nas model.



Rysunek VI.1. Zobrazowanie skorelowania wektorów X i Z_ρ dla $\rho = 0,95, 0,99$ i $0,999864$

Mając dwie procedury, modelowane jako zmienne losowe X i Z_ρ , zbadaliśmy czy, dla danej wartości ρ , uzasadnione jest założenie równomierności rozkładu Z_ρ na podstawie równomierności X .

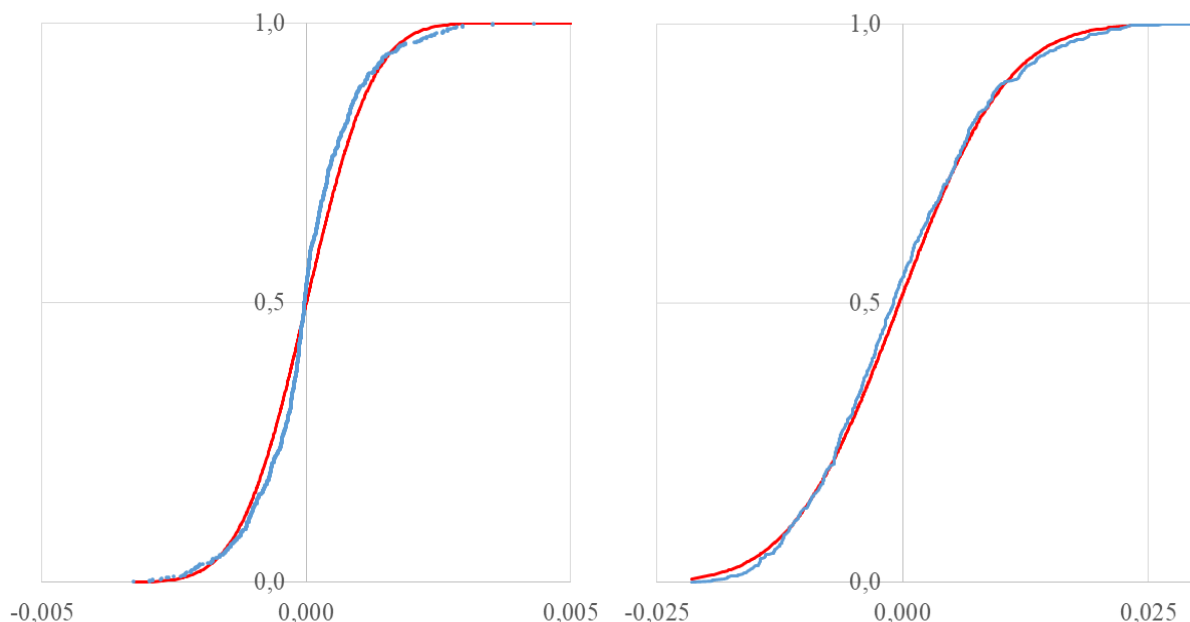
Aby uniknąć problemów proceduralnych dodaliśmy jeszcze niewielką poprawkę:

$$z_i^\rho = \begin{cases} -z_i^\rho, & z_i^\rho < 0, \\ z_i^\rho, & 0 \leq z_i^\rho \leq 1, \\ 2 - z_i^\rho, & z_i^\rho > 1. \end{cases}$$

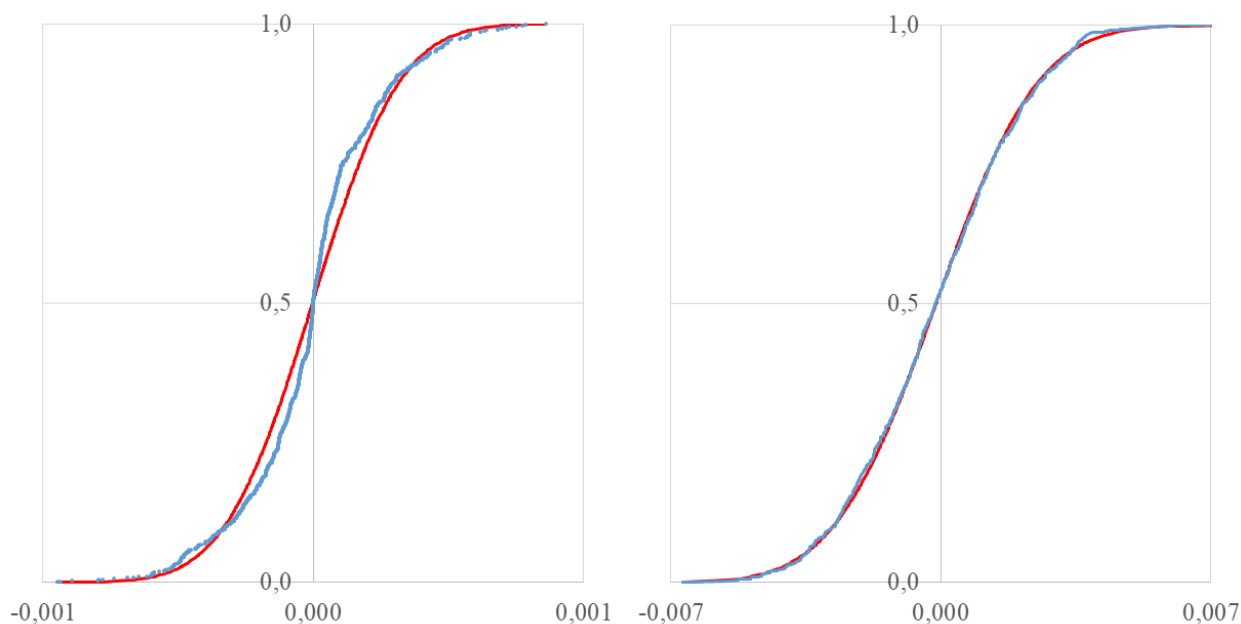
Generowane były 8192-elementowe ciągi liczb rzeczywistych z przedziału $[0;1]$ z 64-bitową precyzją, które traktowane były jako ciąg realizacji zmiennej losowej X . Następnie losowany był kolejny taki ciąg i przy pomocy dystrybucyj odwrotnej standardowego rozkładu normalnego otrzymywany był ciąg realizacji zmiennej Y . Dla każdej z wartości współczynnika ρ badana była

równomierność ciągu odpowiadającego zmiennej Z_ρ . Dalsza analiza dotyczyła różnicy p_{val} z testu Andersona-Darlinga otrzymanych dla wyjściowego ciągu i ciągu skorelowanego. Ostatecznie zdecydowaliśmy się zebrać charakterystyki odpowiadające przedziałom decylowym wyjściowych p_{val} .

Dla dużych wartości współczynnika ρ uznaliśmy za uzasadnione przyjęcie, że różnice p_{val} otrzymanych dla ciągów wyjściowego i skorelowanego mają w przybliżeniu rozkład normalny, dla którego parametry wyznaczane były na podstawie wykonanego eksperymentu – dla każdego ρ zebrano 1 000 000 różnic, co daje przeciętnie 100 000 w każdej próbce.



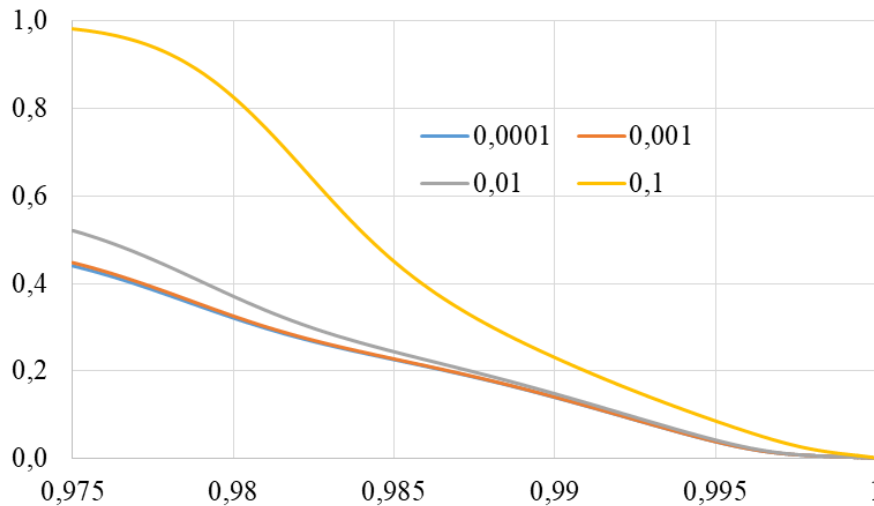
Rysunek VI.2. Dystrybuanty empiryczne i odpowiadające im dystrybuanty rozkładów normalnych dla różnic p_{val} dla $\rho = 0,999$ i przedziałów $(0; 0,1]$ oraz $(0,5; 0,6]$



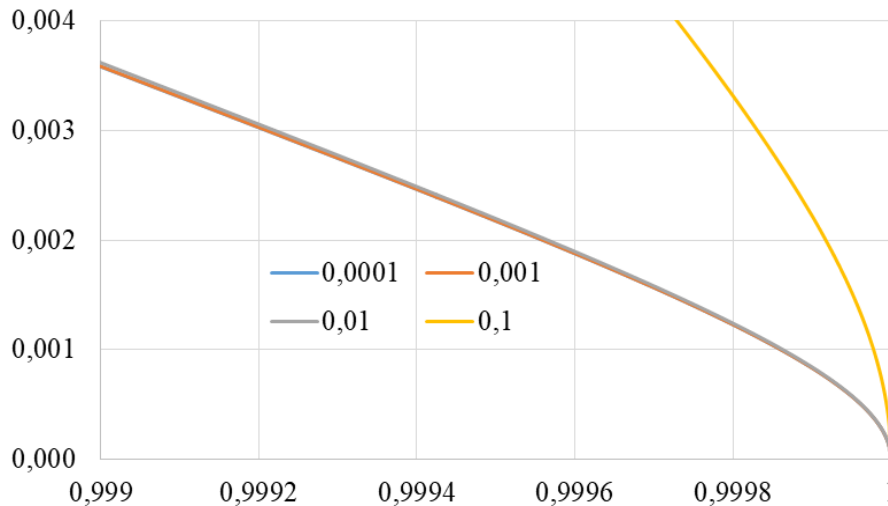
Rysunek VI.3. Dystrybuanty empiryczne i odpowiadające im dystrybuanty rozkładów normalnych dla różnic p_{val} dla $\rho = 0,9999$ i przedziałów $(0; 0,1]$ oraz $(0,5; 0,6]$

Na rysunkach VI.2 i VI.3 nanieśliśmy po około 1000 punktów dystrybuant empirycznych i przyjętego przez nas modelu ciągłego w postaci odpowiednich rozkładów normalnych, dla wybranych przypadków. W obu przypadkach widoczna jest znaczna niedokładność przybliżenia zastosowanym modelem dla przedziału $(0; 0,1]$ oraz duża zgodność dla przedziału $(0,5; 0,6]$.

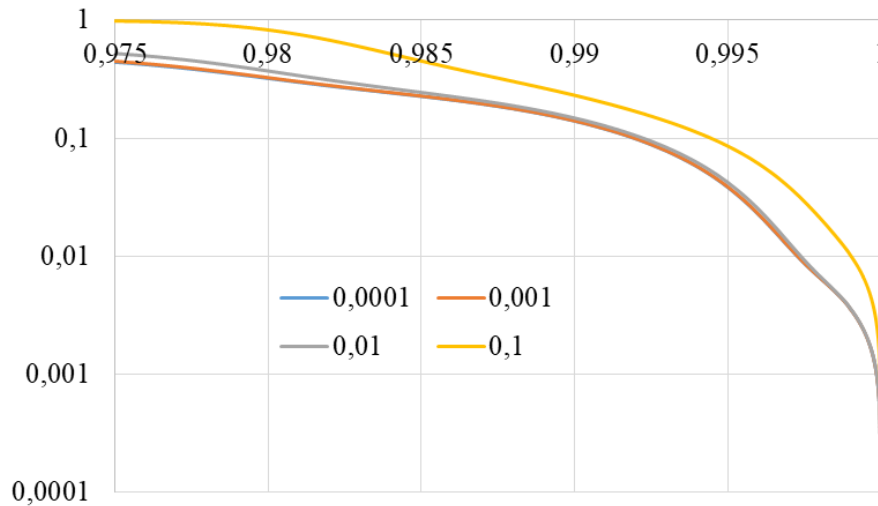
Na rysunkach VI.4, VI.5 i VI.6 prezentujemy zależność pomiędzy prawdopodobieństwem nie przejścia testu równomierności przez procedurę skorelowaną od wartości współczynnika korelacji dla różnych poziomów istotności. Pierwsze dwa wykresy sporządzone zostały w skali liniowej, trzeci w logarytmicznej. Na rysunku VI.5 prezentujemy fragment wykresu z rysunku VI.4, ograniczony do najwyższych wartości współczynnika korelacji.



Rysunek VI.4. Prawdopodobieństwo nieprzejścia testu równomierności przez procedurę skorelowaną w funkcji współczynnika korelacji ρ dla różnych poziomów istotności $1 - p$ w skali liniowej



Rysunek VI.5. Prawdopodobieństwo nieprzejścia testu równomierności przez procedurę skorelowaną w funkcji współczynnika korelacji ρ dla różnych poziomów istotności $1 - p$ dla najwyższych wartości współczynnika korelacji



Rysunek VI.6. Prawdopodobieństwo nieprzejścia testu równomierności przez procedurę skorelowaną w funkcji współczynnika korelacji ρ dla różnych poziomów istotności $1 - p$ w skali logarytmicznej

Jedynie przypadek $p = 0,1$ odróżnia się wyraźnie od pozostałych, co najprawdopodobniej spowodowane jest postacią formuły określającej $q_{\rho,p}$. Pierwszy składnik jest tu, co prawda, równy 0, ale zasadnicze znaczenie ma drugi, który sąsiaduje wprost z progiem p , mając około 2,5-krotnie większe odchylenie standardowe.

Obserwując wykresy analogiczne dla przedstawionych na rysunku VI.1 oraz dane liczbowe przyjęliśmy, że wartością progową współczynnika korelacji będzie 0,999864, jest ona najmniejszą, dla której prawdopodobieństwo nieprzejścia testu równomierności przez procedurę skorelowaną jest mniejsze od 0,001 dla $p = 0,01$, 0,001 i 0,0001. Trzeci z wykresów na rysunku VI.1 odpowiada podanej wartości progowej współczynnika korelacji.

2.3. Wyniki dla regresji dwuwymiarowej uzyskane dla procedur dla ciągów o długości 1 kiB

2.3.1. Korelacja równa 1

Pierwszy sukces, w postaci wykrycia par procedur o korelacji równej 1, został zaprojektowany w przeprowadzonym eksperymencie, lecz nie całkowicie. Taka wartość współczynnika korelacji oznacza, że obie procedury zwracały zawsze identyczne wartości p_{val} . Dotyczy to 254 procedur testów:

- logarytmu prawdopodobieństwa dla wektorów:
 - $t = 1, d = 4, 8,$
 - $t = 2, d = 2, 4,$
 - $t = 4, d = 1, 2,$
 - $t = 8, d = 1,$
- momentu pierwszego rzędu
 - $t = 1, d = 3..30,$
 - $t = 2, d = 2..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$

- momentu drugiego rzędu
 - $t = 1, d = 2..29,$
 - $t = 2, d = 1..14,$
 - $t = 3, d = 1..9,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..5,$
 - $t = 6..7, d = 1..4,$
 - $t = 8..9, d = 1..3,$
 - $t = 10..14, d = 1..2,$
 - $t = 15..29, d = 1,$
- potęgowo-dywergencyjnego – $\delta = 1$
 - $t = 1, d = 2..6,$
 - $t = 2, d = 1..3,$
 - $t = 3, d = 1..2,$
 - $t = 4..6, d = 1,$
- potęgowo-dywergencyjnego – $\delta = 2$
 - $t = 1, d = 4,$
 - $t = 2, d = 2,$
 - $t = 4, d = 1,$
- porządku książek
 - $d = 16, k = 3,$
- stosu książek
 - $d = 8, k = 6..7,$
 - $d = 16, k = 3,$
- wag Hamminga w ciągu bloków
 - $d = 3, t = 2, k = 2, 4,$
- zgodności Pearsona
 - $t = 1, d = 2..7,$
 - $t = 2, d = 1..3,$
 - $t = 3, d = 1..2,$
 - $t = 4..7, d = 1,$

łączących się w 73 grupy, które omówimy poniżej.

Pierwszą grupę stanowią trzy procedury testu logarytmu prawdopodobieństwa dla wektorów:

- $t = 1, d = 4,$
- $t = 2, d = 2,$
- $t = 4, d = 1.$

Łatwo zauważyć, że we wszystkich przypadkach iloczyn $t \cdot d = 4$. Choć idea tego testu mówi o rozpatrywaniu wektorów t elementów d -bitowych, to sam przebieg testu sprowadza wszystkie te przypadki do tego samego bloku czterobitowego.

Analogicznie w drugiej grupie mamy cztery kombinacje t i d dające blok ośmiu bitów. Kolejne 62 grupy składają się z, identycznie powiązanych, procedur testów momentów i potęgowo-dywergencyjnego.

W przypadku trzech par:

- stosu książek $d = 8, k = 6$ i stosu książek $d = 8, k = 7,$

- stosu książek $d = 16, k = 3$ i porządku książek $d = 16, k = 3$,
- wag Hamminga w ciągu bloków $d = 3, t = 2, k = 2$ i wag Hamminga w ciągu bloków $d = 3, t = 2, k = 4$,

na skutek grupowania klas w teście zgodności chi-kwadrat Pearsona, druga z procedur „degenerowana” jest do pierwszej.

Jako przedostatnie omówimy 6 grup procedur testu zgodności Pearsona. W przypadku 15 wymienionych powyżej procedur sytuacja jest identyczna jak dla testów logarytmu prawdopodobieństwa, momentów i potęgowo-dywergencyjnego. Jednak należy zauważyć, że jeśli do tej fazy przeszłyby procedury operujące na większych długościach bloku, to w przypadku konieczności wykonania grupowania w teście zgodności, inaczej będzie ono przebiegać dla różnych kombinacji t i d .

Choć odnieśliśmy się już do każdej z wymienionych na początku punktu procedur, to okazuje się, że mamy jeszcze jeden ciekawy przypadek – to 13 par procedur z testów potęgowo-dyferencyjnego dla $\delta = 1$ i zgodności Pearsona dla tych samych kombinacji t i d . Spojrzenie do punktu II.3.1. uzmysławia, że te dwa testy są równoważne, ale nieco odmiennie wykonywane. Prawdopodobnie niewielki błąd w wyznaczeniu parametrów dla testu potęgowo-dywergencyjnego spowodował, że w jego przypadku do tego etapu dotarła jedna procedura mniej. Na tej też podstawie podjęliśmy decyzję, że to procedury z testu potęgowo-dywergencyjnego uznamy za nadmiarowe.

Podsumowując, ten krok pozwolił na wyeliminowanie 186 procedur, a lista procedur z początku tego punktu została zredukowana do następującej:

- logarytm prawdopodobieństwa dla wektorów – $t = 1, d = 4, 8$,
- momentu pierwszego rzędu – $t = 1, d = 3..30$,
- momentu drugiego rzędu – $t = 1, d = 2..29$,
- potęgowo-dywergencyjny – $\delta = 2 - t = 1, d = 4$,
- stosu książek
 - $d = 8, k = 6$,
 - $d = 16, k = 3$,
- wag Hamminga w ciągu bloków
 - $d = 3, t = 2, k = 2$,
- zgodności Pearsona – $t = 1, d = 2..7$.

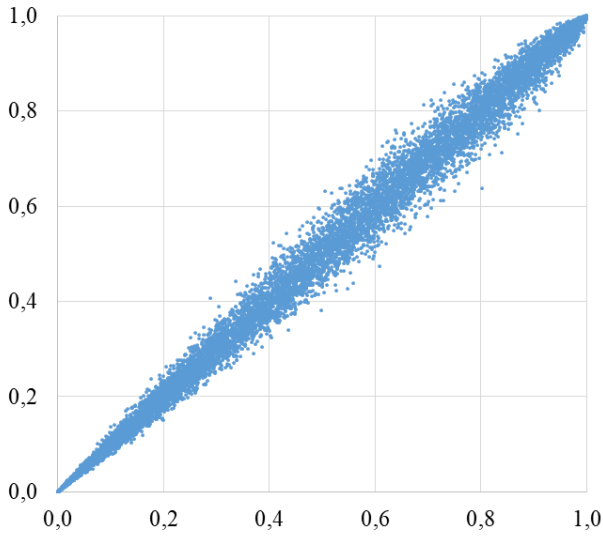
2.3.2. Korelacja bliska 1

Zidentyfikowaliśmy jeden przypadek, w którym dla trzech procedur współczynniki korelacji liniowej wynoszą od 0,9966 do 0,9992, są nimi:

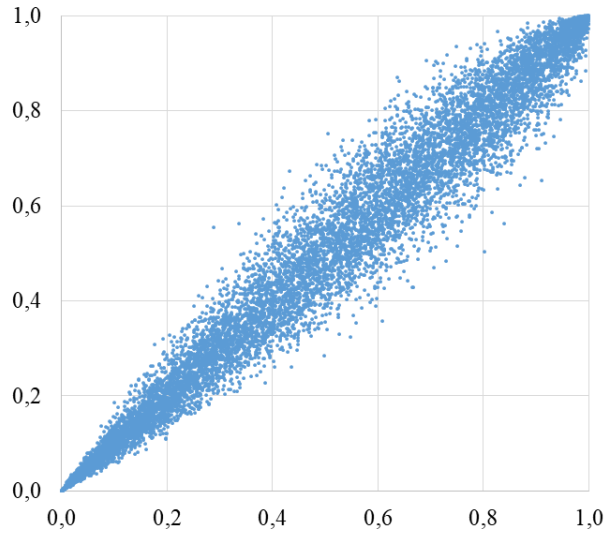
- logarytm prawdopodobieństwa dla wektorów – $t = 4, d = 4$,
- potęgowo-dywergencyjny – $\delta = 2 - t = 1, d = 4$,

- zgodności Pearsona – $t = 1, d = 4$.

Na rysunkach VI.7 i VI.8 naniesiono punkty, których współrzędne są odpowiadającymi sobie parami p_{val} uzyskanymi dla analizowanych procedur.



Rysunek VI.7. Regresja liniowa dla pary testów zgodności Pearsona dla $t = 1, d = 4$ i logarytmu prawdopodobieństwa dla wektorów $t = 4, d = 4$

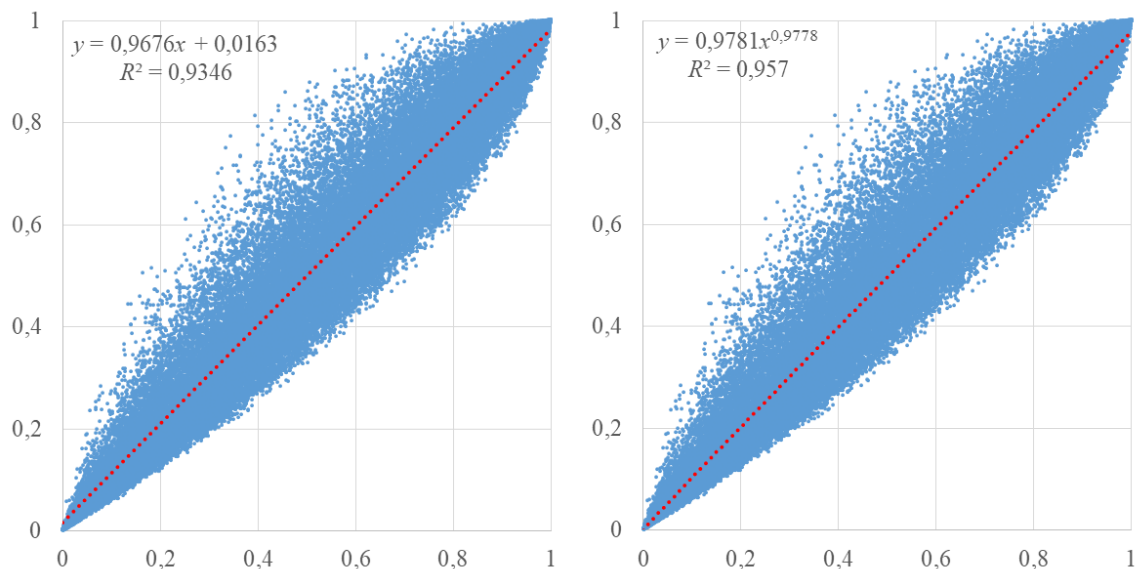


Rysunek VI.8. Regresja liniowa dla pary testów potęgowo-dywergencyjnego dla $\delta = 2, t = 1, d = 4$ i logarytmu prawdopodobieństwa dla wektorów $t = 4, d = 4$

W obu przypadkach siła skorelowania jest dość duża, ale niższa od progu przyjętego w punkcie VI.2.2. Nawet dla lepszego z tych przypadków chmura punktów jest znacznie bardziej rozproszona niż na trzecim wykresie z rysunku VI.1. Wykres dla pary testów zgodności Pearsona i potęgowo-dywergencyjnego jest zbliżony do prezentowanego na rysunku VI.7. Do procedur tych powrócimy jeszcze w punkcie VI.4.3.

Następna, w kolejności siły skorelowania, para to testy Andersona-Darlinga i Cramera-von Misesa dla bloków 16-bitowych. Współczynnik korelacji liniowej wynosi dla tej pary 0,96658, nieco wyższą wartość uzyskujemy dla zależności postaci $U^j = e^{a \ln U^i + b} - 0,97835$ i $U^j = e^{a \ln^3 U^i + b \ln^2 U^i + c \ln U^i + d} - 0,97838$. Na rysunku VI.9 prezentujemy brak realnego znaczenia zastosowania regresji krzywoliniowej na jakość prognozy dla omawianej pary. Na rysunkach naniesiono 50 000 punktów (par p_{val}), natomiast obliczenia współczynników korelacji wykonano dla pełnych wektorów (1 000 000 par), stąd wartości liczbowe na wykresach mogą nieco odbiegać od prezentowanych powyżej.

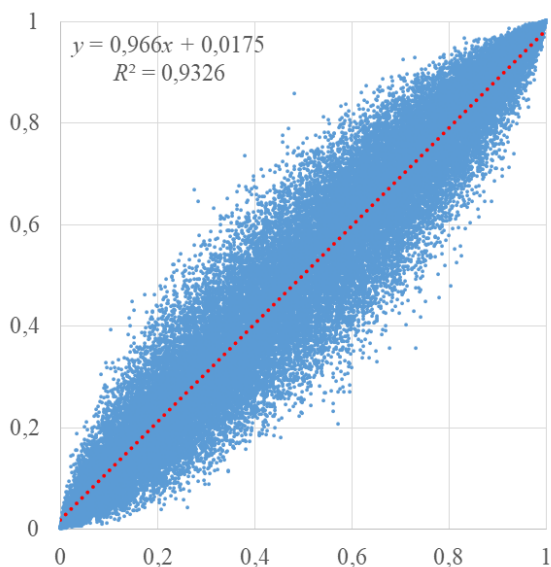
Przytoczone dane liczbowe pokazują, że regresja potęgowa nie daje istotnie lepszego dopasowania do danych, przy czym wynik uzyskane przy użyciu wielomianu stopnia 3 jest tylko nieznacznie lepszy od uzyskanego dla postaci liniowej. Wskazuje to, że w istocie mamy do czynienia z zależnością liniową, a jedynie widoczna na rysunku VI.9 chmura punktów jest wyraźnie niesymetryczna względem prostej $y = x$.



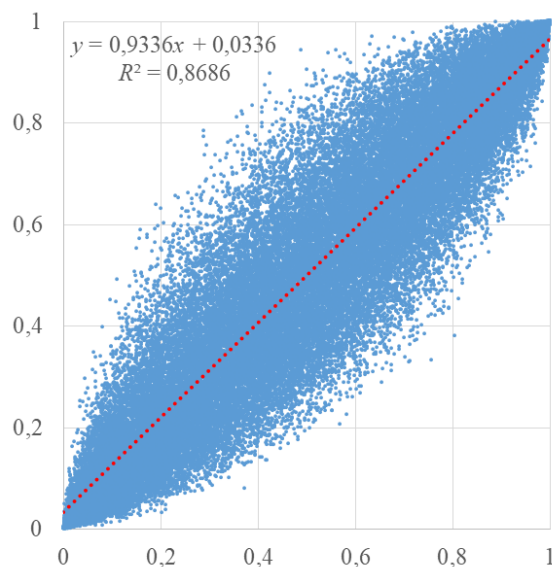
Rysunek VI.9. Regresja liniowa i potęgowa dla pary testów Andersona-Darlinga i Cramera-von Misesa dla bloków 16-bitowych

W obu przypadkach widać, że pomimo wysokich wartości współczynnika determinacji R^2 nie uzyskujemy możliwości określenia wartości p_{val} jednej z procedur na podstawie wyniku dla drugiej. Na przykład, dla p_{val} testu Andersona-Darlinga należących do przedziału $(0,399; 0,401)$, w teście Cramera-von Misesa otrzymaliśmy p_{val} z zakresu $(0,2517; 0,7804)$.

Dokładnie taki sam obraz otrzymaliśmy dla pozostałych 16 par procedur testu Andersona-Darlinga i Cramera-von Misesa dla bloków o długości od 17 do 32 bitów.



Rysunek VI.10. Regresja liniowa dla pary testów nienakładających się 3-bitowych wzorców w bloku $M = 256$ -bitowym



Rysunek VI.11. Regresja liniowa dla pary testów Kołmogorowa i Cramera-von Misesa dla bloków 32-bitowych

Kolejne miejsca, w porządku malejącej siły korelacji, zajęły dwie pary procedur testu nienakładających się 3-bitowych wzorców w bloku $M = 256$ -bitowym:

- wzorce 0×1 i 0×4 ,
- wzorce 0×3 i 0×6 ,

dla których współczynniki korelacji liniowej wynoszą 0,96576 i 0,96581. W tych przypadkach regresja nieliniowa daje tylko nieznacznie lepsze wyniki. Widoczna na rysunku VI.10 chmura punktów jest zdecydowanie bardziej symetryczna.

Kolejne miejsca przypadły 27 parom procedur testów momentów pierwszego i drugiego rzędu dla bloków o długości od 3 do 29 bitów, dla których obraz jest identyczny jak dla testu nienakładających się wzorców (procedura testu momentów drugiego rzędu dla bloku 30-bitowego została wyeliminowana w poprzednim rozdziale).

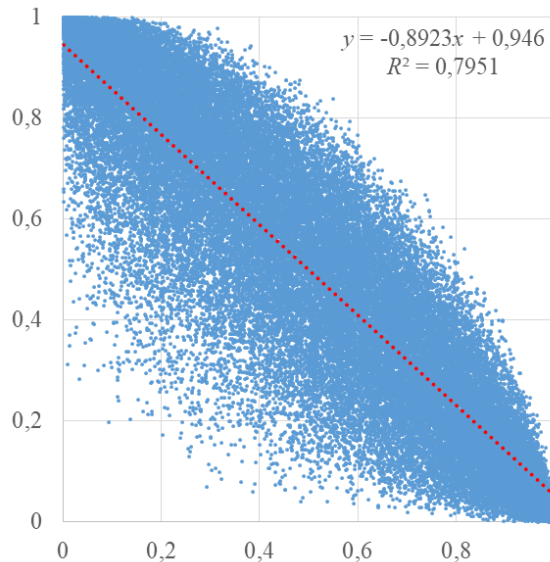
Nieco słabiej skorelowane okazały się procedury testów Kołmogorowa i Cramera-von Misesa dla bloków o długości od 16 do 32 bitów. W tych przypadkach wartość współczynnika korelacji liniowej wahała się od 0,93118 do 0,93177. Na rysunku VI.11 widać, iż chmura punktów jest wyraźniej rozproszona, niż dla wcześniejszych.

Łącznie dla 66 par procedur współczynnik korelacji zawiera się w przedziale $(0,9; 1,0)$. Dla 162 par współczynnik korelacji zawiera się w przedziale $(0,8; 0,9)$, są to procedury testów:

- Andersona-Darlinga i Kołmogorowa dla bloków o długości od 16 do 32 bitów,
- Andersona-Darlinga i momentów pierwszego rzędu dla bloków o długości od 16 do 30 bitów,
- Cramera-von Misesa i momentów pierwszego rzędu dla bloków o długości od 16 do 30 bitów,
- Cramera-von Misesa i momentów drugiego rzędu dla bloków o długości od 16 do 29 bitów,
- momentów pierwszego rzędu i Kołmogorowa-Smirnowa K^+ oraz K^- dla bloków o długości od 16 do 30 bitów,
- momentów drugiego rzędu i Kołmogorowa-Smirnowa K^+ oraz K^- dla bloków o długości od 16 do 29 bitów,
- podziałów i pokerowy dla nienakładających się wektorów – $t = 5, d = 3$ oraz $t = 7, d = 4$,
- stosu książek i porządku książek – $d = 11, k = 3$,
- maksimum z t dla nienakładających się wektorów – korelują procedury dla tej samej wielkości bloku i kolejnych $t - 40$ par.

Spśród wymienionych wybraliśmy trzy przypadki, które prezentujemy poniżej.

Pierwszym jest silna ujemna korelacja dla testów momentów pierwszego lub drugiego rzędu i statystyki K^- testu Kołmogorowa-Smirnowa. Przykład dla jednej z par prezentujemy na rysunku VI.12. Warto zaznaczyć, że statystyka K^+ koreluje dodatnio ze zbliżoną siłą z testami momentów, co tłumaczy przyczynę tej obserwacji.



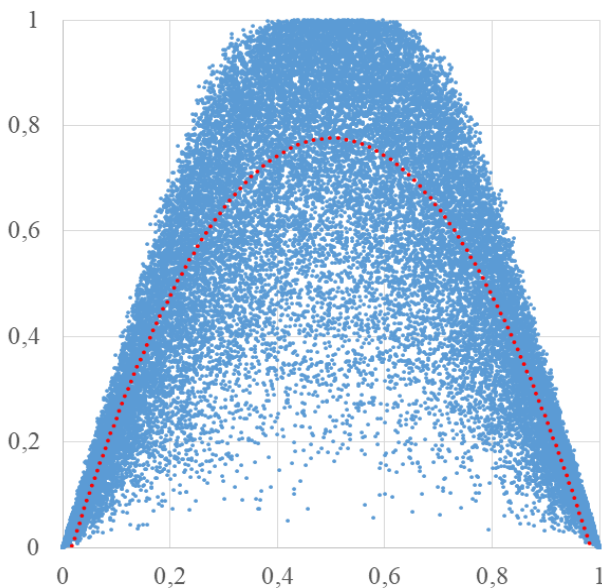
Rysunek VI.12. Regresja liniowa dla pary testów momentów pierwszego rzędu i Kołmogorowa-Smirnowa K^- dla bloku 27-bitowego

Przypadek drugi, który w pierwszej chwili był jedynie ciekawostką i konsekwencją niedopatrzenia, w konsekwencji doprowadził do istotnego poszerzenia wykonanych analiz. Na rysunku VI.12 prezentujemy zależność pomiędzy wektorami p_{val} uzyskanymi dla testu Cramera-von Misesa i momentów pierwszego rzędu dla bloków 26-bitowych.

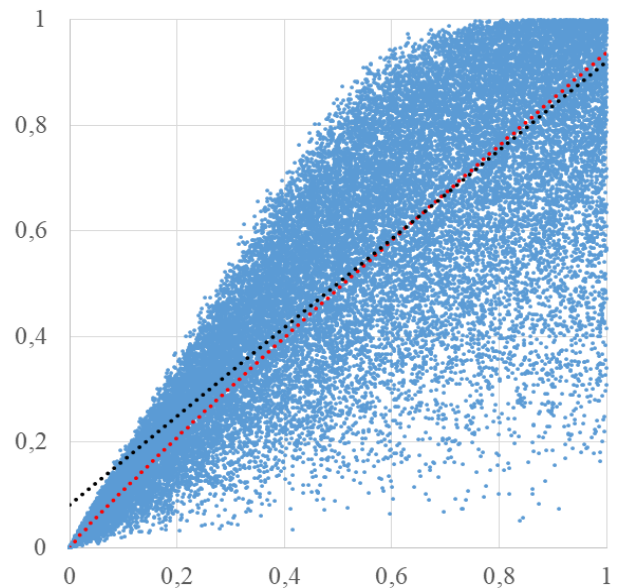
Krzywa regresji ma równanie:

$$U^j = 0,01288(U^i)^3 - 3,35213(U^i)^2 + 3,34017U^i - 0,05580,$$

gdzie U^j są p_{val} dla testu Cramera-von Misesa, zaś U^i są p_{val} testu momentów pierwszego rzędu, współczynnik determinacji wynosi 0,7407.



Rysunek VI.12. Regresja nieliniowa dla pary testów Cramera-von Misesa i momentu pierwszego rzędu dla bloków 26-bitowych



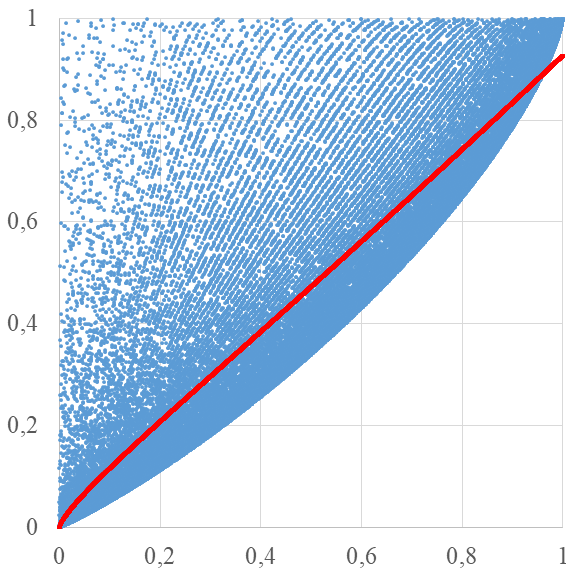
Rysunek VI.13. Regresja liniowa dla pary testów Cramera-von Misesa i przekształconego momentu pierwszego rzędu dla bloków 26-bitowych

Bardzo mały współczynnik przy trzeciej potędze sugeruje wręcz zależność kwadratową i takie jest pierwsze wrażenie patrzącego na rysunek VI.12. Prawdziwą zależność przedstawia jednak

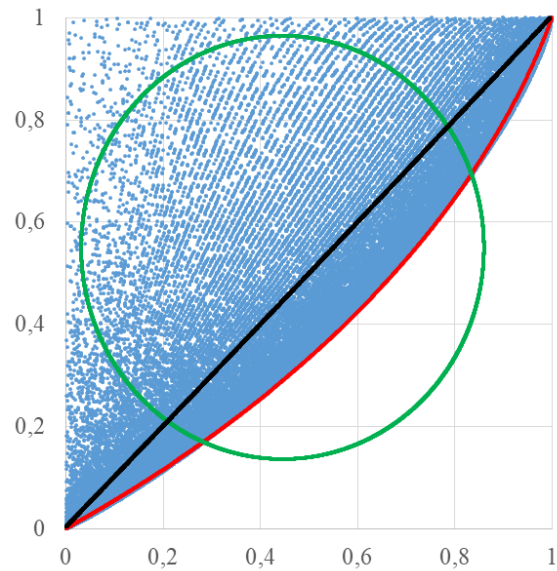
rysunek VI.13, dotyczy on tej samej pary procedur, jednak dla testu momentu pierwszego rzędu wykonano najpierw przekształcenie $U_{||}^i = 1 - 2|U^i - \frac{1}{2}|$. W tym przypadku najlepsze dopasowanie dają regresja liniowa (czarna linia) ze współczynnikiem determinacji 0,7076 oraz potęgowa – 0,7132 (czerwona krzywa), a więc w obu przypadkach wyraźnie słabszym niż wcześniejszy.

Ponieważ celem naszych poszukiwań było znalezienie możliwie wysokich skorelowań, zdecydowaliśmy się poszerzyć obliczenia o zbadanie skorelowania z przekształconymi wektorami, przy czym ograniczyliśmy się jedynie do przypadków, w których wcześniejsze obliczenia dały bardzo niską korelację liniową przy wysokiej korelacji nieliniowej, charakterystyczne dla obserwowanych parabolicznych lub trójkątnych chmur punktów.

Ciekawy układ punktów uzyskaliśmy również dla testów podziałów i pokerowego dla nienakładających się wektorów. Prezentujemy go na rysunku VI.14.



Rysunek VI.14. Regresja nieliniowa dla pary testów pokerowego i podziałów dla nienakładających się wektorów 5 elementów 3-bitowych



Rysunek VI.15. Regresja do okręgu dla pary testów pokerowego i podziałów dla nienakładających się wektorów 5 elementów 3-bitowych

Najwyższą korelację dała tu krzywa:

$$U^j = e^{0,0044 \ln^3 U^i + 0,05549 \ln^2 U^i + 1,00885 \ln U^i - 0,07559},$$

gdzie U^j są p_{val} dla testu pokerowego, zaś U^i – testu podziałów, współczynnik determinacji wynosi 0,75033.

Wykonaliśmy dodatkowe sprawdzenie, czy w takich przypadkach uzasadnione będzie przybliżenie wykorzystujące odcinek okręgu. Na rysunku VI.15, obok punktów odpowiadających korelującym danym, naniesiono trzy krzywe:

- czerwona – fragment elipsy o równaniu:

$$\frac{(U^i + 2,16662)^2}{3,71313} + \frac{(U^j - 1,43672)^2}{10,57039} = 1,$$

wyznaczonej na podstawie punktów $(0; 0)$, $(1; 1)$, $(0,333; 0,20451)$ i $(0,8; 0,64829)$ leżących „na krawędzi” obserwowanej chmury punktów,

- zielona – okrąg:

$$(U^i - 0,44600)^2 + (U^j - 0,54983)^2 = (0,41388)^2,$$

uzyskany jako wynik regresji do okręgu metodą opisaną w punkcie VI.2.1,

- czarna – fragment okręgu:

$$(U^i + 67,201)^2 + (U^j - 68,191)^2 = (95,736)^2,$$

uzyskanego jako wynik 8 kroków metodą regresji geometrycznej do okręgu opisaną w punkcie VI.2.1, dla punktu startowego $(0; 1)$.

Żadna z krzywych widocznych na rysunku VI.15 nie daje możliwości predykcji, ani nawet nie dopasowuje się dobrze do danych:

- krawędź chmury punktów nie jest elipsą,
- standardowa regresja do okręgu daje, z naszego punktu widzenia, rezultat kuriozalny,
- kroki regresji geometrycznej nie zbiegają do żadnego rozwiązania – kolejne dają coraz większe promienie, przyspieszające do nieskończoności,
- z dobrym przybliżeniem czarny łuk można traktować jako odcinek stanowiący granicę ciągu łuków otrzymywanych w kolejnych krokach metody Gaussa-Newtona, jego widoczne na rysunku końce leżą na prostej

$$U^j = 1,002 \cdot U^i + 0,0075.$$

Ostatecznie zdecydowaliśmy, że dla tej pary, jak i innych dających podobny układ punktów, przyjmować będziemy tak otrzymaną prostą jako *najlepiej dopasowaną* do danych, zamiast krzywej widocznej na rysunku VI.14.

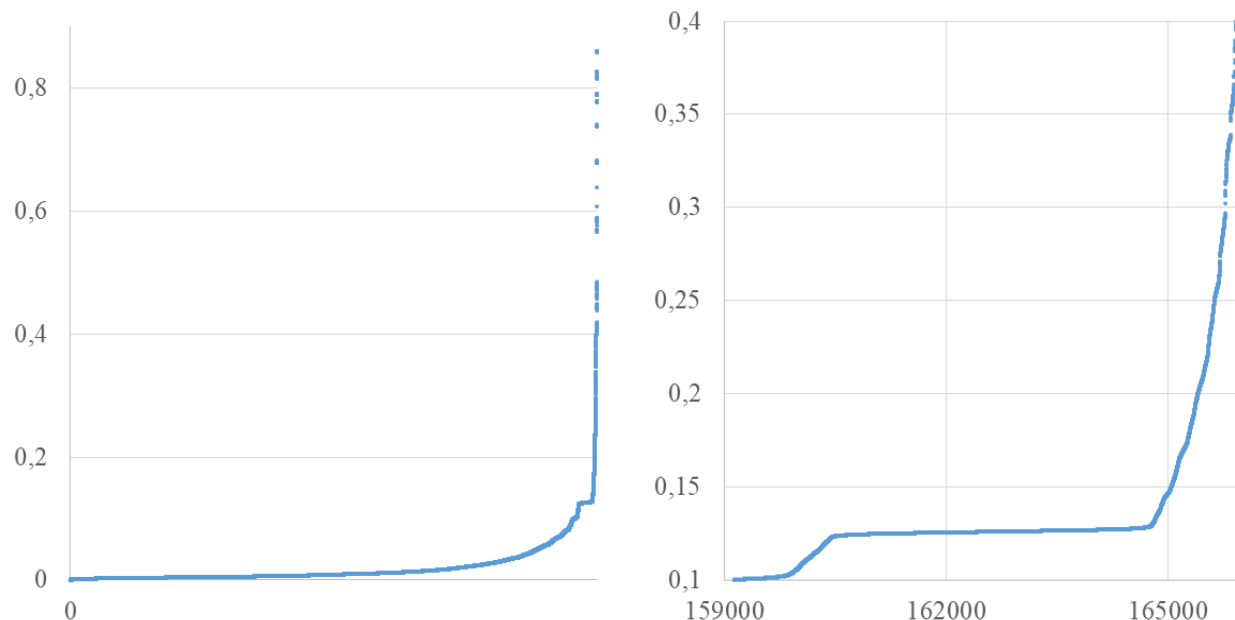
Dodamy jeszcze, że dla 1061 par współczynnik korelacji zawiera się w przedziale $(0,7; 0,8)$, dla kolejnych 1071 jest większy od 0,5, a dla 26053 – większy od 0,1.

2.3.3. Korelacja nieliniowa

Już w punkcie VI.2.3.2 znalazł się przykład pary testów (Andersona-Darlinga i Cramera-von Misesa), dla których współczynnik dopasowania w regresji nieliniowej był wyraźnie wyższy niż dla liniowej. Lepsze dopasowanie, przy bardziej złożonym modelu, jest oczekiwaną obserwacją, istotna będzie dla nas możliwość wskazania przypadków, w których przyjęcie złożonego modelu daje rzeczywistą przewagę. W dalszej części operować będziemy na wartościach bezwzględnych współczynników korelacji.

Z pozostałych na tym etapie 166 616 par, dla 75 008 par współczynnik korelacji któregoś z 11 wariantów nieliniowych jest o co najmniej 0,01 większy od współczynnika korelacji linowej. Dla

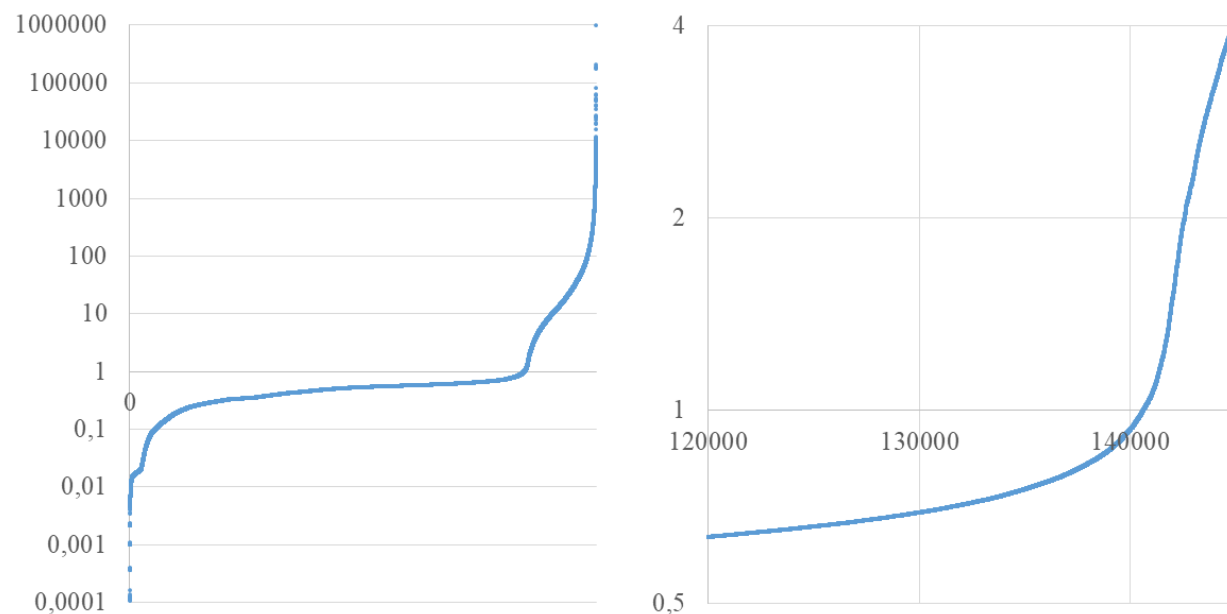
wszystkich par jest wyższy. Najmniejsza poprawa to $6,25 \cdot 10^{-5}$, największa to 0,8604, a przeciętnie wynosi ona 0,02153.



Rysunek VI.16. Wizualizacja poprawy współczynnika korelacji w wyniku zastosowania modelu nieliniowego – poprawa bezwzględna

Na rysunkach VI.16 i VI.17 prezentujemy rozkład poprawy wartości współczynnika korelacji w wyniku zastosowania modelu nieliniowego. W pierwszym przypadku obrazujemy, posortowany niemalejąco, ciąg różnic pomiędzy największym ze współczynników korelacji, a współczynnikiem korelacji liniowej, w drugim – tę poprawę w proporcji do wartości bezwzględnej współczynnika korelacji liniowej. Wykresy po lewej prezentują pełen zakres 166 616 wartości, po prawej fragment zakresu, w którym wydaje się następować jakościowa zmiana tych wskaźników.

Największa względna poprawa wynosi 983 800 razy.



Rysunek VI.17. Wizualizacja poprawy współczynnika korelacji w wyniku zastosowania modelu nieliniowego – poprawa proporcjonalna do współczynnika korelacji liniowej

Analizy przeprowadzone dla danych dla ciągów o długości 1 Mb pokazały bardzo podobny obraz rozkładu siły skorelowań. Na podstawie wykresów, przyjęliśmy następujące kryteria przejścia do modelu nieliniowego:

- bezwzględna poprawa przekracza 0,04,

lub

- poprawa względna przekracza 0,25 i maksymalny współczynnik korelacji przekracza 0,1.

Mamy świadomość, że przyjęte zasady bardzo mocno faworyzowały czysty model liniowy, pamiętać należy jednak, że celem tej części pracy było poszukiwanie takich relacji, które niskim kosztem obliczeniowym pozwalają uzyskać p_{val} pewnej procedury testowej na podstawie p_{val} innej.

2.3.4. Wyniki dla prawdopodobieństw testowych drugiego poziomu dla 128 powtórzeń

W ramach wykonywanych testów przebadaliśmy również korelację uzyskaną dla p_{val} drugiego poziomu. Przypuszczaliśmy, że przy pewnym poziomie skorelowania procedur pierwszego poziomu, procedura dwupoziomowa może wzmocnić tę korelację.

Na bazie próby zgromadzonej na potrzeby pierwotnego badania, wylosowany został milion 128-elementowych zestawów i w ten sposób uzyskaliśmy, dla każdej z rozważanych procedur testowych, nową próbę miliona p_{val} drugiego poziomu. W przeciwieństwie do pierwotnych danych, próba ta obciążona była niewielką korelacją wynikającą z przeciętnie 128-krotnego użycia każdego z p_{val} pierwszego poziomu.

Prawdopodobieństwo, że wybrane dwie t -tki o elementach ze zbioru n -elementowego będą posiadały k identycznych elementów wynosi:

$$p_k = \frac{\binom{t}{k} \binom{n-t}{t-k}}{\binom{n}{t}}.$$

Ponieważ dla $n = 10^6$ i $t = 128$ wyrażenie to sprawia problemy numeryczne, więc posłużyliśmy się następującym oszacowaniem:

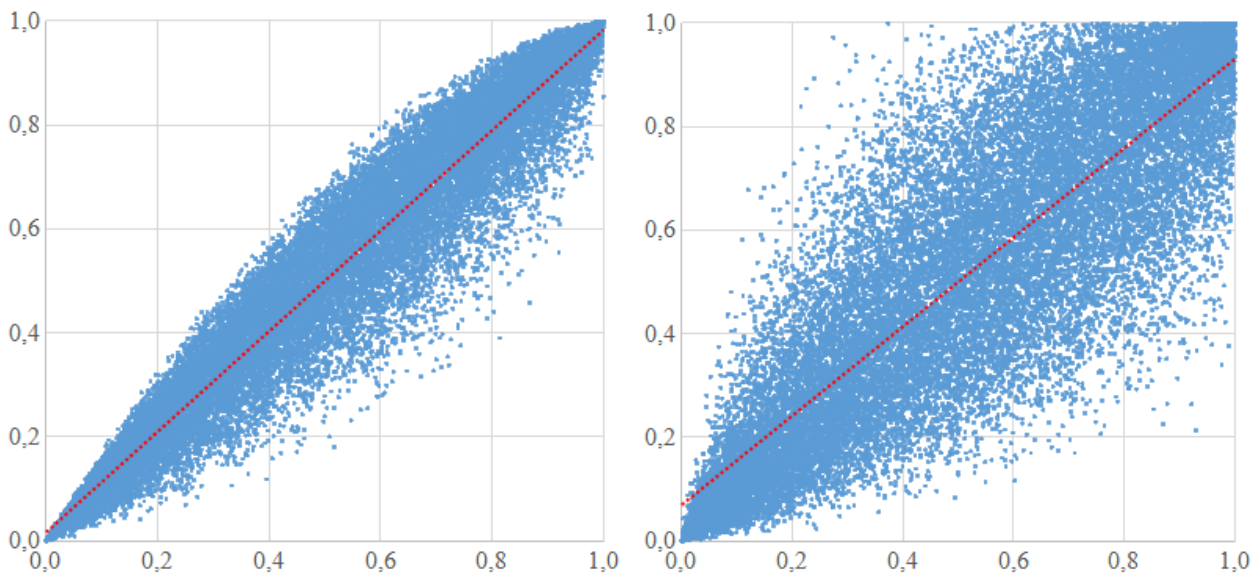
$$\left(1 - \frac{t}{n-t+k+1}\right)^{t-k} \prod_{i=1}^k \frac{(n-i+1)^2}{(n-t+i)i} < p_k < \left(1 - \frac{t}{n}\right)^{t-k} \prod_{i=1}^k \frac{(n-i+1)^2}{(n-t+i)i},$$

które w tych warunkach okazało się być bardzo ciasne.

Uzyskana wartość $p_0 \approx 0,98375$ oznacza, że przytłaczająca większość elementów nowej próby nie jest wzajemnie skorelowana. Z kolei $p_6 \approx 2,0867 \cdot 10^{-14}$, przy liczbie par wynoszącej $5 \cdot 10^{11}$, oznacza, że nie należy spodziewać się więcej niż pięciu powtarzających się elementów w żadnej z par. Na tej podstawie stwierdziliśmy, że zastosowana metoda budowania próby nie powinna wnosić istotnych dodatkowych korelacji.

Uzyskane wyniki okazały się nie być zadowalające. Współczynniki determinacji są zdecydowanie gorsze od pierwotnych.

Najwyższy wynik uzyskała para testów Andersona-Darlinga i Cramera-von Misesa dla bloków 20-bitowych, dla której współczynnik determinacji wyniósł 0,83489. Jako ilustrację osłabienia siły korelacji prezentujemy na rysunku VI.18 regresję liniową dla p_{val} pierwszego i drugiego poziomu dla tej pary testów.



Rysunek VI.18. Regresja liniowa p_{val} pierwszego i drugiego poziomu dla pary testów Andersona-Darlinga i Cramera-von Misesa dla bloków 20-bitowych

Dla omówionej pary, w analogiczny sposób, wyznaczyliśmy korelację dla próbki p_{val} drugiego poziomu przy liczbie powtórzeń wynoszącej 1024. Ponieważ wynik był nieco słabszy od powyższego, zaniechaliśmy kontynuowania badania w tym kierunku.

2.4. Wyniki dla regresji dwuwymiarowej uzyskane dla procedur dla ciągów o długości 1 Mb

2.4.1. Korelacja równa 1

Na podstawie wyników uzyskanych dla ciągów o długości 1 kiB, opisanych w punkcie VI.2.2, większość korelujących procedur została usunięta na etapie doboru zestawu początkowego procedur. Dla 164 procedur testowych zidentyfikowaliśmy łącznie 140 par, w których współczynnik korelacji równy był 1. Są to procedury testów:

- autokorelacji dla ciągu binarnego – przesunięcie o 1 bit,
- całkowitej liczby serii,
- logarytmu prawdopodobieństwa dla wektorów $t = 1, d = 2$,
- podziałów dla nienakładających się wektorów $t = 3, d = 2..4$,
- pokerowego dla nienakładających się wektorów $t = 3, d = 2..4$,
- porządku książek
 - $d = 14..17, k = 12..13$,
 - $d = 18, k = 2$,
 - $d = 19, k = 2..3$,
 - $d = 20, k = 2..4$,
 - $d = 21, k = 2..5$,
 - $d = 22..25, k = 2..6$,
 - $d = 26, k = 2..4$,

- potęgowo-dywergencyjnego dla $\delta = 1$ i $d = 2..12$,
- stosu książek
 - $d = 18, k = 2$,
 - $d = 19, k = 2..3$,
 - $d = 20, k = 2..4$,
 - $d = 21, k = 2..5$,
 - $d = 22..25, k = 2..6$,
 - $d = 26, k = 2..4$,
- zgodności Pearsona
 - $t = 1, d = 2..30$,
 - $t = 2, d = 1..6$,
 - $t = 3, d = 1..4$,
 - $t = 4, d = 1..3$,
 - $t = 5..6, d = 1..2$,
 - $t = 7..30, d = 1$.

Wymienione procedury łączą się w 69 grup, które omówimy poniżej.

Rozpocznijmy od serii podobnych do siebie grup procedur. Przypadki takie omówiliśmy już w punkcie VI.2.2. W każdej grupie podkreślona jest ta z procedur, którą pozostawiliśmy w pakiecie. Mamy procedury testów:

a)

- logarytmu prawdopodobieństwa dla wektorów $t = 1, d = 2$,
- potęgowo-dywergencyjnego dla $\delta = 1$ i $d = 2$,
- zgodności Pearsona
 - $t = 1, d = 2$,
 - $t = 2, d = 1$,

b) dla $i = 3, 5, 7, 11$,

- potęgowo-dywergencyjnego dla $\delta = 1$ i $d = i$,
- zgodności Pearsona
 - $t = 1, d = i$,
 - $t = i, d = 1$,

c) dla $i = 2, 3$,

- potęgowo-dywergencyjnego dla $\delta = 1$ i $d = i^2$,
- zgodności Pearsona
 - $t = 1, d = i^2$,
 - $t = i, d = i$,
 - $t = i^2, d = 1$,

d) dla $i = 3, 4, 5$,

- potęgowo-dywergencyjnego dla $\delta = 1$ i $d = 2i$,
- zgodności Pearsona
 - $t = 1, d = 2i$,
 - $t = 2, d = i$,
 - $t = i, d = 2$,
 - $t = 2i, d = 1$,

e)

- potęgowo-dywergencyjnego dla $\delta = 1$ i $d = 12$,
- zgodności Pearsona
 - $t = 1, d = 12$,
 - $t = 2, d = 6$,
 - $t = 3, d = 4$,
 - $t = 4, d = 3$,
 - $t = 6, d = 2$,
 - $t = 12, d = 1$,

Kolejną serię stanowią pary procedur testu zgodności Pearsona $t = 1, d = i$ oraz $t = i, d = 1$, dla $i = 13..30$.

Jako kolejne zidentyfikowaliśmy:

- 3 pary procedur testów pokerowego i podziałów,
- 33 pary procedur testów porządku książek i stosu książek,

korelujące przy tych samych parametrach.

Ostatnią serię stanowią cztery pary procedur testu porządku książek dla $d = 14..17$ i $k = 12$ oraz 13.

Na koniec pozostała nam para procedur testów autokorelacyjnego dla ciągu binarnego i całkowitej liczby serii. Jeśli zauważymy, że obydwa testy w istocie zliczają ilukrotnie różnią się dwa kolejne bity ciągu, potrzeba eliminacji jednego z nich staje się oczywista.

Dla ciągów o długości 1 Mb udało się zaobserwować różnicowanie się testów zgodności Pearsona dla różnych kombinacji t i d dających ten sam iloczyn. Dla małych wartości iloczynu $t \cdot d$, wszystkie kombinacje są równoważne, gdyż nie następuje grupowanie klas. Dla dużych, grupowanie może być bardzo znaczące i zależeć od kombinacji t i d . Doskonałym przykładem będą tu $t \cdot d = 12$ i 24. W pierwszym przypadku wartość oczekiwana liczby obserwacji dla każdej z możliwych wartości 12-bitowego bloku to $1000000 / 12 / 2^{12} > 20,3$ i grupowanie nie następuje, natomiast dla bloku 24-bitowego wynosi ona zaledwie $1000000 / 24 / 2^{24} < 0,0025$ i muszą być tworzone klasy po co najmniej 4027 wartości bloku.

Ten krok pozwolił na wyeliminowanie 94 procedur.

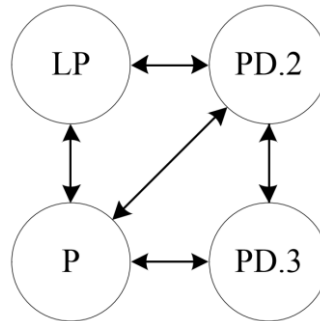
2.4.2. Korelacja bliska 1

Analizując wyniki korelacji wielowymiarowej (patrz punkt VI.4.4.) uznaliśmy, że potraktowanie poniższych jako szczególnych przypadków korelacji wielowymiarowej pozwoli uzyskać lepsze wyniki, toteż w tym punkcie omawiamy jedynie charakter powiązań, ale nie wskazujemy procedur eliminowanych.

Zidentyfikowaliśmy 44 pary procedur, dla których współczynnik korelacji liniowej przekracza próg określony w punkcie VI.2.2. Nie znaleźliśmy przypadku, w którym któraś z nieliniowych korelacji przekraczała ten próg, podczas gdy liniowa nie. Łącznie mamy tu 62 procedury, z których 34 tworzą jedynie pary, lecz zidentyfikowaliśmy również bardziej złożone przypadki.

Największą grupę stanowi ósemka procedur złożona z testów logarytmu prawdopodobieństwa, zgodności Pearsona i potęgowo-dywergencyjnego dla $\delta = 2..7$, dla elementów 6-bitowych. Tworzą one prosty łańcuch skorelowań, w takiej kolejności, w jakiej zostały tu wymienione.

Nieco ciekawszy układ dają procedury testów logarytmu prawdopodobieństwa, zgodności Pearsona i potęgowo-dywergencyjnego dla $\delta = 2$ oraz 3, dla elementów 5-bitowych. Schemat skorelowania ma postać przedstawioną na rysunku VI.19. Warto zauważyć, że jeśli przeprowadzalibyśmy eliminację w tym punkcie, to z tej grupy należałoby pozostawić jedynie jedną procedurę.

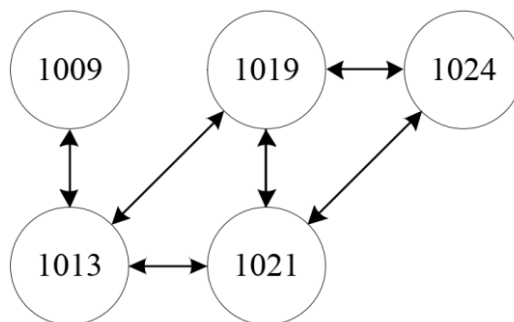


Rysunek VI.19. Graf skorelowania czwórki procedur dla elementów 5-bitowych

Dwie kolejne grupy ponownie charakteryzowało skorelowanie łańcuchowe, złożone były z procedur testu sumy kwadratów odstępów bez zapętlenia dla bloków 26-bitowych, dla odstępów:

- 929, 937, 941, 947 i 953,
- 967, 971, 977, 983, 991 i 997.

Ostatnią grupę również stanowiły procedury testu sumy kwadratów odstępów bez zapętlenia dla bloków 26-bitowych, dla odstępów 1009, 1013, 1019, 1021 i 1024-elementowych, z tą różnicą, że skorelowane były zgodnie ze schematem przedstawionym na rysunku VI.20.



Rysunek VI.20. Graf skorelowania grupy procedur testu sumy kwadratów odstępów bez zapętlenia dla bloków 26-bitowych

Wspomniane 17 par wysoko skorelowanych procedur to procedury testów:

- całkowitej liczby serii oraz całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- logarytmu prawdopodobieństwa dla wektorów $t = 1$ elementów $d = 3$ oraz 4-bitowych oraz zgodności Pearsona dla tych samych parametrów,
- nienakładających się 2-bitowych wzorców w bloku $M = 524288$ bitowym, wzorce 0×1 oraz 0×2 ,

- nienakładających się 3-bitowych wzorców w bloku $M = 524288$ bitowym, wzorce:
 - 0×1 oraz 0×4 ,
 - 0×3 oraz 0×6 ,
- sumy logarytmów odstępów z zapętlaniem 2-elementowych odstępów dla bloków $d = 25..32$ bitowych oraz sumy logarytmów odstępów bez zapętlania dla tych samych parametrów,
- sumy logarytmów odstępów z zapętlaniem 3-elementowych odstępów dla bloków $d = 23..25$ bitowych oraz sumy logarytmów odstępów bez zapętlania dla tych samych parametrów.

2.4.3. Korelacja nieliniowa

W związku z zastosowaniem procedury, opisaniej w punkcie VI.2.1, ograniczającej nakład obliczeń, gromadzone były dane dotyczące jedynie par, dla których skorelowanie osiągnęło wystarczająco wysoki poziom. Ostateczny wybór modelu dokonany został na podstawie formuł podanych w punkcie VI.2.3.3. W związku z tym, w przeciwieństwie do procedur dla ciągów o długości 1 kiB, możemy przedstawić tu jedynie ogólne statystyki.

Spośród 537 717 par procedur, dla których współczynnik korelacji wynosi co najmniej 0,1, jedynie dla ok. 8,5‰ (4 562 przypadki) przyjęty został model nieliniowy. Było to:

- $U^j = e^b (U^i)^a$ – 9-krotnie,
- $U^j = b \cdot a^{U^i}$ – 15-krotnie,
- $U^j = a \ln U^i + b$ – 4454 razy,
- $U^j = a(U^i)^3 + b(U^i)^2 + c(U^i) + d$ – 69 razy,
- $\ln U^j = a \ln^3 U^i + b \ln^2 U^i + c \ln U^i + d$ – 7-krotnie,
- $\ln U^j = a(U^i)^3 + b(U^i)^2 + c(U^i) + d$ – 8-krotnie,
- $U^j = a \ln^3 U^i + b \ln^2 U^i + c \ln U^i + d$ – brak.

2.5. Wpływ długości ciągu na siłę korelacji

Wpływ długości badanego ciągu na siłę korelacji nie był istotny z punktu widzenia celu tej pracy – budowy pakietu testów losowości, tym niemniej zgromadzone dane pozwalają na poczynienie pewnych obserwacji.

Udało nam się zidentyfikować 57 519 par procedur, dla których posiadamy stabilizowane współczynniki korelacji dla obu długości ciągu. Kolejność procedur w parze była brana pod uwagę, porównywane były jedynie wartości dla modelu wybierane wg algorytmu z punktu VI.2.3.3.

W 13 796 odnotowaliśmy wzrost, co do wartości bezwzględnej, współczynnika korelacji. Średnio wynosił on 0,03571, zaś maksymalnie 0,39357 – dla pary testu autokorelacji dla ciągu bloków 12-bitowych, przesunięcie o 324 oraz bloków 6-bitowych, przesunięcie o 648. Jedno spojrzenie na tę parę testów uzmysławia dlaczego są skorelowane i dlaczego wraz z wydłużaniem się ciągu

będzie rosło ich skorelowanie. Różne pary procedur testu autokorelacji zajęły w tym zestawieniu 108 początkowych pozycji. Łącznie znaleźliśmy 8811 par procedur tego testu, w których równe są iloczyny długości bloku i przesunięcia oraz 1400, z różnicą nieprzekraczającą 2, dla których wzrasta współczynnik korelacji.

Na przeciwnym biegunie znalazły się pary procedur testów Andersona-Darlinga, Cramera-von Misesa i Kołmogorowa zestawiane z obiema statystykami testu Kołmogorowa-Smirnowa, notujące spadek współczynnika korelacji sięgający 0,54545. Zajęły one 94 ostatnie pozycje.

Pary procedur testu autokorelacji stanowiły ponad 53% przeanalizowanych w tym punkcie.

Zidentyfikowaliśmy łącznie 16 par testów, dla których wszystkie (lub niemal wszystkie) pary procedur wykazały wzrost skorelowania. W 14 spośród nich występuje test wag Hamminga w ciągu bloków:

- z testami Andersona-Darlinga, Cramera-von Misesa, Kołmogorowa i maksimum z t dla nienakładających się wektorów zarówno w roli procedury wyjaśniającej, jak i wyjaśnianej,
- z testami momentu pierwszego i drugiego rzędu, potęgowo-dywergencyjnym i zgodności Pearsona dla wektorów w roli procedury wyjaśnianej,
- z testem wypełniania drzewa binarnego na podstawie ciągu bitów – zgodności dla pozycji kolizji w roli procedury wyjaśniającej,
- z samym sobą.

Pozostałe dwie pary to:

- testy niezależności wag Hamminga dla bloku M -bitowego i pokerowy dla nienakładających się wektorów,
- wag Hamminga dla bloku M -bitowego i uniwersalny Maurera.

Przytoczone powyżej pary nie wyczerpują wszystkich przypadków wzrostu skorelowania. Bardzo częstą sytuacją był wzrost dla par procedur operujących na tej samej długości bloku, zaś spadek dla pozostałych przypadków.

3. Grupy skorelowanych procedur

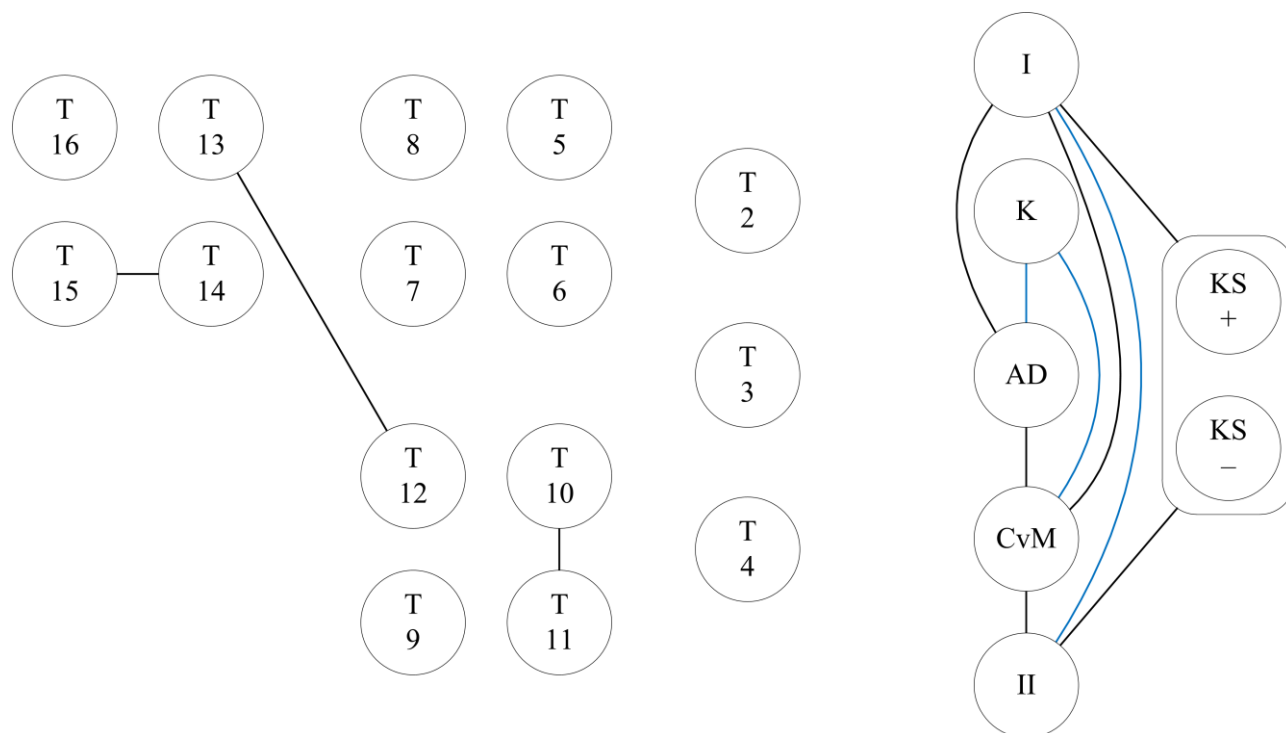
Wydaje się, że ciekawym sposobem zobrazowania korelacji i jej siły będą prezentowane w tym punkcie grafy.

Wierzchołkami grafów są procedury, a łącząca je krawędź występuje, gdy odpowiadający im współczynnik korelacji przekracza określoną progową wartość. Dla uproszczenia posłużyliśmy się grafami nieskierowanymi.

3.1. Grupy skorelowanych procedur dla ciągów o długości 1 kiB

Dla procedur dla ciągów 1 kiB naszą szczególną uwagę zwróciło 16 grup procedur, których grafy korelacji były bardzo podobne. Pojawiały się drobne różnice polegające na występowaniu lub nie pojedynczych krawędzi, co wynikało z przekraczania albo nie założonych progów.

Na rysunkach VI.20 – VI.22 prezentujemy grafy charakteryzujące korelacje dla 16 grup po 22 procedury operujące na blokach długości od 17 do 32 bitów, dla współczynników korelacji nie mniejszych od, odpowiednio, 0,8, 0,7 i 0,5. Na rysunku VI.20 kolorem niebieskim wyróżniono przypadki, gdy współczynnik korelacji przekraczał 0,9.



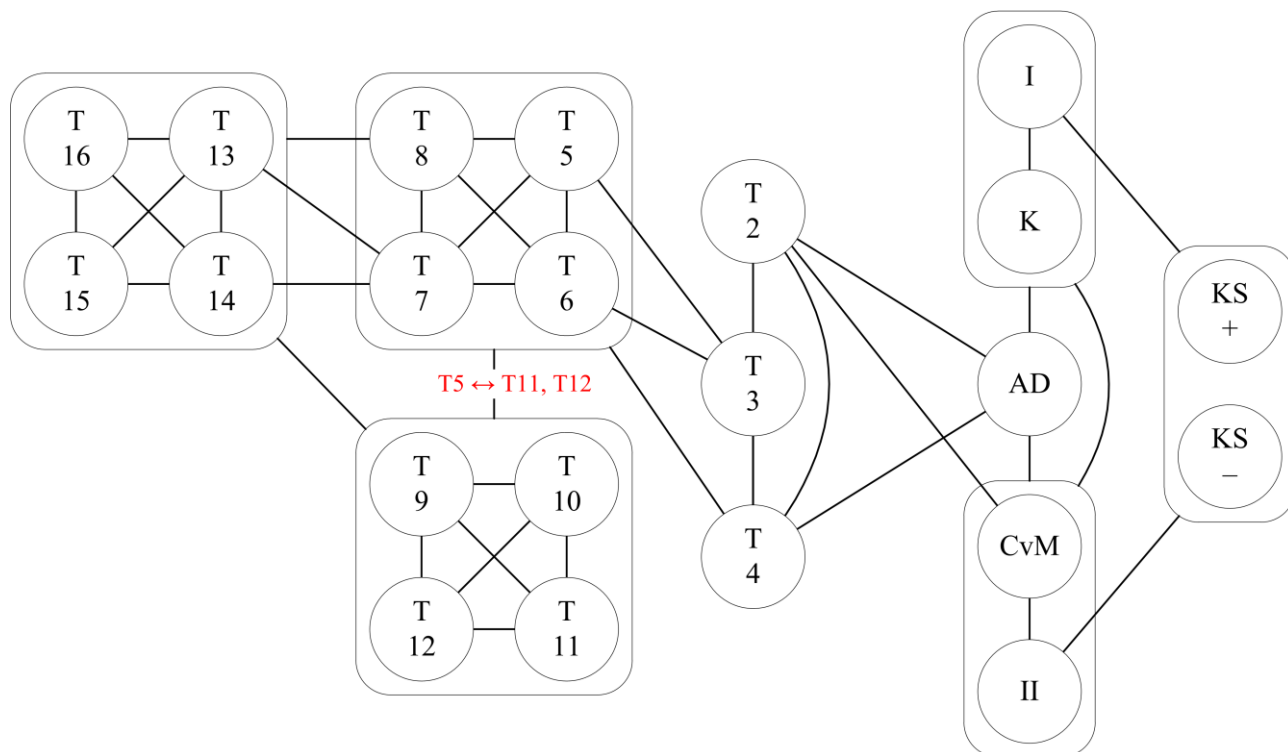
Rysunek VI.20. Graf skorelowania grupy procedur dla ciągów 1 kiB – skorelowanie co najmniej 0,8

Na rysunkach VI.20 – VI.22, poszczególne symbole oznaczają następujące procedury testowe:

- T2 – T16 – maksimum z t dla nienakładających się wektorów $t = 2..16$ elementów d -bitowych,
- I – momentu pierwszego rzędu dla elementów d -bitowych,
- II – momentu drugiego rzędu dla elementów d -bitowych,
- AD – Andersona-Darlinga dla bloków d -bitowych,
- CvM - Cramera-von Misesa dla bloków d -bitowych,
- K – Kołmogorowa dla bloków d -bitowych,
- K⁻ – Kołmogorowa-Smirnowa – statystyka K⁻ dla bloków d -bitowych,
- K⁺ – Kołmogorowa-Smirnowa – statystyka K⁺ dla bloków d -bitowych,

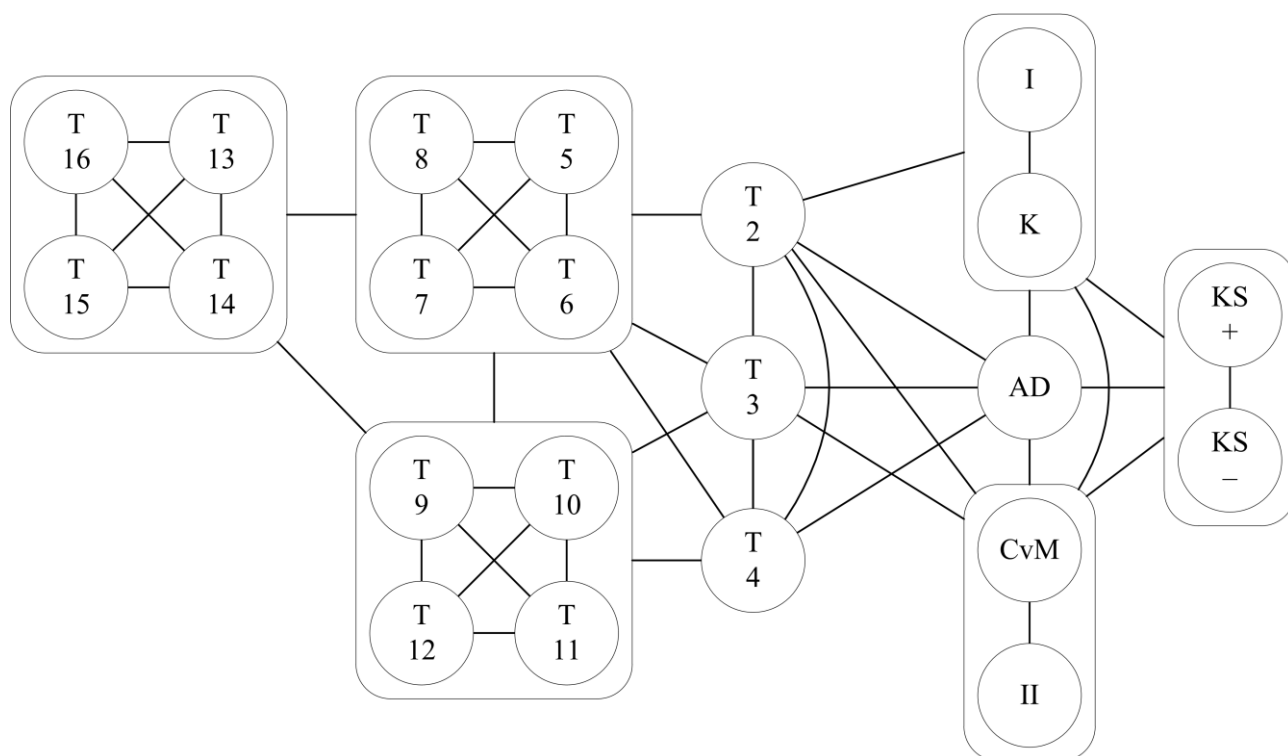
dla jednej z długości bloku $d = 17..32$.

Graf z rysunku VI.20 potwierdza wcześniejsze spostrzeżenie o mocno korelujących ze sobą procedurach testów opartych o dystrybuantę empiryczną, uwidoczniał również silny związek z nimi testów momentów pierwszego i drugiego rzędu.



Rysunek VI.21. Graf skorelowania grupy procedur dla ciągów 1 kiB – skorelowanie co najmniej 0,7

Na rysunkach VI.21 i VI.22 zastosowaliśmy grupowanie wierzchołków. Poza jednym wyjątkiem (pary procedur dla testu Kołmogorowa-Smirnowa) procedury należące do danej grupy skorelowane są ze wszystkimi procedurami z danej grupy oraz tym samym podzbiorem pozostałych procedur – w takich przypadkach krawędź została doprowadzona do obwiedni grupy.



Rysunek VI.22. Fragment grafu skorelowania grupy procedur dla ciągów 1 kiB – skorelowanie co najmniej 0,5

Grafy z rysunków VI.21 i VI.22 wskazują, że poszczególne procedury testu maksimum z t dla nienakładających się wektorów są bardzo mocno powiązane. Dla poziomu 0,5 dwóm podzbiom procedur dla $t = 3..12$ oraz dla $t = 5..16$ odpowiadają grafy pełne. Rolę pomostu pomiędzy tymi

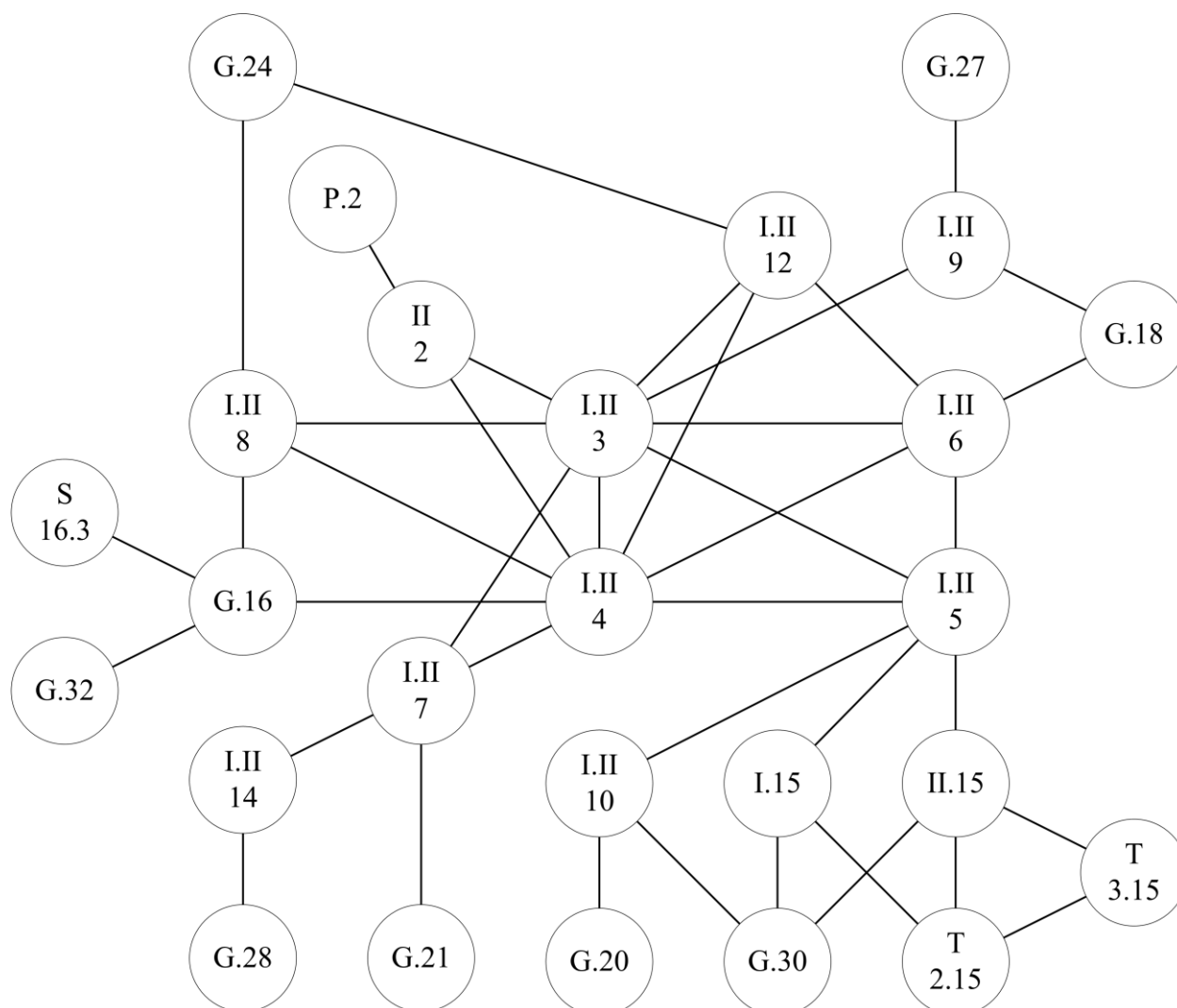
procedurami, a testami wykorzystującymi dystrybucję empiryczną i testami momentów stanowią procedury dla $t = 2, 3$ i 4 .

Ostatni, z przygotowanych grafów, przedstawia skorelowanie wewnątrz grupy 224 procedur. Większą część stanowi 9 grup procedur opisanych grafami jak na rysunku VI.22, odpowiadających blokowi długości $d = 16, 18, 20, 21, 24, 27, 28, 30$ i 32 – reprezentowane na rysunku VI.23 wierzchołkami opisanymi jako $G.d$. Pozostałe wierzchołki odpowiadają procedurom testów momentów pierwszego lub drugiego rzędu oraz:

- S.16.3 – stosu książek dla $d = 16$ i $k = 3$,
- P.2 – zgodności Pearsona dla elementów $d = 2$ bitowych,
- T.2.15 i T.3.15 – maksimum z t dla nienakładających się wektorów $t = 2$ i 3 elementów $d = 15$ bitowych.

Dowiązania poszczególnych grup do reszty grafu nie mają, jak na wcześniejszych rysunkach, postaci każdy-do-każego – korelacje występują dla procedur testu maksimum z t dla $t = 2, 3$ i 4 oraz dla testów momentów pierwszego i drugiego rzędu.

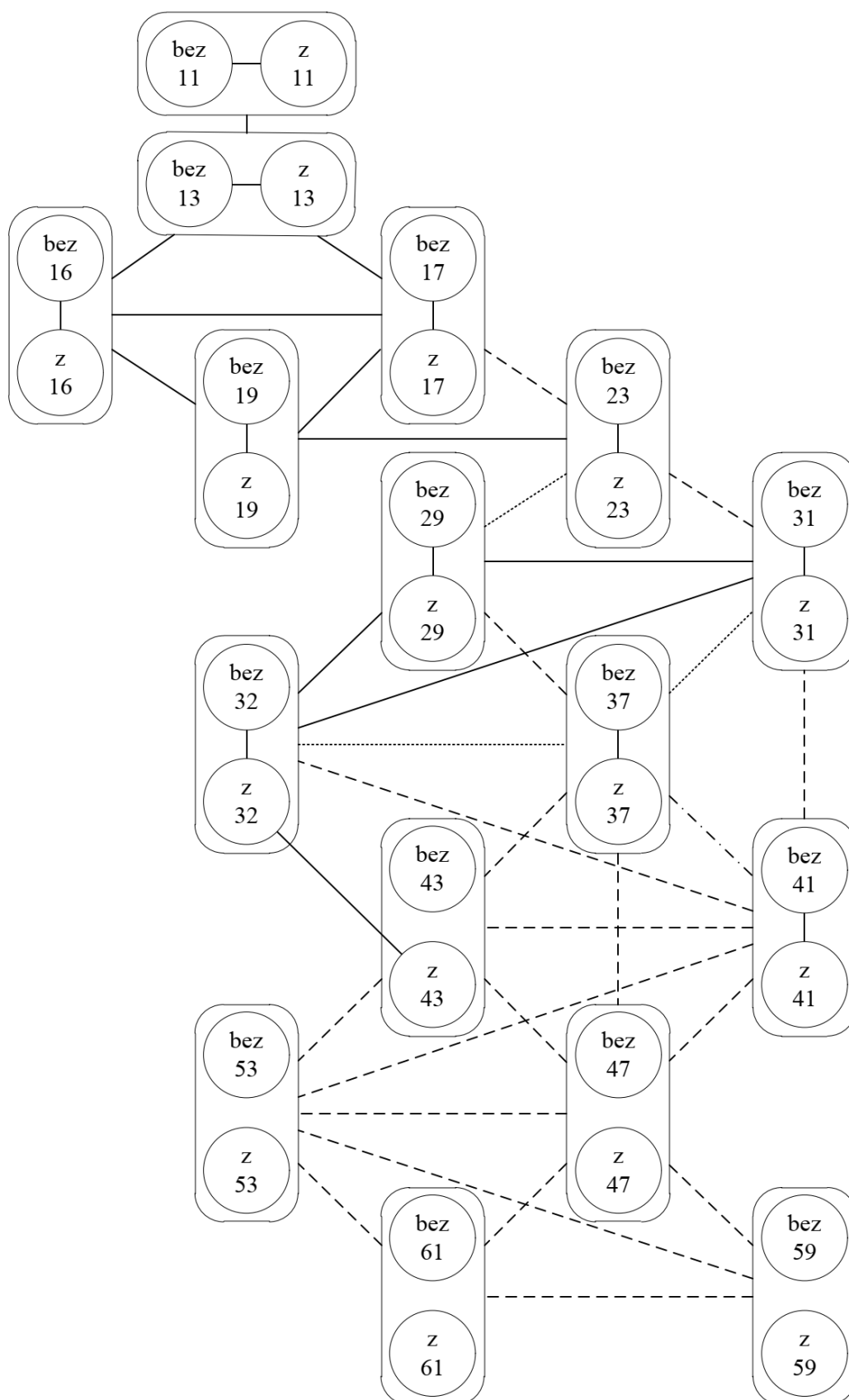
Ogólnym wnioskiem z przeprowadzonej w tym punkcie analizy jest korelowanie procedur operujących na blokach jednakowej długości lub o długościach podzielnych.



Rysunek VI.23. Największa grupa skorelowanych procedur dla ciągów 1 kiB – skorelowanie co najmniej 0,5

3.2. Grupy skorelowanych procedur dla ciągów o długości 1 Mb

Pośród procedur dla ciągów 1 Mb naszą uwagę zwróciły 124 grupy procedur skorelowanych na poziomie co najmniej 0,9 (ale mniej niż 1), których grafy korelacji układały się w opisane poniżej klasy.



Rysunek VI.24. Graf skorelowania grupy procedur testów sumy kwadratów odstępów – skorelowanie co najmniej 0,9, odstępy od 11 do 61, długości bloku od 22 do 32 bitów

Najciekawszy w tej części jest graf z rysunku VI.24 łączący grupy procedur testu sumy kwadratów odstępów „z” i „bez” zapętlenia dla odstępów z zakresu od 11 do 61. Procedury połączyliśmy w pary, dzięki czemu schemat stał się bardziej przejrzysty, za to musieliśmy posłużyć się trzema rodzajami krawędzi:

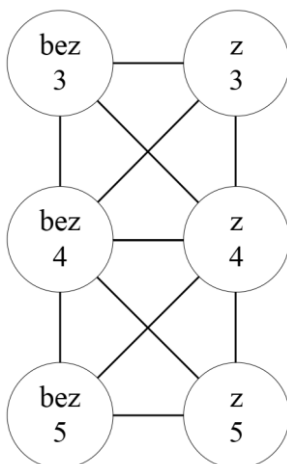
- linia ciągła pomiędzy dwoma węzłami oznacza, że każda procedura skorelowana jest z obiema z drugiej pary,
- linia przerywana – skorelowane są pomiędzy sobą procedury „z” oraz „bez” z obu par,
- linia kropkowana – skorelowane są jedynie procedury „z” z obu par.

Dla wszystkich długości bloku od 22 do 32 obserwujemy takie powiązanie procedur, przy czym dla większych długości bloku występują dodatkowe skorelowane procedury, w pewnym sensie przedłużające sekwencję skorelowań widoczną na rysunku. Dla bloku 31-bitowego kończy się ona dopiero dla odstepu 109-elementowego w teście sumy kwadratów odstępów z zapętleniem.

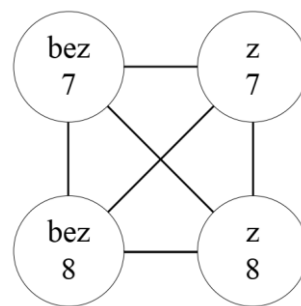
Dla mniejszych wartości odstepu mamy niemal drabinową strukturę skorelowań, natomiast wraz ze zwiększaniem się odstepu przybywa skorelowanych z daną par, maleje jednak siła korelacji.

Dla bloków od 17 do 21 bitów, odpowiednie grafy są mniej lub bardziej okrojonymi wersjami przedstawionego na rysunku VI.24.

Swoistym uzupełnieniem do powyższych są dwie klasy procedur, których grafy przedstawione są na rysunkach VI.25 i VI.26.



Rysunek VI.25. Graf skorelowania grupy procedur testu sumy kwadratów odstępów – skorelowanie co najmniej 0,9, odstepy od 3 do 5, długości bloku od 19 do 32 bitów



Rysunek VI.26. Graf skorelowania grupy procedur testu sumy kwadratów odstępów – skorelowanie co najmniej 0,9, odstepy od 7 do 8, długości bloku od 18 do 32 bitów

Na rysunku VI.25 prezentujemy graf obejmujący 6 procedur testu sumy kwadratów odstępów „z” i „bez” zapętlenia, dla odstępów równych 3, 4, 5. Każdej z długości bloku od 19 do 32 bitów odpowiada jeden taki graf. Bardzo charakterystyczna jest tu swoista nieprzechodniość siły skorelowania – oczywiście procedury dla odstępów równych 3 i 5 również są skorelowane, lecz wartości współczynnika korelacji nie przekraczają 0,8.

Na rysunku VI.26 prezentujemy graf obejmujący kolejne 4 procedury testu sumy kwadratów odstępów „z” i „bez” zapętlenia dla odstępów równych 7 i 8. Przypadek ten jest skromniejszy od poprzedniego, ale zachodzi również dla bloków długości 18 bitów.

Można domniemywać, że jeśli w rozpatrywanym zbiorze znalazłyby się również procedury odpowiadające odstępom równym 6, 9 i 10, to grafy z rysunków VI.24 – VI.26 stanowiłyby graf spójny.

W identyczny, do pokazanego na rysunku VI.26, sposób korelują również czwórki procedur testu sumy logarytmów odstępów 2 i 3-elementowych „z” i „bez” zapętlenia dla bloków od 25 do 32 bitów.

Kolejną istotną klasą skorelowanych procedur są procedury testów:

- zgodności Pearsona,
- potęgowo-dowergencyjnego dla $\delta = 2..6$,
- logarytmu prawdopodobieństwa

dla bloków $d = 5..12$ bitowych. W każdej z 8 grup wszystkie procedury korelują pomiędzy sobą. Powrócimy jeszcze do nich w punkcie VI.4.4.

Dla $d = 4$ powyższe procedury nadal korelują, lecz graf nie jest już pełny.

Ostatnią dużą grupę stanowi 17 procedur testu sumy kwadratów odstępów bez zapętlenia dla bloków 26-bitowych – odstępów: 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021 i 1024. Procedury te korelują na zasadzie każda z każdą i były dla nas największym wyzwaniem podczas eliminacji przy użyciu regresji wielowymiarowej (punkt VI.4.4).

Dla bloków o długościach od 16 do 32 bitów wystąpiły trójki skorelowanych łańcuchowo procedur testów Andersona-Darlinga z Cramera-von Misesa, a ta z Kołmogorowa.

Odpowiadające sobie procedury testów momentów pierwszego i drugiego rzędu również uzyskiwały współczynniki korelacji przekraczające 0,9.

Test nienakładających się wzorców charakteryzuje się dwoma parami korelujących procedur, wynika to z dość oczywistej własności wzorców postaci:

- 0...01 oraz 10...0,
- 1...10 oraz 01...1,

dla których liczby obserwacji w bloku mogą różnić się maksymalnie o 2.

Nie było również zaskoczeniem wysokie skorelowanie procedur testów uniwersalnego Maurera i Corona dla tych samych długości bloku.

Jako ostatnią wymienimy w tym punkcie ciekawą grupę czterech procedur:

- test autokorelacyjny dla ciągu binarnego – przesunięcie o 1 bit,
- test całkowitej liczby serii,
- test całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- test uniwersalny Maurera dla bloku 1-bitowego.

Zabrakło tu jeszcze dwóch procedur, które znajdowały się w początkowym zestawie, ale nie przeszły kroku eliminacji opisanego w poprzednim rozdziale – testów częstości 0 i 1 oraz 0 i 1 łącznie. Wszystkie te procedury wykonują, de facto, analizę rozkładu lub liczby serii w badanym ciągu.

Łącznie, aż 951 procedur jest skorelowanych na poziomie co najmniej 0,9 z przynajmniej jedną procedurą, przy czym tylko 172 z nich tworzy układy stanowiące jedynie parę. Największą grupę stanowi 47 procedur testu sumy kwadratów odstępów dla bloków 31-bitowych.

1054 procedury są skorelowane na poziomie co najmniej 0,8 z przynajmniej jedną inną procedurą, przy czym zaledwie 144 z nich tworzą układy stanowiące jedynie parę. Największą grupę stanowi 59 procedur testu sumy kwadratów odstępów z zapętlaniem i bez dla odstępów od 2 do 109 dla bloków 31-bitowych. Dla pozostałych długości bloków liczby skorelowanych procedur wahają się od 29 do 44.

Na poziomie co najmniej 0,5 skorelowanych jest 2320 procedur, z czego 178 połączyło się w pary. Największa grupa złożona jest aż z 476 procedur poniższych testów (w nawiasach podana jest liczba procedur z danego testu):

- Andersona-Darlinga (10),
- Cramera-von Misesa (10),
- iloczynów dla nienakładających się wektorów (137),
- Kołmogorowa (10),
- Kołmogorowa-Smirnowa (20),
- Kuipera (10),
- logarytmu prawdopodobieństwa (1),
- maksimum z t dla nienakładających się wektorów (136),
- momentu pierwszego rzędu (23),
- momentu drugiego rzędu (22),
- porządku książek (45),
- potęgowo-dywergencyjnego (1),
- stosu książek (31),
- sumy kwadratów odstępów bez zapętlenia (18),
- zgodności Pearsona dla wektorów (2).

Są to bardzo zróżnicowane testy, choć widać też oczywiście podgrupy, które omówiliśmy wcześniej

Na poziomie co najmniej 0,1 skorelowanych jest 253 893 (49%) procedur, z czego 3158 połączone jest w jedną grupę, a 114 854 – w pary.

139 040 procedur znalazło się w skorelowanych grupach o licznosciach powyżej dwóch i choć nie oznacza to, że każda z nich posiadać będzie co najmniej dwie potencjalne procedury wy-

jaśniające, to tak duża liczba wzajemnie wysoko korelujących grup procedur dobrze rokuje w kontekście poszukiwań opisanych w kolejnym punkcie.

4. Wielowymiarowa regresja liniowa

W większości przypadków skorelowanych procedur ujawnionych w poprzednim punkcie, siła skorelowania była zbyt mała, aby na tej podstawie eliminować procedury. W tym punkcie badamy, czy połączenie informacji z kilku procedur pozwoli dokładniej estymować wartości p_{val} wybranej procedury.

4.1. Regresja wielowymiarowa

Poszukiwać będziemy współczynników kombinacji liniowej:

$$U^0 = b_0 + b_1 U^1 + \dots + b_k U^k,$$

gdzie U^i , $i = 0..k$, są kolumnowymi wektorami p_{val} , zaś b_i tzw. parametrami strukturalnymi.

Jeśli określimy macierz $X = [\underline{1} | U^1 | \dots | U^k]_{n \times (k+1)}$, gdzie $\underline{1}$ jest n -elementowym wektorem kolumnowym jedynek, to wektor współczynników b_i wyznaczamy z formuły:

$$[b_i] = (X^T X)^{-1} X^T U^0.$$

Dla wyznaczonej kombinacji obliczany będzie współczynnik determinacji

$$R^2 = \frac{b^T X^T U^0 - n(\bar{U}^0)^2}{(U^0)^T U^0 - n(\bar{U}^0)^2}.$$

Nie będziemy wyznaczać skorygowanego współczynnika determinacji ze względu na bardzo liczną próbę.

4.2. Dobór procedur do kombinacji

Jak pokazaliśmy w punkcie VI.1.1, mamy do czynienia z ogromnymi zbiorami wzajemnie skorelowanych zmiennych losowych. Na sposób doboru konkretnych procedur i ich ról w wielowymiarowym modelu liniowym mają wpływ następujące czynniki:

- niejednorodność, a wręcz ogromne zróżnicowanie, siły korelacji pomiędzy różnymi parami procedur,
- brak symetryczności dla par, dla których najlepszy okazał się model nieliniowy,
- brak pierwotnie określonych ról – zmienne wyjaśniane, zmienne wyjaśniające.

Sposobem przezwyciężenia wszystkich tych trudności okazała się metoda Hellwiga doboru zmiennych objaśniających, do której dodaliśmy jeden stopień swobody w postaci sprawdzenia wszystkich kandydatów do roli zmiennej wyjaśnianej.

Dla danego podzbioru procedur testowych $l_0, l_1, l_2, \dots, l_L$, gdzie l_0 jest numerem procedury wyjaśnianej, natomiast pozostałe $l_1, l_2, \dots, l_L$ – numerami procedur wyjaśniających, wyznaczamy integralną pojemność informacyjną:

$$H_{l_0, l_1, l_2, \dots, l_L} = \sum_{i=1}^L \frac{\bar{\rho}_{l_i, l_0}^2}{\sum_{j=1}^L |\bar{\rho}_{l_j, l_i}|},^{183}$$

gdzie $\bar{\rho}_{l_i, j}$ jest współczynnikiem korelacji liniowej dla odpowiedniego modelu przyjętego zgodnie z metodą opisaną w punkcie VI.2.3.3, w zależności od tego, który z nich ma dla danej pary zastosowanie. Dla procedur wyjaśniających korelacje obliczane są nie pomiędzy pierwotnymi wektorami, a wektorami zastępczymi wyznaczonymi według formuły uzyskanej na podstawie ich skorelowania z procedurą wyjaśnianą.

Dla podzbioru procedur mającego największą pojemność informacyjną wykonywana jest regresja wielowymiarowa opisana w poprzednim punkcie. Jeśli uzyskany współczynnik determinacji okazuje się wystarczający, procedura testowa l_0 jest zapamiętywana jako możliwa do wyeliminowania, a cały proces powtarzany jest dla kolejnej procedury w roli l_0 .

Finalnie otrzymujemy zbiór procedur możliwych do wyeliminowania, wraz z listami ich procedur wyjaśniających. Dla grupy wzajemnie wyjaśniających się procedur, poszukujemy maksymalnie liczego podzbioru procedur możliwych do jednoczesnego wyeliminowania, powtarzając dla nich opisaną powyżej procedurę, lecz dla odpowiednio „zmanipulowanej” listy wyeliminowanych procedur. Spośród wszystkich takich maksymalnych doborów wybierany jest dający największą sumę współczynników determinacji.

4.3. Wyniki regresji wielowymiarowej dla procedur dla ciągów o długości 1 kiB

W tej fazie rozpatrywane było 13 171 procedur, dla których udało się zidentyfikować co najmniej dwie procedury, których współczynnik korelacji przekraczał 0,01. Choć wartość ta, sama w sobie, jest absurdalnie mała, to jednak założyliśmy, że w modelu wielowymiarowym właśnie taki wkład może być brakującym elementem do osiągnięcia zadowalającego poziomu determinacji.

Największą liczbę potencjalnych procedur wyjaśniających, 538, miał test wag Hamminga w ciągu bloków 2-bitowych, liczba wymiarów 3, liczba klas 3. Średnia liczba procedur wyjaśniających wyniosła 25,3, a mediana 12.

W tej grupie procedur udało się znaleźć tylko jeden przypadek, w którym uzyskana wartość współczynnika korelacji była wystarczająco wysoka. Dotyczy on trójki procedur, które omawialiśmy już w punkcie VI.2.3.2:

- logarytm prawdopodobieństwa dla wektorów – $t = 4, d = 4$,
- potęgowo-dywergencyjny – $\delta = 2 - t = 1, d = 4$,

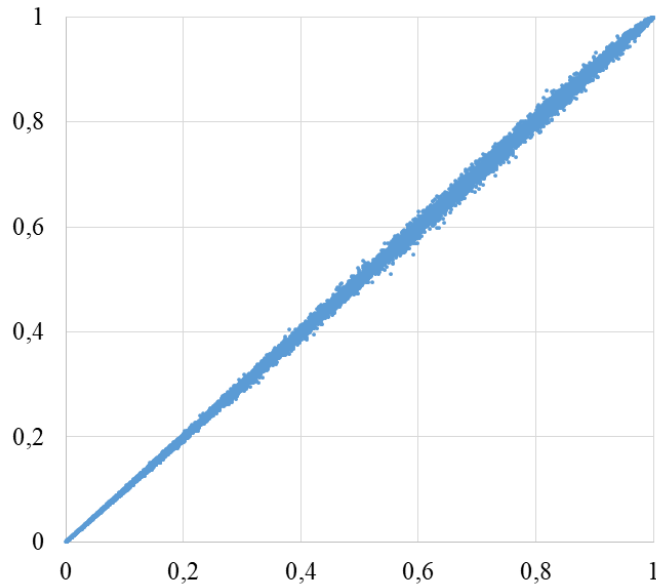
¹⁸³ L. Kowalski, *Statystyka*, BEL Studio, 2001, wydanie I, s. 208

- zgodności Pearsona – $t = 1, d = 4$.

Uzyskaliśmy następującą formułę:

$$U^Z = 0,503384U^P + 0,497344U^L - 0,000234,$$

gdzie U^Z , U^P i U^L są odpowiadającymi p_{val} testów zgodności Pearsona, potęgowo-dywergencyjnego i logarytmu prawdopodobieństwa. Zależność pomiędzy zmienną zastępczą, odpowiadającą prawej stronie powyższej formuły, a p_{val} testu zgodności Pearsona prezentujemy na rysunku VI.27, uzyskany współczynnik korelacji to 0,999998.

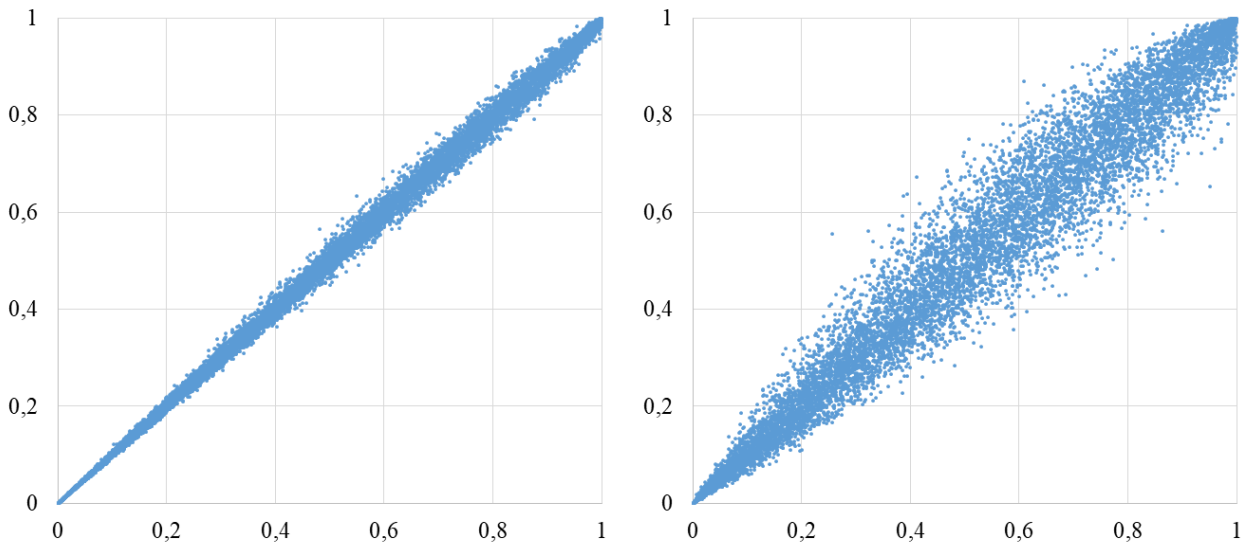


Rysunek VI.27. Regresja liniowa dla testu zgodności Pearsona dla $t = 1, d = 4$ i zastępczej zmiennej wyjaśniającej uzyskanej przy użyciu regresji wielowymiarowej

O tym jak duży zysk dała, w tym przypadku, regresja wielowymiarowa przekonać nas może zestawienie powyższego z wykresem z rysunku VI.7.

W przypadku liniowej korelacji dwuwymiarowej istnieje pełna symetria – jeśli procedura A jest wystarczająco skorelowana z B , to również B jest z A , a zatem można arbitralnie wybrać, którą z nich usunąć z pakietu. Nie jest to już prawdą w modelach nieliniowych, a także w przypadku regresji liniowej wielowymiarowej. Uzasadnienie dla tego stwierdzenia znaleźć można w samej metodzie doboru procedur wyjaśniających, pozwolimy sobie jednak unaocznić to przykładem, gdyż omawiany powyżej przypadek jest do tego idealny.

Równania liniowe, jakim opisana została zależność pomiędzy procedurami testów logarytmu prawdopodobieństwa, potęgowo-dywergencyjnego i zgodności Pearsona, można trywialnie przekształcić tak, aby wyrażało p_{val} dowolnej z procedur w funkcji pozostałych. W ten sposób otrzymujemy jednak wyraźnie słabsze dopasowania, widoczne na rysunku VI.28.



Rysunek VI.28. Próba odwrócenia regresji liniowej dla testu zgodności Pearsona w celu wyjaśniania testów potęgowo-dywergencyjnego i logarytmu prawdopodobieństwa

Dla procedury testu potęgowo-dywergencyjnego optymalny zestaw procedur wyjaśniających składa się z omawianej procedury testu zgodności Pearsona oraz testu autokorelacji dla ciągu bloków 20-bitowych – przesunięcie o 141 bloków. Współczynnik korelacji wynosi 0,999165, a więc jest nieco za niski, gdyż odpowiada mu prawdopodobieństwo nie przejścia testu równomierności przez skorelowaną procedurę trzykrotnie przekraczające założony próg (patrz punkt IV.2.2.). Jest to drugi w kolejności rezultat w tej części.

Dla procedury testu logarytmu prawdopodobieństwa regresja wielowymiarowa okazała się nieefektywna.

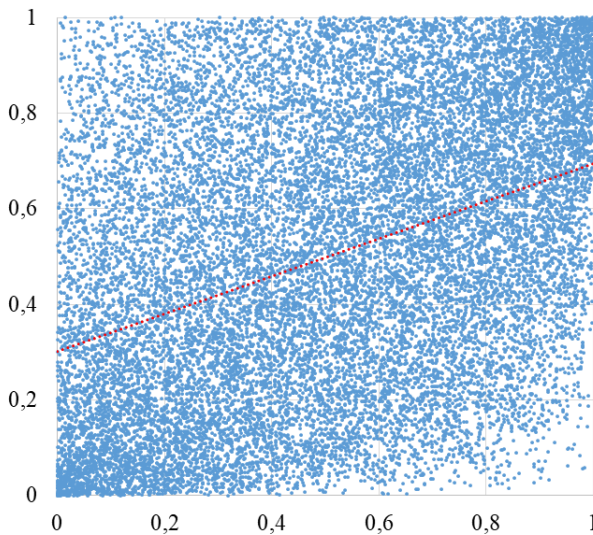
Kolejne miejsca zajęła seria procedur testu Cramera-von Misesa dla bloków o długości od 16 do 32, ze współczynnikami korelacji wyższymi od 0,979, a następnie Andersona-Darlinga dla tych samych długości bloku, ze współczynnikami korelacji powyżej 0,966.

Największy zysk, dzięki regresji wielowymiarowej, uzyskaliśmy dla testu częstości wag Hamminga dla bloku 5-bitowego, wyjaśnianego przez osiem innych:

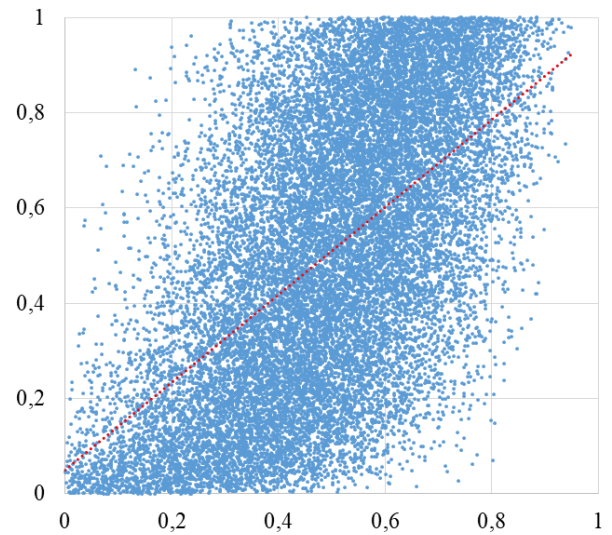
- autokorelacji dla ciągu bloków 3-bitowych – przesunięcie o 577 bloków,
- autokorelacji dla ciągu bloków 11-bitowych – przesunięcie o 253 bloki,
- momentu drugiego rzędu dla elementów 2-bitowych,
- najdłuższej serii 0 w bloku 4-bitowym,
- niezależności wag Hamminga dla bloku 5-bitowego
- wag Hamminga w ciągu bloków 2-bitowych, liczba wymiarów 2, liczba klas 3,
- wag Hamminga w ciągu bloków 5-bitowych, liczba wymiarów 2, liczba klas 2,
- zgodności Pearsona dla wektorów 5-bitowych.

Na rysunku VI.29 prezentujemy zależność pomiędzy p_{val} testów częstości i niezależności wag Hamminga dla bloku 5-bitowego, która jest najwyżej skorelowana z rozpatrywaną. Dla tej pary współczynnik korelacji wynosi 0,39127.

Na rysunku VI.30 prezentujemy zależność pomiędzy p_{val} testu częstości wag Hamminga i zastępczej zmiennej losowej wyznaczonej przy użyciu regresji wielowymiarowej. W tym przypadku współczynnik korelacji wynosi 0,60435. Udało się bardzo wyraźnie zawęzić chmurę punktów, tym niemniej brak jest możliwości predykcji.



Rysunek VI.29. Regresja liniowa dla pary testów niezależności wag Hamminga dla bloku 5-bitowego i częstości wag Hamminga dla bloku 5-bitowego



Rysunek VI.30. Regresja liniowa dla testu częstości wag Hamminga dla bloku 5-bitowego i zastępczej zmiennej wyjaśniającej uzyskanej przy użyciu regresji wielowymiarowej

Największą liczbę procedur wyjaśniających uzyskał test autokorelacji dla ciągu bloków 32-bitowych – przesunięcie o 8 bloków – 50. Przeciętna liczba procedur wyjaśniających to 8,07. Dla 285 procedur nie uzyskano co najmniej dwóch procedur wyjaśniających.

Podsumowując wyniki zaprezentowane w tym punkcie, należy przyznać, że przy użyciu regresji wielowymiarowej udało się wyeliminować jedynie jedną z procedur dla ciągów o długości 1 kiB. Przywołane przykłady pokazały jednak potencjał metody i potwierdziły początkowe przypuszczenie o możliwości zastąpienia części procedur testowych skumulowaną wiedzą wynikającą z przeprowadzenia innych.

4.4. Wyniki regresji wielowymiarowej dla procedur dla ciągów o długości 1 Mb

Punkt ten zaczniemy od kilku danych statystycznych, ilustrujących przekrój otrzymanych wyników. W dalszej części omówimy najlepsze uzyskane dopasowania.

Rozpatrywane było 137 918 procedur, dla których udało się zidentyfikować co najmniej dwie procedury, których współczynnik korelacji przekraczał 0,1.

Największą liczbę potencjalnych procedur wyjaśniających – 149, miał test zgodności Persona dla elementów 2-bitowych. Średnia liczba procedur wyjaśniających wyniosła 3,57, a mediana 3.

Łącznie 172 154 procedury wystąpiły w roli potencjalnych wyjaśniających, najwięcej, bo 445 razy, test momentu pierwszego rzędu dla bloków 1-bitowych, a więc klasyczny test częstości jedynek, a tuż za nim, test momentu pierwszego rzędu dla bloków 2-bitowych. Test momentów pierwszego i drugiego rzędu zajmuje 19 początkowych miejsc w tym zestawieniu i 57 spośród 77 począt-

kowych. Centralne miejsce tych procedur w pakiecie jest dla nas on tyle istotne, że wykorzystana implementacja uwzględnia opracowane przez nas poprawki.¹⁸⁴

Dla 107 650, a więc 78% procedur udało się znaleźć kombinację co najmniej dwóch procedur dającą większą wartość współczynnika korelacji od maksymalnej pojedynczej.

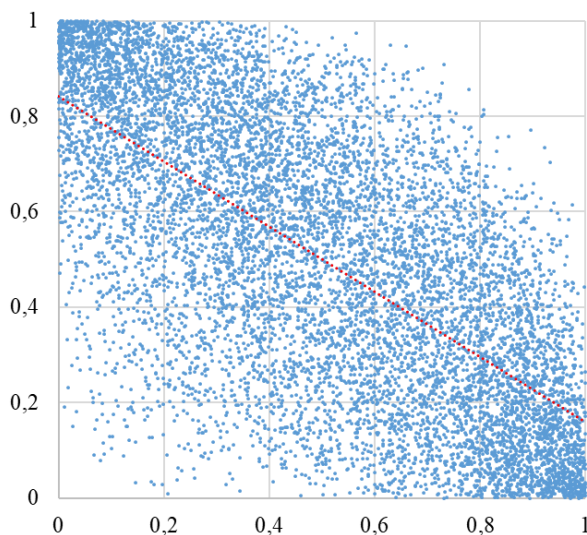
Największą liczbę wykorzystanych procedur wyjaśniających uzyskał ostatecznie test wypełniania drzewa binarnego o wysokości 4 na podstawie ciągu bitów – zgodności dla pozycji kolizji. Było to 48 procedur wywiedzionych z 12 testów. Niestety uzyskany współczynnik korelacji był bardzo niski, wyniósł 0,25219.

Średnia liczba procedur wyjaśniających wyniosła 2,72, zaś mediana była równa 2.

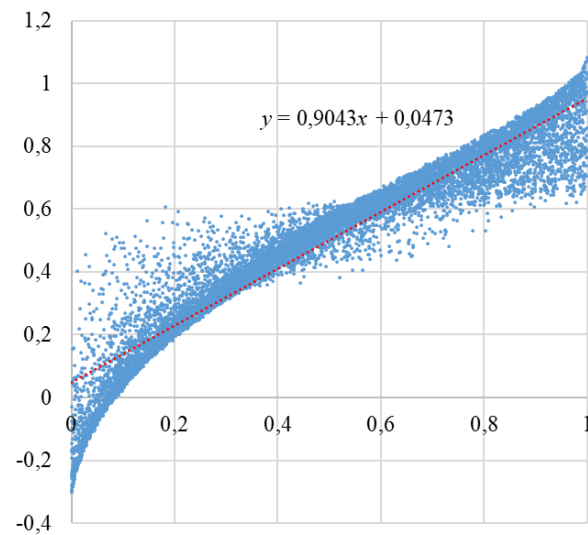
Największy bezwzględny zysk uzyskaliśmy dla testu Kołmogorowa-Smirnowa – statystyka K^- dla bloków 32-bitowych, wyjaśnianego przez:

- test Kołmogorowa-Smirnowa – statystyka K^- dla bloków 16-bitowych,
- test Kołmogorowa-Smirnowa – statystyka K^+ dla bloków 32-bitowych (najmocniej skorelowany z wyjaśnianym),
- test Kuipera dla bloków 32-bitowych,
- test momentu pierwszego rzędu dla bloków 16-bitowych.

Dla zastępczej procedury uzyskaliśmy tu współczynnik korelacji liniowej wynoszący 0,95277, wobec 0,69150 dla najmocniej skorelowanej pojedynczej procedury. Przykład ten posłuży nam również do unaocznienia największego problemu regresji wielowymiarowej, którego analog sygnalizowaliśmy już w punkcie VI.2.2.



Rysunek VI.31. Regresja liniowa dla pary testów Kołmogorowa-Smirnowa dla bloku 32-bitowego – statystyki K^+ i K^-



Rysunek VI.32. Regresja liniowa dla testu Kołmogorowa-Smirnowa – statystyka K^- dla bloku 32-bitowego i niezmodyfikowanej zastępczej zmiennej wyjaśniającej uzyskanej przy użyciu regresji wielowymiarowej

¹⁸⁴ K. Mańk, *Dokładne dystrybuanty statystyk w testach momentów 1 i 2 rzędu*, Biuletyn WAT, Kryptologia IV, Warszawa, 2004

Na rysunku VI.31 prezentujemy regresję liniową pomiędzy testem Kołmogorowa-Smirnowa – statystyka K^- dla bloków 32-bitowych, a najsilniej z nim skorelowanym bliźniakiem ze statystyką K^+ . Korelacja, choć znaczna, jest dalece niewystarczająca w świetle wyników z punktu VI.2.2. Na sąsiednim rysunku VI.32 pokazujemy ponownie regresję liniową, tym razem z niezmodyfikowaną zastępczą procedurą wyjaśniającą, uzyskaną przy użyciu regresji wielowymiarowej.

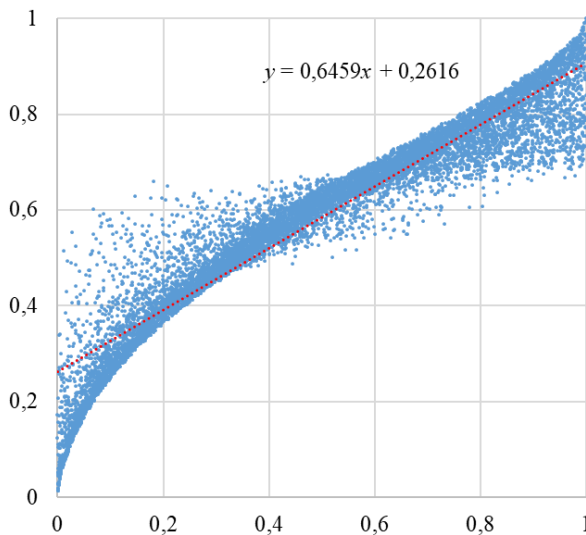
Od razu zauważyć można, że wartości przyjmowane dla procedury zastępczej wykraczają poza przedział $[0;1]$, nie można więc ich traktować jako p_{val} . Rozważyliśmy trzy podejścia na sprowadzenie uzyskiwanych wartości do przedziału $[0;1]$.

Jako pierwsze analizowaliśmy przeskalowanie według formuły:

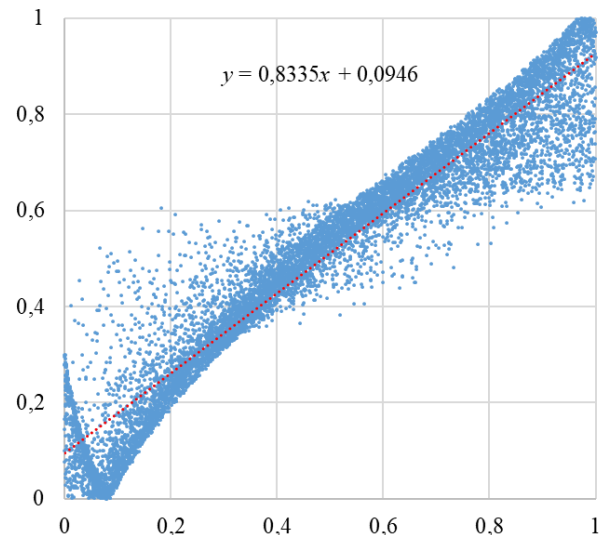
$$z_i^p = \frac{z_i - z_{\min}}{z_{\max} - z_{\min}},$$

gdzie $z_{\min}$ i $z_{\max}$ są odpowiednio najmniejszą i największą możliwą do uzyskania wartością dla procedury zastępczej. Niestety wiążą się z nim dwa poważne problemy – nie znamy wartości $z_{\min}$ i $z_{\max}$ oraz operacja ta wyraźnie osłabia uzyskany wynik, co pokazujemy na rysunku VI.33. Na potrzeby jego sporządzenia przyjęliśmy, że $z_{\min} = \min_i \{z_i\}$ i $z_{\max} = \max_i \{z_i\}$.

Na rysunkach VI.32 i VI.33 nanieśliśmy równania uzyskanych prostych regresji. O ile w pierwszym przypadku jest ono bliskie idealnego: $y = x$, na podstawie drugiego można powiedzieć, że zniweczony został cały zysk z regresji wielowymiarowej.



Rysunek VI.33. Regresja liniowa dla testu Kołmogorowa-Smirnowa – statystyka K^- dla bloku 32-bitowego i zastępczej zmiennej wyjaśniającej uzyskanej przy użyciu regresji wielowymiarowej z przeskalowaniem



Rysunek VI.34. Regresja liniowa dla testu Kołmogorowa-Smirnowa – statystyka K^- dla bloku 32-bitowego i zastępczej zmiennej wyjaśniającej uzyskanej przy użyciu regresji wielowymiarowej z odbiciem

Na rysunku VI.34 pokazujemy rezultat zastosowania drugiego z podejść – odbicia, wykorzystującego przekształcenie znane już z punktu VI.2.2:

$$z_i^o = \begin{cases} -z_i, & z_i < 0, \\ z_i, & 0 \leq z_i \leq 1, \\ 2 - z_i, & z_i > 1. \end{cases}$$

Jest to najprostsza w realizacji metoda i wydaje się być skuteczniejsza od poprzedniej. Co więcej, w praktyce będzie wykorzystywana w sytuacji znacząco lepszego dopasowania, co ograniczy wpływ dokonywanych modyfikacji na otrzymywane przybliżenie.

Ostatnie z rozważonych podejść polega na obcięciu wartości do przedziału $[0;1]$ poprzez zastępowanie wartości przekraczających zakres odpowiednio 0 albo 1:

$$z_i^c = \begin{cases} 0, & z_i \leq 0, \\ z_i, & 0 < z_i < 1, \\ 1, & z_i \geq 1. \end{cases}$$

Choć w prezentowanym przypadku daje ono najlepszy rezultat, to odrzuciliśmy je, gdyż miałyby destrukcyjny wpływ na wyniki testu drugiego poziomu – stosowany w tej roli test Andersona-Darlinga jest bardzo wrażliwy na pojawianie się 0 i 1 w wektorze p_{val} .

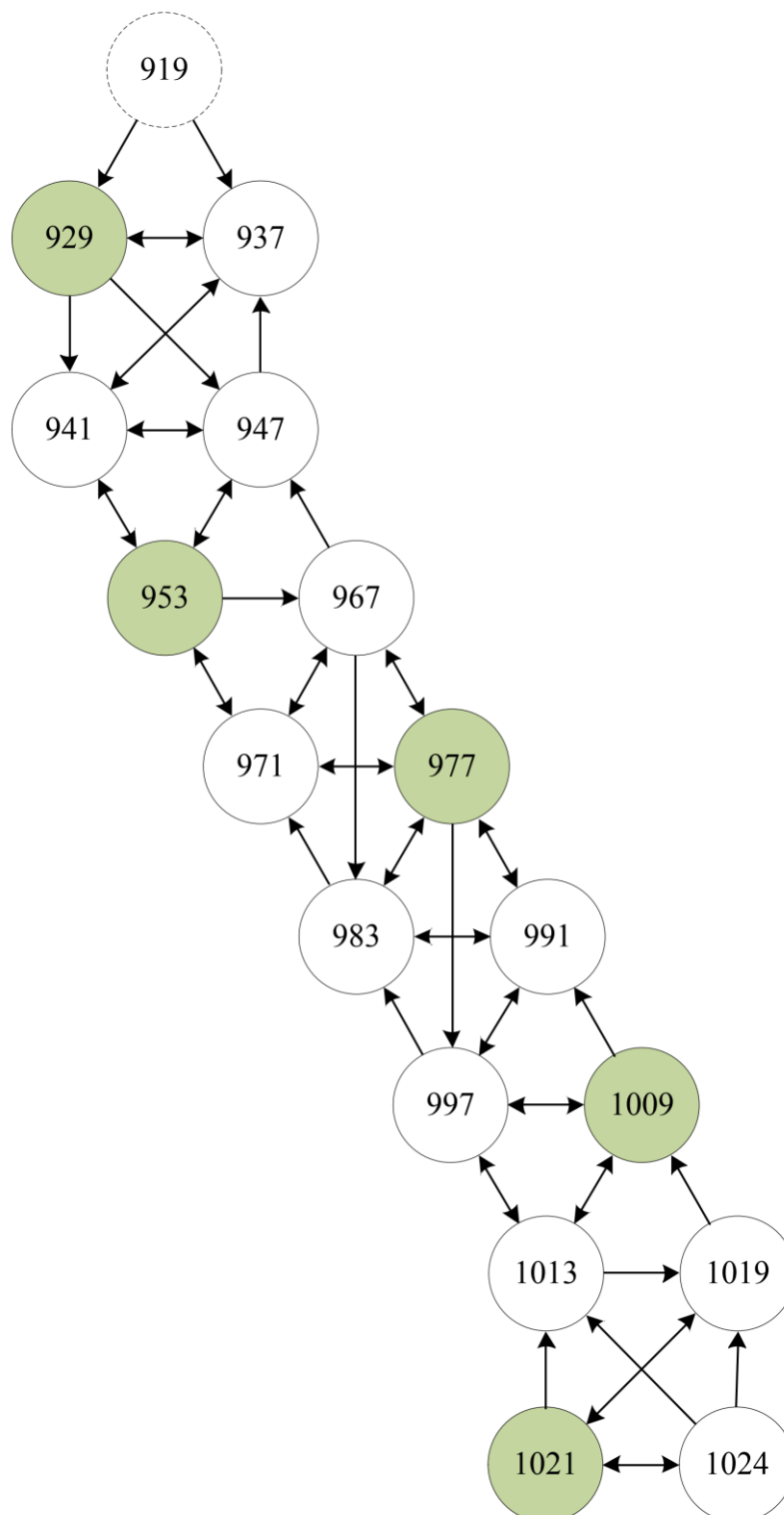
Średni zysk – różnica pomiędzy współczynnikiem korelacji uzyskanym z wykorzystaniem regresji wielowymiarowej, a najlepszym wynikiem z regresji dwuwymiarowej, wyniósł 0,048, średni względny zysk – 16,9%.

Udało się znaleźć 96 procedur, dla których wartość współczynnika korelacji (dwu- lub wielowymiarowej) była wystarczająco wysoka. Połączyły się one w 13 grup, w których procedura wyjaśniana pojawiała się również jako procedura wyjaśniająca dla innej lub innych w tej grupie, 16 wzajemnie wyjaśniających się par oraz jedną wyjaśnianą przez dwie inne.

W przypadku grup, musieliśmy wykonać dodatkowe sprawdzenia w celu ustalenia optymalnego zestawu eliminowanych procedur. Pierwotne poszukiwania skupione były na wyznaczeniu optymalnego zestawu procedur wyjaśniających, co nie wykluczało istnienia innych zestawów, pozwalających uzyskać wystarczająco wysoką wartość współczynnika korelacji.

Najliczniejszą grupę stanowiło 16 procedur testu sumy kwadratów odstępów bez zapętlania dla bloków 26-bitowych, dla odstępów 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021 i 1024-elementowych, które splotły się poprzez wzajemne współwyjaśnianie. Na rysunku VI.35, na podstawie pierwotnych wyników regresji wielowymiarowej, zobrazowaliśmy zależności występujące pomiędzy nimi oraz dodatkową procedurą tego testu dla odstepu równego 919. Warto zaznaczyć, że tylko 4 spośród 16 uzyskanych współczynników determinacji nie było równych 1 (z dokładnością rozdzielczości 64-bitowej arytmetyki zmiennoprzecinkowej), a ich suma była jedynie o $3,8 \cdot 10^{-5}$ niższa od 16. Jednocześnie, za pomocą tylko tego zestawu wyjaśnień, moż-

liwe byłyby wyeliminowanie maksymalnie pięciu procedur – jeden z takich zestawów zaznaczyliśmy na rysunku VI.35 kolorem zielonym.



Rysunek VI.35. Graf obrazujący zależności: procedura wyjaśniana – wyjaśniająca w regresji wielowymiarowej dla grupy procedur testu sumy kwadratów odstępów bez zapętlenia dla bloków 26-bitowych

Niewielka modyfikacja oprogramowania pozwoliła nam znaleźć zdecydowanie efektywniejszą kombinację formuł wyjaśniających. Możliwe okazało się wyeliminowanie aż 13 procedur, dla 84 różnych trójek procedur pozostawianych. Spośród nich wybraliśmy tę, która oferowała największą sumę współczynników determinacji – pozostawiliśmy procedury dla odstępów 947, 997 i 1021. Obok nich, w roli wyjaśniających, wystąpiła procedura dla odstępu równego 919.

Formuły wyrażające p_{val} poszczególnych procedur mają postać:

- $U^{929} = 0,640908U^{919} + 0,359437U^{947} - 0,000172$,
- $U^{937} = 0,352599U^{919} + 0,647746U^{947} - 0,000172$,
- $U^{941} = 0,210016U^{919} + 0,790237U^{947} - 0,000126$,
- $U^{953} = -0,110103U^{919} + 1,048380U^{947} + 0,061751U^{997} - 0,000014$,
- $U^{967} = -0,165127U^{919} + 0,769463U^{947} + 0,579658U^{997} - 0,184133U^{1021} - 0,000071$,
- $U^{971} = -0,154657U^{919} + 0,659619U^{947} + 0,707944U^{997} - 0,213047U^{1021} - 0,000073$,
- $U^{977} = 0,129280U^{919} + 1,306520U^{997} - 0,435563U^{1021} - 0,000115$,
- $U^{983} = -0,162823U^{947} + 1,093560U^{997} - 0,256259U^{1021} - 0,000058$,
- $U^{991} = 0,052710U^{947} + 1,092920U^{997} - 0,145606U^{1021} - 0,000013$
- $U^{1009} = 0,497078U^{997} + 0,503159U^{1021} - 0,000118$,
- $U^{1013} = 0,329575U^{997} + 0,670634U^{1021} - 0,000105$,
- $U^{1019} = U^{1021}$,
- $U^{1024} = U^{1021}$.

gdzie U^j oznacza p_{val} dla procedury odpowiadającej odstępowi równemu j .

Ponieważ w przypadku dwóch ostatnich procedur najlepszy wynik dała liniowa regresja dwuwymiarowa, to te procedury zostaną wyeliminowane z pakietu, natomiast wykonanie pozostałych zastąpione zostanie obliczaniem powyższych formuł.

Przyjmując, że najgorszą możliwą kombinacją jest uzyskanie granicznej wartości korelacji we wszystkich 13 regresjach, to przedstawiona kombinacja jest zaledwie o niecałe 6,3% gorsza od pierwotnej. Najgorsza ze wspomnianych 84 byłaby o dalsze 18,4% słabsza.

Analiza formuł dla procedur wyjaśnianych przez te same zestawy procedur pozwala dość dobrze identyfikować ich położenie względem procedur wyjaśniających. Bliższym sobie wartościom odstępów odpowiadają wyższe współczynniki uzyskane z regresji. Dwoma skrajnymi przypadkami są tu odstęp 953 i 1009. Dla pierwszego z nich największy współczynnik jest niemal dziesięciokrotnie większy od drugiego w kolejności – można więc powiedzieć, że stanowi jedynie małą poprawkę. Jeśli jednak weźmiemy pod uwagę, że odpowiada on odstępowi 947 – bardzo bliskiemu omawianemu, pozostałe dwa zaś stosunkowo odległym, to implikacja staje się oczywista. Z kolei 1009 lokuje się dokładnie w połowie pomiędzy odstępami 997 i 1021, co skutkuje wartościami współczynników bliskimi 0,5.

Procedury testu sumy kwadratów odstępów utworzyły cztery kolejne grupy.

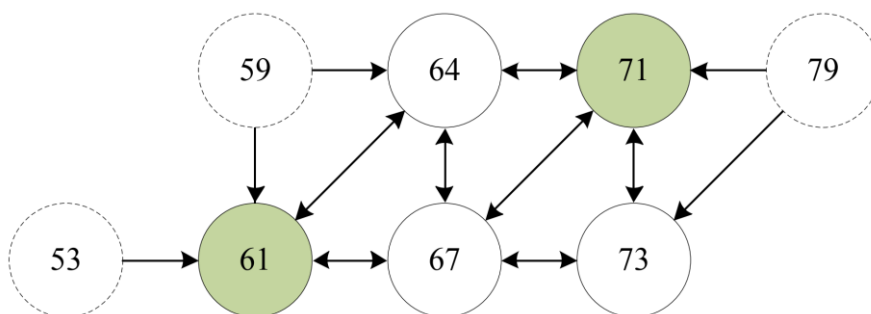
Testy sumy kwadratów odstępów bez zapętlania dla przesunięć 61- i 64-elementowych dla bloków 31-bitowych wyjaśniały się wzajemnie przy wykorzystaniu czterech kolejnych procedur te-

go testu. Niestety niemożliwe okazało się jednocześnie wyjaśnienie obu, w związku z czym wybór padł na pierwszą z nich ze względu na większą wartość współczynnika determinacji.

Identyczna sytuacja wystąpiła dla tej samej pary przesunięć w teście z zapętlaniem dla bloków 32-bitowych.

Testy sumy kwadratów odstępów z zapętlaniem, dla bloku 31-bitowego, dla przesunięć 61, 64, 67, 71 i 73-elementowych powiązane były relacjami współwyjaśniania jak pokazujemy na rysunku VI.36. Trzy dodatkowe procedury należą do tego samego testu, odpowiadają przesunięciom 53, 59 i 79. W tym przypadku, pomimo wdrożonych poszukiwań, nie udało się nam znaleźć możliwości eliminacji więcej niż dwóch spośród wymienionych pięciu procedur. Dla dwóch zaznaczonych na rysunku zielonym kolorem, sumaryczny współczynnik determinacji był najwyższy. Ich wykonanie zastąpią poniższe formuły:

- $U^{61} = -0,052396U^{53} + 0,619383U^{59} + 0,592723U^{64} - 0,160118U^{67} - 0,000206$,
- $U^{71} = -0,131845U^{64} + 0,446600U^{67} + 0,773307U^{73} - 0,088446U^{79} - 0,000194$.



Rysunek VI.36. Graf obrazujący zależności procedura wyjaśniana – wyjaśniająca w regresji wielowymiarowej dla grupy procedur testu sumy kwadratów odstępów z zapętlaniem dla bloków 31-bitowych

Kolejna grupa trzech powiązanych procedur również należała do tego testu. Były to przesunięcia równe 101, 103 i 107, uzupełniane przez procedury dla przesunięć 97 i 109. Niemożliwe okazało się wyeliminowanie wszystkich trzech procedur. Pozostać musiała środkowa – był to jedyny sposób na eliminację dwóch procedur. Ponieważ pierwotnie były one powiązane łańcuchowo, więc udało się wykorzystać najlepsze kombinacje procedur wyjaśniających dla obu eliminowanych, uzyskując relacje:

- $U^{101} = 0,324348U^{97} + 0,676807U^{103} - 0,000579$,
- $U^{107} = 0,324922U^{103} + 0,676118U^{109} - 0,000521$.

Jako następne omówimy dziesięć grup procedur operujących na blokach o długości od 3 do 12 bitów.

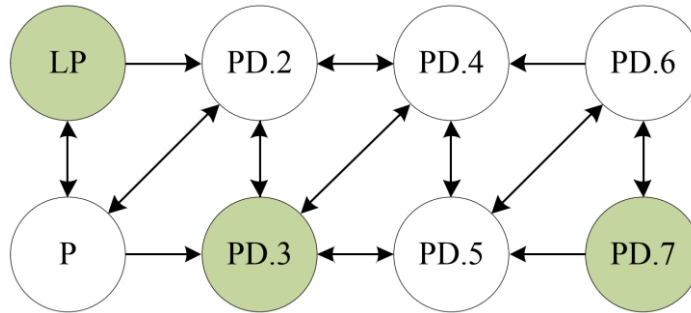
Dwie pierwsze, zaledwie dwuelementowe grupy, stanowią pary testów zgodności Pearsona i logarytmu prawdopodobieństwa dla $d = 3$ oraz 4, wysoko liniowo skorelowane. W obu przypadkach niemożliwe jest ich osobne wyjaśnienie, a więc zdecydowaliśmy o eliminacji obu procedur testu logarytmu prawdopodobieństwa.

Dla $d = 5$ grupa, oprócz testów zgodności Pearsona i logarytmu prawdopodobieństwa, składa się też z dwóch procedur testu potęgowo-dywergencyjnego dla $\delta = 2$ oraz 3. Najefektywniejsza okazała się eliminacja, przy użyciu testu zgodności Pearsona trzech pozostałych procedur.

Globalnie drugą pod względem wielkości była grupa testów operujących na blokach 6-bitowych:

- zgodności Pearsona – $t = 1, d = 6$ – oznaczony jako P,
- potęgowo-dywergencyjny – $\delta = 2..7 - t = 1, d = 6$ – oznaczane jako PD2..7,
- logarytm prawdopodobieństwa dla wektorów – $t = 1, d = 6$ – oznaczony jako LP.

Ich powiązania zobrazowaliśmy na rysunku VI.37. Kolorem zielonym oznaczyliśmy trzy procedury możliwe do wyeliminowania na bazie pierwotnych danych.



Rysunek VI.37. Graf obrazujący zależności procedura wyjaśniana – wyjaśniająca w regresji wielowymiarowej dla grupy procedur operującej na blokach 6-bitowych

Tak jak dla wcześniejszych grup, przeprowadziliśmy wtórną analizę możliwości eliminacji poszczególnych procedur. Ostatecznie uzyskaliśmy cztery kombinacje pozwalające na eliminację sześciu procedur, spośród których największą sumę współczynników determinacji dawało pozostawienie testu potęgowo-dywergencyjnego dla $\delta = 6$ i logarytmu prawdopodobieństwa. Dla eliminowanych procedur uzyskaliśmy następujące relacje:

- $U^{6P} = U^{6LP}$,
- $U^{6PD.2} = 0,334099U^{6PD.6} + 0,666423U^{6LP} - 0,000250$,
- $U^{6PD.3} = 0,501058U^{6PD.6} + 0,499542U^{6LP} - 0,000304$,
- $U^{6PD.4} = 0,667768U^{6PD.6} + 0,332771U^{6LP} - 0,000280$,
- $U^{6PD.5} = U^{6PD.6}$,
- $U^{6PD.7} = U^{6PD.6}$.

Trzy procedury można więc całkowicie wyeliminować z pakietu, a wykonanie trzech kolejnych zastąpić obliczeniem powyższych formuł.

Trzecie miejsce ex aequo zajęły analogiczne do powyższego zestawu testów operujących na blokach 7 oraz 8-bitowych. Choć pierwotne wyniki regresji wielowymiarowej każdą z tych grup rozbiły na dwie części, to bazując na wynikach dla $d = 6$, pogłębioną analizę prowadziliśmy na całych ósemkach.

W obu przypadkach najefektywniejsze okazało się pozostawienie dwóch procedur testu potęgowo-dywergencyjnego dla $\delta = 4$ i 5 oraz testu logarytmu prawdopodobieństwa. Dla $d = 7$ uzyskaliśmy następujące relacje:

- $U^{7P} = 0,250621U^{7PD.4} + 0,749861U^{7LP} - 0,000249$,
- $U^{7PD.2} = 0,501064U^{7PD.4} + 0,499586U^{7LP} - 0,000348$,
- $U^{7PD.3} = 1,502790U^{7PD.4} - 0,602514U^{7PD.5} + 0,0997217U^{7LP} - 0,000003$,
- $U^{7PD.6} = -0,997838U^{7PD.4} + 1,997510U^{7PD.5} + 0,000181$,
- $U^{7PD.7} = -1,992110U^{7PD.4} + 2,991120U^{7PD.5} + 0,000543$.

Z kolei dla $d = 8$ poniższe:

- $U^{8P} = 0,251315U^{8PD.4} + 0,749800U^{8LP} - 0,000595$,
- $U^{8PD.2} = 0,502248U^{8PD.4} + 0,499239U^{8LP} - 0,000797$,
- $U^{8PD.3} = 1,505600U^{8PD.4} - 0,605084U^{8PD.5} + 0,0994773U^{8LP} - 0,000003$,
- $U^{8PD.6} = -0,996343U^{8PD.4} + 1,995600U^{8PD.5} + 0,000402$,
- $U^{8PD.7} = -1,986020U^{8PD.4} + 2,983790U^{8PD.5} + 0,001205$.

Podobieństwo obu zestawów formuł jest bardzo duże. Część współczynników jest, w odpowiadających sobie wzorach, zgodna z dokładnością do trzech miejsc po przecinku.

Dla $d = 9$ sytuacja przedstawiała się nieco inaczej, najprawdopodobniej decydujący wpływ na to miała eliminacja dwóch procedur testu potęgowo-dywergencyjnego już w poprzednim rozdziale. Spośród wyjściowego zestawu złożonego z procedury testu zgodności Pearsona i trzech procedur testu potęgowo-dywergencyjnego dla $\delta = 2..4$ udało się, na podstawie poniższych formuł, wyeliminować trzy:

- $U^{9P} = 0,502199U^{9PD.2} + 0,498640U^{9LP} - 0,000441$,
- $U^{9PD.3} = 0,666323U^{9PD.2} + 0,335322U^{9PD.5} - 0,000869$,
- $U^{9PD.4} = 0,332286U^{9PD.2} + 0,669354U^{9PD.5} - 0,000868$,

posiłkując się dodatkowo kolejną procedurą testu potęgowo-dywergencyjnego dla $\delta = 5$.

Z tych samych powodów, jeszcze skromniejsza była grupa dla $d = 10$. W jej skład weszły procedura testu zgodności Pearsona i dwie procedury testu potęgowo-dywergencyjnego dla $\delta = 2$ oraz 3. Spośród nich udało się wyeliminować pierwszą i trzecią:

- $U^{10P} = 0,504825U^{10PD.2} + 0,497033U^{10LP} - 0,000966$,
- $U^{10PD.3} = 0,498536U^{10PD.2} + 0,503243U^{10PD.4} - 0,000929$.

Dla $d = 11$ okazało się możliwe wyjaśnienie obu pierwotnie wskazanych procedur, dla testu zgodności Pearsona otrzymaliśmy:

- $U^{11P} = 0,345815U^{11PD.3} + 0,662009U^{11LP} - 0,004030$,

zaś dla testu potęgowo-dywergencyjnego dla $\delta = 2$:

- $U^{11PD.2} = 0,681915U^{11PD.3} + 0,325688U^{11LP} - 0,003918.$

Najłatwiejszy i najskromniejszy w tej serii grup był przypadek dla bloku 12-bitowego, w którym jedyną możliwą do eliminacji była procedura testu potęgowo-dywergencyjnego dla $\delta = 2$:

- $U^{12PD.2} = 0,491293U^{12P} + 0,516266U^{12PD.3} - 0,003873.$

Jako następne omówimy 11 par testu sumy logarytmów odstępów bez i z zapętlaniem wzajemnie się wyjaśniających dla następujących kombinacji parametrów:

- 2-elementowy odstęp dla bloków $d = 23..25$ bitowych,
- 3-elementowy odstęp dla bloków $d = 25..32$ bitowych.

Dla wszystkich par możliwe było wyeliminowanie tylko jednej z procedur, wobec czego wybór padł na procedury z zapętlaniem.

Na koniec pozostała para silnie liniowo skorelowanych testów: autokorelacyjnego dla ciągu binarnego – przesunięcie o 1 bit oraz całkowitej liczby serii z uwzględnieniem proporcji 0 i 1. Ta procedura testu autokorelacyjnego posłużyła nam już w punkcie VI.2.4.1 do eliminacji drugiego z testów całkowitej liczby serii. Eliminacja obu procedur nie jest możliwa, ponownie pozostawiamy test autokorelacyjny.

Łącznie, przy użyciu regresji dwu- i wielowymiarowej, udało się wyjaśnić ponad 150 procedur, w większości zostały one wyeliminowane z pakietu, a wykonanie pozostałej około $\frac{1}{3}$ można było zastąpić obliczeniem prostej kombinacji liniowej p_{val} uzyskanych dla innych procedur.

5. Podsumowanie analizy skorelowania procedur

Analizy opisane w tym rozdziale jednoznacznie wykazały istnienie skorelowania dla niemal wszystkich rozważanych procedur, co więcej w wielu przypadkach siła skorelowania jest tak duża, że uzasadnia rezygnację z wykonywania części z nich.

Obserwacje dotyczące grup wzajemnie korelujących procedur doprowadziły nas do poszukiwania korelacji wielowymiarowych. Zaproponowana przez nas metoda pozwala na dalsze ograniczenie nakładu obliczeniowego, bez zawężania zakresu testowania, poprzez wyznaczanie p_{val} wyjaśnianej procedury na podstawie p_{val} innych procedur pakietu.

W punktach 5 załączników A i B zamieściliśmy zestawy procedur rekomendowanych do badania ciągów o długościach 1 kiB i 1 Mb.

Podsumowanie

Podsumowanie rozpoczniemy od kilku zdań na temat pakietu testów TeStat, który powstawał równoległe do tej pracy. Wszystkie testy losowości, których zachowanie było analizowane zostały zaimplementowane w języku `c++` na jego potrzeby. W kilku przypadkach autorami implementacji byli studenci Wydziału Cybernetyki Wojskowej Akademii Technicznej uczestniczący wówczas w realizacji projektu ROTOR, również część implementacji mojego autorstwa powstała w ramach realizacji tego projektu.

Pakiet implementuje 20 głównych bibliotek oraz kilka pomocniczych, objętość kodów źródłowych to przeszło 79 MiB, zawartych w przeszło 157 tysiącach linii kodu.

Od samego początku pakiet projektowany był pod kątem uzyskania wysokiej wydajności podczas wykonywania wielopoziomowego badania losowości ciągów binarnych. Zrezygnowaliśmy z monstrualnego logowania danych pośrednich charakterystycznego dla pakietu NIST, czy też uniwersalności interfejsu, aby „każdy mógł tego użyć” – co widoczne jest w Diehardzie. TeStat przeznaczony jest dla świadomych programistów uczestniczących w badaniach prymitywów kryptograficznych, dla których pakiet testów losowości jest pierwszym, stosunkowo niekłopotliwym sitem, pozwalającym zastąpić pracę analityka brutalną mocą jednostek obliczeniowych. W zamian za te „niedogodności” pakiet oferuje gigantyczny zestaw procedur testowych, sprawdzonych pod kątem jakości uzyskiwanych p_{val} i zredukowanej nadmiarowości.

Nadrzędnym celem była wiarygodność uzyskiwanych wyników, toteż już na etapie gromadzenia specyfikacji poszczególnych testów staraliśmy się docierać do możliwie aktualnych wyników ich dotyczących, wielokrotnie wyznaczając powtórnie podawane w literaturze parametry rozkładów i inne stałe, tak aby osiągać wysoką precyzję wyznaczanych p_{val} . W kilku przypadkach udało się uzyskać również wyniki teoretyczne.

Za najistotniejsze wyniki uzyskane niejako przy okazji budowania pakietu uznajemy:

- wyznaczenie dokładnych wartości parametrów rozkładu statystyki testowej w testach momentów pierwszego i drugiego rzędu dla ciągu bitów,¹⁸⁵
- uogólnienie test autokorelacyjnego dla ciągu binarnego i wyznaczenie parametrów rozkładu,¹⁸⁶
- współopracowanie metody wyznaczenia parametrów rozkładu w teście wolnych cel,¹⁸⁷

¹⁸⁵ K. Mańk, *Dokładne dystrybuanty statystyk w testach momentów 1 i 2 rzędu*, Biuletyn WAT, Kryptologia IV, Warszawa, 2004, s. 47-65

¹⁸⁶ K. Mańk, *Test autokorelacyjny dla ciągu binarnego*, Studia Bezpieczeństwa Narodowego, IV, 6, Kryptologia i Cyberbezpieczeństwo, Warszawa, 2014

¹⁸⁷ T. Jarocki, *Wyznaczenie dokładnych momentów pierwszego i drugiego rzędu w testach wolnych cel Marsagli*, praca dyplomowa, WAT, 2009

- przyspieszenie i uproszczenie obliczania statystyki testowej w teście częstości dla nakładających się wektorów,¹⁸⁸
- przyspieszenie realizacji testów najmniejszej odległości, m najbliższych par i Bickela-Breimana,¹⁸⁹
- propozycja i wyznaczenie rozkładu testu dla 2 i 3 liczb,
- wyznaczenie dokładnych wartości parametrów dla testu Maurera,¹⁹⁰
- wyznaczenie parametrów rozkładu statystyki testowej w teście pochodnej ciągu liczb,¹⁹¹
- opracowanie metody wyznaczania rozkładów w testach wypełniania drzewa binarnego na podstawie ciągu bitów i ciągu liczbowego.¹⁹²

Potyczki z testami losowości rozpoczęliśmy od zgromadzenia materiału źródłowego i poddania go możliwie wnikliwej i krytycznej analizie, co staraliśmy się odzwierciedlić w trzech początkowych rozdziałach pracy, począwszy od absolutnych podstaw w rozdziale I, poprzez opisanie poszczególnych testów losowości i zauważonych możliwościach ich poprawy w rozdziale II, a kończąc na zajrzeniu w potencjał „konkurencji” w postaci przeglądu wybranych pakietów, zawartego w rozdziale III.

Trzy kolejne rozdziały to opis, najistotniejszych w kontekście tej rozprawy, metod wykorzystanych do zbudowania pakietu testów losowości, pokrywającego możliwie szeroki zakres cech przypisywanych ciągowi losowemu, wiarygodnego i jednoznacznego w sensie uzyskiwanych wydyktów oraz zredukowanej korelacji pomiędzy wykorzystywanymi procedurami testowymi.

W dwóch dołączonych do pracy załącznikach zestawiliśmy listy procedur testowych, które mogą, naszym zdaniem, stanowić propozycje pakietów przydatnych dla różnych wariantów wykorzystania. W szczególności zestawy zawarte w punktach 5 obu załączników są najistotniejszym praktycznym wynikiem tej pracy.

Uważny czytelnik zauważył, że w wielu przypadkach posługiwaliśmy się tylko wybranymi wielkościami bloków czy przesunięć, takimi jak liczby pierwsze i potęgi dwójki, wyszliśmy z prostego założenia – jeśli test funkcjonuje poprawnie dla wartości parametrów m i $m + 2$, to będzie poprawnie funkcjonował dla $m + 1$. Miało to oszczędzić nieco czasu obliczeniowego, a stało się przyczynkiem do mniej spektakularnych wyników części korelacyjnej. Tym niemniej, wyniki uzyskane w rozdziale VI dla tak okrojonego zestawu pokazały, że w wielu przypadkach parametry

¹⁸⁸ K. Mańk, *Efficient Evaluation of Frequency Test for Overlapping Vectors Statistic*, CECC 2024, Book of Abstracts, M. Kutyłowski, J. Pomykała, Warszawa, 2024

¹⁸⁹ K. Mańk, *Speeding Up Minimum Distance Randomness Tests*, Journal of Telecommunications and Information Technology, 2022, vol. 2

¹⁹⁰ K. Mańk, *Wyznaczenie dokładnych wartości parametrów dla testu Maurera*, Biuletyn WAT, LVII, 4, Warszawa, 2008

¹⁹¹ K. Mańk, *Wyznaczenie rozkładu średniej wartości odstępów w teście pochodnej ciągu liczb*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2017

¹⁹² K. Mańk, *Upgrading Frequency Test for Overlapping Vectors and Fill Tree Tests*, IJET, 2/2025

wzięte zostały zbyt gęsto, a zaproponowana przez nas metoda może być z powodzeniem stosowana w procesie budowania pakietu testów losowości.

Biorąc pod uwagę powyższe, nie będzie zaskoczeniem, że jednym z planowanych kolejnych kroków będzie, realizowane kosztem wolnego czasu jednostki obliczeniowej, rozwijanie zbudowanego pakietu poprzez włączenie, odfiltrowanie i odkorelowanie nowych testów i procedur dla ciągów długości 1 kiB i 1 Mb. W następnym kroku pakiet powinien zostać poszerzony o procedury dla ciągów długości ok. 10 MB, charakterystyczne dla pakietu DIEHARD.

Pozornie podobnym, lecz w istocie realizacji wyraźnie różniącym się zagadnieniem jest konstrukcja pakietów testów typu on-line, przeznaczonych do implementacji w strukturach programowalnych i monitorujących w sposób ciągły funkcjonowanie fizycznych źródeł losowości. Ograniczone badania, poczynione w trakcie realizacji kilku prac badawczych, pokazały występowanie analogicznych zjawisk – począwszy od porażającego braku zgodności częstości sygnalizacji z założonym poziomem istotności, na 100% zgodności momentów sygnalizacji kończąc.

Podsumowując samą pracę, mamy nadzieję, że jej struktura i zawartość sprawnie przeprowadziła czytelnika przez etapy wykonanych prac, które doprowadziły nas do tego punktu.

Bibliografia

- [AES01] *Advanced Encryption Standard (AES)*, Federal Information Processing Standards, 2001, doi:10.6028/NIST.FIPS.197
- [A04] A. Alhakim, *On the Eigenvalues and Eigenvectors of an Overlapping Markov Chain*, Probability Theory and Related Fields, 128, 2004, s. 589-605
- [AKM04] A. Alhakim, J. Kawczak, S. Molchanov, *On the class of nilpotent Markov chains, I: The spectrum of covariance operator*, Markov Processes and Related Fields 10, 2004, s. 629-652.
- [A08] A. Alhakim, *A Spectral Look at the Overlapping Serial Test*, 2008
- [AD52] T.W. Anderson, D.A. Darling, *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes*, Annals of Mathematical Statistics, 23, 2, 1952, s. 193-212
- http://projecteuclid.org/DPubS/Repository/1.0/Disseminate?view=body&id=pdf_1&handle=euclid.aoms/1177729437
- [Bt37] M.S. Bartlett, *Properties of Sufficiency and Statistical Tests*, Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences, vol. 160, no. 901, 1937, s. 268-282
- [Ba04] L. Bassham, *Validation Testing and NIST Statistical Test Suite*, 2004
- [BB83] P.J. Bickel, L. Breiman, *Sums of Functions of Nearest Neighbor Distances, Moment Bounds, Limit Theorems and a Goodness of Fit Test*, The Annals of Probability, 11, 1, 1983, s. 185-214
- [B95] P. Billingsley, *Probability and Measure*, 3rd ed., John Wiley & Sons ISBN 0-471-00710-2, 1995, p. 357, www.colorado.edu/amath/sites/default/files/attachedfiles/billingsley.pdf
- [BM82] M. Blum, S. Micali, *How to generate cryptographically strong sequences of pseudo random bits*, 23rd Annual Symposium on Foundations of Computer Science (sfcs 1982), Chicago, IL, USA, 1982, s. 112-117
- [BSW04] G. Boero, J. Smith, K.F. Wallis, *Decomposition of Pearson's chi-squared test*, Journal of Econometrics 123, Elsevier, 2004, s. 189-193
- [B03] M. Boesgaard, M. Vesterager, T. Pedersen, J. Christiansen, O. Scavenius. *Rabbit: A New High-Performance Stream Cipher*, Proceedings Fast Software Encryption 2003, Lecture Notes in Computer Science 2887, Springer, 2003, s. 307-329
- [BA14] S. Bouflass, A. Azhari, *On a new properties of random number sequences, a randomness test and a new RC4's key scheduling algorithm*, Cryptology ePrint Archive, 2014, pozycja 2014/466

- [BF74] M.B. Brown, A.B. Forsythe, *Robust tests for the equality of variances*, Journal of the American Statistical Association, 69 (346), 1974, s. 364–367
- [B06] R.G. Brown, *DIEHARDER: A GNU Public Licensed Random Number Tester*, Duke University Physics Department, 2006, <http://www.phy.duke.edu/~rgb>
- [CP06] C. De Cannière, B. Preneel, *Trivium - A Stream Cipher Construction Inspired by Block Cipher Design Principles*, eSTREAM submitted papers, 2006
- [CN98] J.S. Coron, D. Naccache, *An Accurate Evaluation of Maurer's Universal Test*, Proceedings of SAC '98 (Lecture Notes in Computer Science), Springer-Verlag, 1998, s. 57-71
- [Co98] J.S. Coron, *On the Security of Random Sources*, Gemplus Corporate Product R&D Division, Technical Report IT02-1998
- [Cr80] N. Cressie, *The Asymptotic Distribution of the Scan Statistic under Uniformity*, Annals of Probability, 8, 4, 1980, s. 828-840
- [Cu02] H. Cui, *Average projection type weighted Cramér- von Mises statistics for testing some distributions*, Science in China Series A: Mathematics, 45, 2002, s. 562-577
- [DR99] J. Daemen, V. Rijmen, *AES Proposal: Rijndael*, 1999
- [DEKS10] A. Doğanaksoy, B. Ege, O. Koçak, F. Sulak, *Cryptographic Randomness Testing of Block Ciphers and Hash Functions*, Cryptology ePrint Archive, 2010, pozycja 2010/564
- [D&15] A. Doğanaksoy, F. Sulak, M. Uğuz, O. Seker, Z. Akcengiz, *New Statistical Randomness Tests Based on Length of Runs*, Hindawi Publishing Corporation, Mathematical Problems in Engineering, Volume 2015
- [DGL00] J.H. Drew, A.G. Glen, L.M. Leemis, *Computing the Cumulative Distribution Function of the Kolmogorov-Smirnov Statistic*, Computational Statistics and Data Analysis, 34, 2000, s. 1-15
- [FIPS1] *Federal Information Processing Standard Publication 140-1*, FIPS PUB 140-1, 1994
- [FIPS2] *Federal Information Processing Standard Publication 140-2*, FIPS PUB 140-2, 2001
- [F02] M. Fischler, *Distribution of Minimum Distance Among N Random Points in d Dimensions*, DOI: 10.2172/794005, FERMILAB-TM-2170, 2002
- [GGS96] W. Gander, G.H. Golub, R. Streb, *Least-Squares Fitting of Circles and Ellipses*, Bulletin of the Belgian Mathematical Society, Supplement to Volume 3, 1996, s. 63-84
- [GMW97] P.E. Gill, W. Murray, M.H. Wright, *Practical Optimization*, Academic Press Limited, wyd. 11, 1997
- [GK57] B. Gnienenko B., A.N. Kołmogorow, *Rozkłady graniczne sum zmiennych losowych niezależnych*, PWN, Warszawa, 1957
- [G67] S.W. Golomb, *Shift Register Sequences*, San Francisco, Holden-Day, 1967

- [Go53] I.J. Good, *The serial test for sampling number and other tests for randomness*, Proceedings Cambridge Philosophical Society, 49, 1953, s. 276-284
- [GPST17] C. Georgescu, A. Petrescu-Nita, E. Simion, A. Toma, *NIST randomness tests (in)dependence*, Cryptology ePrint Archive, 2017, pozycja 2017/336
- [GO81] L.J. Guibas, A.M. Odlyzko, *Periods in Strings*, Journal of Combinatorial Theory, 20, 1981, s. 19-42
- [H07] K. Hamano, T. Kaneko, *Correction of Overlapping Template Matching Test Included in NIST Randomness Test Suite*, IEICE Transactions of Electronics, Communications and Computer Sciences 2007, E90-A(9), s. 1788-1792
- [H79] L. Holst, *Asymptotic Normality of Sum-Functions of Spacings*, The Annals of Probability, 7, 6, s. 1066-1072
- [J09] T. Jarocki, *Wyznaczenie dokładnych momentów pierwszego i drugiego rzędu w testach wolnych cel Marsagli*, praca dyplomowa, WAT, 2009
- [K06] G.K. Kanji, *100 Statistical Tests*, 3rd Edition, SAGE Publications, 2006
- [KS73] M.G. Kendall, A. Stuart, *The Advanced Theory of Statistics 2*, Ch. Griffin & Company Limited, London, 1973
- [KS01] W. Killmann, W. Schindler, *AIS-31 – Functionality classes and evaluation methodology for true (physical) random number generators*, Version 3.1, 2001
- [KPS94] P. Kirschenhofer, H. Prodinger, W. Szpankowski, *Digital Search Trees Again Revisited: The Internal Path Length Perspective*, SIAM Journal on Computation, 23, 1994, s. 598-616
- [K81] D.E. Knuth, *The Art of Computer Programming*, Addison-Wesley, 1981
- [K97] D.E. Knuth, *Sztuka programowania, t.2. Algorytmy seminumeryczne*, WNT, 2002
tłumaczenie trzeciego wydania: *The Art of Computer Programming*, 1997
- [KP13] P. Kowalik, *Metoda wskaźników pojemności informacyjnej Hellwiga jako zadanie nieliniowego programowania binarnego*, Rola informatyki w naukach ekonomicznych i społecznych, tom 2, Kielce, 2013, s. 215-224
- [KP17] P. Kowalik, *Selection of independent variables in econometric models as a binary programming problem and its application to spreadsheet-based calculations*, *Probability in Action*, vol. 3, Politechnika Lubelska, 2017,
- [KFC02] K.A. Kowalska, D. Fogliano, J.G. Coello, *On the revision of NIST 800-22 Test Suites*, Crypta Labs, London, United Kingdom, Cryptology ePrint Archive, 2022, pozycja 2022/540
- [KL01] L. Kowalski, *Statystyka*, BEL Studio, 2001, wydanie I
- [E97] P. L'Ecuyer, *Tests Based on Sum-Functions of Spacings for Uniform Random Numbers*, Journal of Statistical Computation and Simulation, 59, 1997, s. 251-269

- [ECC97] P. L'Ecuyer, A. Compagner, J.F. Cordeau, *Entropy Tests for Random Number Generators*, GERAD report G-96-41, wersja marzec 1997,
<http://www.iro.umontreal.ca/~lecuyer/myftp/papers/entrop.ps>
- [ESW98] P. L'Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, 1998
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.3118&rep=rep1&type=pdf>
- [ECS00] P. L'Ecuyer, J.F. Cordeau, R. Simard, *Close-Point Spatial Tests and Their Application to Random Number Generators*, *Operations Research*, 48, 2, 2000, s. 308-317
- [ES02] P. L'Ecuyer, R. Simard, *TestU01: A C Library for Empirical Testing of Random Number Generators*, 2002, <http://www.iro.umontreal.ca/~lecuyer/>
- [ESW02] P. L'Ecuyer, R. Simard, S. Wegenkittl, *Sparse Serial Tests of Uniformity for Random Number Generators*, *SIAM Journal on Scientific Computing*, 24, 2, 2002, s. 652-668
- [ES07a] P. L'Ecuyer, R. Simard, *TestU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators*, Universite de Montreal, 2002, wersja z 23 kwietnia 2007, <http://www.iro.umontreal.ca/~lecuyer/>
- [ES07b] P. L'Ecuyer, R. Simard, *ProbDist: A Software Library of Probability Distributions and Goodness-of-Fit Statistics in ANSI C*, Universite de Montreal, 2002, wersja z 23 kwietnia 2007, <http://www.iro.umontreal.ca/~lecuyer/>
- [Ma04] K. Mańk, *Dokładne dystrybuanty statystyk w testach momentów 1 i 2 rzędu*, Biuletyn WAT, Kryptologia IV, Warszawa, 2004, s. 47-65
- [Ma05] K. Mańk, *Statistical testing of random numbers and random bit sequence generators*, IEEE International Conference on Technologies for Homeland Security and Safety, TEHOSS 2005, s. 269-272
- [Ma08] K. Mańk, *Wyznaczenie dokładnych wartości parametrów dla testu Maurera*, Biuletyn WAT, LVII, 4, Warszawa, 2008, s. 117-149
- [Ma08r] K. Mańk, *Testy losowości wykorzystujące statystykę chi-kwadrat*, referat w ramach XXXVII Ogólnopolskiej Konferencji Zastosowań Matematyki, Zakopane-Kościelisko, 2008
- [Ma09] K. Mańk, *Analiza nowych rozwiązań w dziedzinach kryptologii i statystyki pod kątem uruchomienia programu badawczego własnego pod tytułem „Zbieżność rozkładów i skorelowanie statystyk testów losowości wykorzystywanych w badaniach mocy algorytmów kryptograficznych”*, Praca końcowa, Warszawa, 2009
- [Ma10] K. Mańk, *Ciekawe przypadki w testowaniu losowości*, Metody biometryczne i kryptograficzne w zintegrowanych systemach bezpieczeństwa - konferencja, Warszawa, 2010

- [Ma13] K. Mańk, *Test częstości dla nakładających się wektorów*, Cyberprzestępczość i ochrona informacji – Bezpieczeństwo w internecie tom II, Wydawnictwo Wyższej Szkoły Menedżerskiej w Warszawie, ISBN 978-83-7520-131-4, 2013
- [Ma14] K. Mańk, *Test autokorelacyjny dla ciągu binarnego*, Studia Bezpieczeństwa Narodowego, IV, 6, Kryptologia i Cyberbezpieczeństwo, Warszawa, 2014, s. 199-226
- [Ma17] K. Mańk, *Wyznaczenie rozkładu średniej wartości odstępów w teście pochodnej ciągu liczb*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2017
- [Ma18] K. Mańk, *Implementacja testów opartych o dyskretną transformatę Fouriera*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2018
- [Ma19a] K. Mańk, *Implementacja i analiza testu największej liczby elementów w przedziale*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2019
- [Ma19b] K. Mańk, *Dostosowanie testu największej liczby elementów w przedziale do wykorzystania w pakiecie testów losowości*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2019
- [Ma22] K. Mańk, *Speeding Up Minimum Distance Randomness Tests*, Journal of Telecommunications and Information Technology, 2022, vol. 2, s. 103-109
<https://doi.org/10.26636/jtit.2022.159222>
- [Ma24] K. Mańk, *Efficient Evaluation of Frequency Test for Overlapping Vectors Statistic*, CECC 2024, Book of Abstracts, M. Kutylowski, J. Pomykała, Warszawa, 2024, s. 70-73
- [Ma25] K. Mańk, *Upgrading Frequency Test for Overlapping Vectors and Fill Tree Tests*, International Journal of Electronics and Telecommunication, 2/2025
- [M84] G. Marsaglia, *A Current View of Random Number Generators*, The Proceedings of the Computer Science and Statistics: 16th Symposium on the Interface, Atlanta, Elsevier Press, 1984
- [M95] G. Marsaglia, *DIEHARD Battery of Tests of Randomness*, 1995,
<http://stat.fsu.edu/pub/diehard/>
- [MM04] G. Marsaglia, J.C.W. Marsaglia, *Evaluating the Anderson-Darling Distribution*, Journal of Statistical Software, 9, 2, 2004, s. 1-5
- [MTW03] G. Marsaglia, W.W. Tsang, J. Wang, *Evaluating Kolmogorov's Distribution*, Journal of Statistical Software, 8, 18, 2004, s. 1-4
- [MZ93] G. Marsaglia, A. Zaman, *Monkey Tests for Random Number Generators*, Computers and Mathematics with Applications, 9, 1993
- [MT02] G. Marsaglia, W.W. Tsang, *Some Difficult-to-pass Tests of Randomness*, Journal of Statistical Software 7, 3, 2002, s. 1-8
- [MK94] M. Matsumoto, Y. Kurita, *Twisted GFSR Generators II*, ACM Transactions of Modeling and Computer Simulation, 4, 3, 1994, s. 254-266

- [MU92] U.M. Maurer, *A Universal Statistical Test for Random Bit Generators*, Journal of Cryptology, 5, 1992, Springer-Verlag, s. 89-105
- [MOV96] A. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996
- [MNG19] P.R. Mishra, B. Nandan, N. Gaba, *On NIST's Compression Estimate Test*, Cryptology ePrint Archive, 2019, pozycja 2019/967
- [Mo40] A.M. Mood, *The Distribution Theory of Runs*, Annals of Mathematical Statistics, 11, 1940, s. 367-392
- [N91] H. Niederreiter, *The Linear Complexity Profile and the Jump Complexity of Keystream Sequences*, Lecture Notes in Computer Science, 473, 1991, s. 174-188
- [PRS12] F. Pareschi, R. Rovatti, G. Setti, *On statistical tests for randomness included in the NIST SP800-22 test suite and based on the binomial distribution*, IEEE Transactions on Information Forensics and Security 7(2), 2012, s. 491-505
- [PTVF92] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C*, Cambridge University Press, 1992
- [RC88] T.R.C. Read, N.A.C. Cressie, *Goodness-of-Fit Statistics for Discrete Multivariate Data*, Springer Series in Statistics, Springer-Verlag, Nowy Jork, 1988
- [RR03] E. Rivals, S. Rahmann, *Combinatorics of Periods in Strings*, Journal of Combinatorial Theory, 104, 2003, s. 95-113
- [NIST01] A. Rukhin i inni, *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*, NIST Special Publication 800-22, 2001
- [R80] B.Ya. Ryabko, *Information Compression by a Book Stack*, Problems of Information Transmission, v.16, N 4, 1980, s. 16-21
- [RSS03] B.Ya. Ryabko, V.S. Stognienko, Yu.I. Shokin, *A new test for randomness and its application to some cryptographic problems*, Journal of Statistical Planning and Inference, 2003, s. 365-376
- [RM03] B.Ya. Ryabko, V.A. Monarev, *Using Information Theory Approach to Randomness Testing*, Cryptology ePrint Archive, 2003, pozycja 2003/127
- [RR03] B.Ya. Ryabko, J. Rissanen, *Fast Adaptive Arithmetic Code for Large Alphabet Sources with Asymmetrical Distributions*, IEEE Communications Letters, v. 7, no. 1, 2003, s.33-35
- [SH11] I. Semaev, M.M. Hassanzadeh, *New Statistical Box-Test and It's Power*, Cryptology ePrint Archive, 2011, pozycja 2011/297
- [S11] I. Semaev, *Local limit theorem for large deviations and statistical box-tests*, Cryptology ePrint Archive, 2011, pozycja 2011/298

- [SHB97] L.N. Shchur, J.R. Heringa, H.W. Blöte, *Simulation of a Directed Random-Walk Model: The Effect of Pseudo-Random Number Correlations*, Physica A, 1997, s. 579-592
- [S08] R. Simard, *SSJ User's Guide – Package GoF*, 2008,
<http://www.iro.umontreal.ca/~simardr/ssj/doc/pdf/guidegof.pdf>
- [SUH04] K. Song-Ju, K. Umeno, A. Hasegawa, *Corrections of the NIST Statistical Test Suite for Randomness*, Cryptology ePrint Archive, Report 2004/018
- [SB00] J. Soto, L. Bassham, *Randomness Testing of the Advanced Encryption Standard Finalist Candidates*, NIST IR 6483, National Institute of Standards and Technology, 2000
<http://csrc.nist.gov/CryptoToolkit/aes/round2/conf3/papers/30-jsoto.pdf>
- [St70] M. A. Stephens, *Use of the Kolmogorov-Smirnov, Cramer-von Mises and Related Statistics without Extensive Tables*, Journal of the Royal Statistical Society, Series B, 32, 1, 1970, s. 115-122
- [Sz18] P. Szabuńko, *Wykonanie implementacji i analiza teoretyczna testu złożoności sekwencji wykorzystującego algorytm kompresji Lempel-Ziv*, opracowanie w ramach realizacji projektu p.k. ROTOR, 2018
- [US90] G. Ugrin-Sparac, *On a Distribution Encountered in the Renewal Process Based on Uniform Distribution*, Glasnik Matematicki, vol. 25 (45), 1990, s. 221-233
- [US91] G. Ugrin-Sparac, *Stochastic Investigation of Pseudo-Random Number Generators*, Computing 46, 1991, s. 53-65
- [Wa14] Y. Wang, *On the Design of LIL Tests for (Pseudo) Random Generators and Some Experimental Results*, Cryptology ePrint Archive, 2014, pozycja 2014/031
- [WW78] M.S. Waterman, D.E. Whiteman, *Estimation of probability densities by empirical density functions*, International Journal of Mathematical Education in Science and Technology, 9.2, 1978, s. 127-137
- [W96] S. Wegenkittl, *Empirical Testing of Pseudorandom Number Generators*, praca dyplomowa, Naturwissenschaftlichen Fakultät der Universität Salzburg, 1996
- [W98] S. Wegenkittl, *Generalized ϕ -Divergence and Frequency Analysis in Markov Chains*, rozprawa doktorska, Naturwissenschaftlichen Fakultät der Universität Salzburg, 1998
- [WZ97] R. Wieczorkowski, R. Zieliński R., *Komputerowe generatory liczb losowych*, WNT, Warszawa, 1997
- [WH36] J. Wishart, H.O. Hirshfeld, *A Theorem Concerning the Distribution of Joins Between Line Segments*, Journal of the London Mathematical Society, 11, 1936
- [WJ98] J. Wiśniewska, *Comparing quality of pseudo- and true- random numbers obtained from different sources*, Symulacja w Badaniach i Rozwoju, Vol. 9, No. 1-2/2018, s. 73-82
- [Y82] A.C. Yao, *Theory and Application of Trapdoor Functions*, 23rd IEEE Symposium on Foundations of Computer Science, Chicago, 1982, s. 80-91

- [Za74] J. Zacharski, *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*, Warszawa, 1974
- [Z16] S. Zhu, Y. Ma, J. Lin, J. Zhuang, J. Jing, *More Powerful and Reliable Second-level Statistical Randomness Tests for NIST SP 800-22*, Cryptology ePrint Archive, 2016, pozycja 2016/863
- [ZL78] J. Ziv, A. Lempel, *Compression of Individual Sequences via Variable-Rate Coding*, IEEE Transactions on Information Theory, 24, 5, 1978, s. 530-536

Załącznik A

Zestawy procedur dla ciągów długości 1 kiB

1. Zestaw początkowy

1.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32$,
- autokorelacji dla ciągu bloków
 - $d = 2, t = 1..2047$,
 - $d = 3, t = 1..1364$,
 - $d = 4, t = 1..1023$,
 - $d = 5, t = 1..818$,
 - $d = 6, t = 1..682$,
 - $d = 7, t = 1..584$,
 - $d = 8, t = 1..511$,
 - $d = 9, t = 1..454$,
 - $d = 10, t = 1..409$,
 - $d = 11, t = 1..371$,
 - $d = 12, t = 1..340$,
 - $d = 13, t = 1..314$,
 - $d = 14, t = 1..292$,
 - $d = 15, t = 1..272$,
 - $d = 16, t = 1..251$,
 - $d = 17, t = 1..240$,
 - $d = 18, t = 1..227$,
 - $d = 19, t = 1..215$,
 - $d = 20, t = 1..204$,
 - $d = 21, t = 1..194$,
 - $d = 22, t = 1..185$,
 - $d = 23, t = 1..177$,
 - $d = 24, t = 1..170$,
 - $d = 25, t = 1..163$,
 - $d = 26, t = 1..157$,
 - $d = 27, t = 1..151$,
 - $d = 28, t = 1..145$,
 - $d = 29, t = 1..140$,
 - $d = 30, t = 1..136$,
 - $d = 31, t = 1..131$,
 - $d = 32, t = 1..127$,
- Corona – $d = 3..5$,
- Cramera-von Misesa – $d = 16..32$,
- gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka ośmiościenna – zgodności dla liczby zwycięstw,
 - kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka sześcienna – zgodności dla liczby zwycięstw,
- kolekcjonera – $d = 2..4$,
- Kołmogorowa – $d = 16..32$,
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,

- Kuipera – $d = 16..32$,
- odstępów dni urodzin
 - $d = 11..12, n = 16, 32$,
 - $d = 13..15, n = 32, 64$,
 - $d = 16..18, n = 64, 128$,
 - $d = 19..21, n = 128, 256$,
 - $d = 22..24, n = 256, 512$,
- pochodnej ciągu bloków – $d = 2..32$
 - pierwsza pochodna,
 - druga pochodna,
- porządku książek

○ $d = 4, k = 1..3$, ○ $d = 5, k = 1..4$, ○ $d = 6, k = 1..5$, ○ $d = 7, k = 1..6$, ○ $d = 8, k = 1..7$, ○ $d = 9, k = 1..8$, ○ $d = 10, k = 1..9$,	○ $d = 11, k = 1..10$, ○ $d = 12, k = 1..11$, ○ $d = 13, k = 1..12$, ○ $d = 14..17, k = 1..13$, ○ $d = 18, k = 1..11$, ○ $d = 19..26, k = 1..10$,
--	--
- stosu książek

○ $d = 4, k = 1..3$, ○ $d = 5, k = 1..4$, ○ $d = 6, k = 1..5$, ○ $d = 7, k = 1..6$, ○ $d = 8, k = 1..7$, ○ $d = 9, k = 1..8$, ○ $d = 10, k = 1..9$,	○ $d = 11, k = 1..10$, ○ $d = 12, k = 1..11$, ○ $d = 13, k = 1..12$, ○ $d = 14..17, k = 1..13$, ○ $d = 18, k = 1..11$, ○ $d = 19..26, k = 1..10$,
--	--
- serii malejących z macierzą kowariancji – $d = 2..32$,
- serii rosnących z macierzą kowariancji – $d = 2..32$,
- serii malejących z pomijaniem w przód – $d = 2..32$,
- serii malejących z pomijaniem w tył – $d = 2..32$,
- serii rosnących z pomijaniem w przód – $d = 2..32$,
- serii rosnących z pomijaniem w tył – $d = 2..32$,
- sumy kwadratów odstępów $m = 1, 2, 3, 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32$ elementowych bloków $d = 17..32$ bitowych
 - bez zapętlania,
 - z zapętlaniem,

- sumy logarytmów odstępów $m = 1, 2, 3, 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32$ elementowych bloków $d = 17..32$ bitowych
 - bez zapętlenia,
 - z zapętleniem,
- ściskania – $d = 10..16$,
- uniwersalny Maurera – $d = 1..5$,
- wag Hamminga w ciągu bloków
 - $d = 2, t = 2..8, k = 3$,
 - $d = 3, t = 2..8, k = 2, 4$,
 - $d = 4, t = 2..8, k = 3, 5$,
 - $d = 5, t = 2..8, k = 2, 4, 6$,
 - $d = 6, t = 2..8, k = 3, 5, 7$,
 - $d = 7, t = 2..8, k = 2, 4, 6, 8$,
 - $d = 8, t = 2..8, k = 3, 5, 7, 9$,
 - $d = 9, t = 2..8, k = 2, 4, 6, 8, 10$,
 - $d = 10$
 - $t = 2..7, k = 3, 5, 7, 9, 11$,
 - $t = 8, k = 3, 5, 7, 9$,
 - $d = 11$
 - $t = 2..7, k = 2, 4, 6, 8, 10, 12$,
 - $t = 8, k = 2, 4, 6, 8, 10$,
 - $d = 12$
 - $t = 2..7, k = 3, 5, 7, 9, 11, 13$,
 - $t = 8, k = 3, 5, 7, 9, 11$,
 - $d = 13$
 - $t = 2..7, k = 2, 4, 6, 8, 10, 12, 14$,
 - $t = 8, k = 2, 4, 6, 8, 10$,
 - $d = 14$
 - $t = 2..6, k = 3, 5, 7, 9, 11, 13, 15$,
 - $t = 7, k = 3, 5, 7, 9, 11, 13$,
 - $t = 8, k = 3, 5, 7, 9$,
 - $d = 15$
 - $t = 2..6, k = 2, 4, 6, 8, 10, 12, 14, 16$,
 - $t = 7, k = 2, 4, 6, 8, 10, 12, 14$,
 - $t = 8, k = 2, 4, 6, 8, 10$,

- $d = 16$
 - $t = 2..6, k = 3, 5, 7, 9, 11, 13, 15, 17,$
 - $t = 7, k = 3, 5, 7, 9, 11, 13,$
 - $t = 8, k = 3, 5, 7, 9,$
- $d = 20$
 - $t = 2..6, k = 3, 5, 7, 9, 11, 13, 15, 17, 19, 21,$
 - $t = 7, k = 3, 5, 7, 9, 11, 13,$
 - $t = 8, k = 3, 5, 7, 9,$
- $d = 32$
 - $t = 2..5, k = 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33,$
 - $t = 6, k = 3, 5, 7, 9, 11, 13, 15, 17, 19, 21,$
 - $t = 7, k = 3, 5, 7, 9, 11, 13,$
 - $t = 8, k = 3, 5, 7, 9,$
- wypełniania drzewa binarnego o wysokości $h = 3..11$ na podstawie ciągu bloków $d = 4..32$ bitowych
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

1.2. Procedury dla ciągów wektorów nienakładających się

- dwupoziomowy Bickela-Breimana – $t = 2..16, d = 16..32,$
- iloczynów – $t = 2..16, d = 2..32,$
- kolizji
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- kolizji indeksów maksymalnych elementów – $t = 2..16, d = 2..32,$
- liczby komórek z co najmniej $b = 1..31$ punktami dla wektorów:
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- liczby komórek z dokładnie $b = 0..31$ punktami dla wektorów:
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$

- $t = 5, d = 1..6,$
- $t = 6, d = 1..5,$
- $t = 7, d = 1..4,$
- logarytm prawdopodobieństwa dla wektorów:
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- m najbliższych par – metryka euklidesowa i nieskończoność:
 - $m = 2$
 - $t = 8, d = 29..32,$
 - $t = 9, d = 26..32,$
 - $t = 10, d = 23..32,$
 - $t = 11, d = 21..32,$
 - $t = 12, d = 19..32,$
 - $t = 13, d = 18..32,$
 - $t = 14, d = 17..32,$
 - $t = 15, d = 16..32,$
 - $t = 16, d = 16..32,$
 - $m = 3$
 - $t = 5, d = 26..32,$
 - $t = 6, d = 22..32,$
 - $t = 7, d = 19..32,$
 - $t = 8, d = 17..28,$
 - $t = 9, d = 16..25,$
 - $t = 10, d = 16..22,$
 - $t = 11, d = 16..20,$
 - $t = 12, d = 16..18,$
 - $m = 4$
 - $t = 3, d = 28..32,$
 - $t = 4, d = 21..32,$
 - $t = 5, d = 17..25,$
 - $t = 6, d = 16..21,$
 - $t = 7, d = 16..18,$
 - $t = 8, d = 16,$
 - $m = 5$
 - $t = 2, d = 29..32,$
 - $t = 3, d = 19..27,$
 - $t = 4, d = 16..20,$
 - $t = 5, d = 16,$
 - $m = 6$
 - $t = 2, d = 21..28,$
 - $t = 3, d = 16..18,$
 - $m = 7, t = 2, d = 17..20,$
 - $m = 8, t = 2, d = 16,$
- maksimum z $t - t = 2..16, d = 2..32,$
- momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 1..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$

- momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- permutacji – $t = 2..10, d = 2..32,$
- podziałów – $t = 2..16, d = 2..32,$
- pokerowy – $t = 2..16, d = 2..32,$
- potęgowo-dywergencyjny – $\delta = 1..7$
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- zgodności bitowej – $t = 2..16, d = 16..32,$
- zgodności Pearsona
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1.$

1.3. Procedury dla ciągów wektorów nakładających się

- częstości dla nakładających się wektorów
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$
 - $t = 4, d = 2,$
- logarytm prawdopodobieństwa
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$
 - $t = 4, d = 2,$
- nakładających się permutacji – $t = 3..6, d = 2..32,$
- potęgowo-dywergencyjny – $\delta = 1..7$
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$

- $t = 4, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego
 - $t = 1, d = 4,$
 - $t = 2, d = 3,$
 - $t = 2, d = 5,$
 - $t = 4, d = 2.$

1.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..4096,$
- całkowitej liczby serii,
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- częstości serii 0 i 1 łącznie,
- częstości serii 0 i 1 osobno zliczanych,
- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- korelacji wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- liczby skoków w profilu liniowym,
- liczby skoków w profilu liniowym dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320$ bitowe,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146, 158, 166, 178, 194, 202, 206, 214, 218, 226, 254, 256, 262, 274, 278, 298, 302, 314, 326, 334, 346, 358, 362, 382, 386$:
 - liczby powrotów do 0,
 - liczby przejść przez 0 ze zmianą strony,
 - łącznego „czasu” odchylenia w prawo,
 - maksymalnego odchylenia ścieżki,
 - momentu pierwszego osiągnięcia wartości $y = 1..100,$
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64, 128, 256$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64, 128, 256$ bitowym,
- nakładających się t bitowych wzorców 0, 1 w bloku M bitowym
 - $t = 4, M = 16, 32, 64, 128, 256,$
 - $t = 5, M = 32, 64, 128, 256,$
 - $t = 6, M = 64, 128, 256,$
 - $t = 7, M = 128, 256,$
 - $t = 8, M = 256,$

- nienakładających się d bitowych wzorców w bloku M bitowym (wzorce w zapisie heksadecymalnym, ograniczone do początkowych 21):
 - $d = 2, M = 16, 32, 64, 128, 256$, wzorce: 1, 2,
 - $d = 3, M = 16, 32, 64, 128, 256$, wzorce: 1, 3, 4, 6,
 - $d = 4, M = 16, 32, 64, 128, 256$, wzorce: 1, 3, 7, 8, C, E,
 - $d = 5, M = 32, 64, 128, 256$, wzorce: 1, 3, 5, 7, B, F, 10, 14, 18, 1A, 1C, 1E,
 - $d = 6, M = 64, 128, 256$, wzorce: 1, 3, 5, 7, B, D, F, 13, 17, 1F, 20, 28, 2C, 30, 32, 34, 38, 3A, 3C, 3E,
 - $d = 7, M = 128, 256$, wzorce: 1, 3, 7, 9, B, D, F, 13, 15, 17, 1B, 1D, 1F, 23, 27, 2B, 2F, 37, 3F, 40,
 - $d = 8, M = 256$, wzorce: 1, 3, 5, 7, 9, B, D, F, 13, 15, 17, 19, 1B, 1D, 1F, 23, 25, 27, 2B, 2D, 2F,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- niezależności wag Hamminga – wariant 2

○ $M = 4, d = 1..2,$	○ $M = 64, d = 1..32,$
○ $M = 5, d = 1..3,$	○ $M = 67, d = 1..34,$
○ $M = 7, d = 1..4,$	○ $M = 71, d = 1..36,$
○ $M = 8, d = 1..4,$	○ $M = 73, d = 1..37,$
○ $M = 11, d = 1..6,$	○ $M = 79, d = 1..40,$
○ $M = 13, d = 1..7,$	○ $M = 83, d = 1..42,$
○ $M = 16, d = 1..8,$	○ $M = 89, d = 1..45,$
○ $M = 17, d = 1..9,$	○ $M = 97, d = 1..49,$
○ $M = 19, d = 1..10,$	○ $M = 101, d = 1..51,$
○ $M = 23, d = 1..12,$	○ $M = 103, d = 1..52,$
○ $M = 29, d = 1..15,$	○ $M = 107, d = 1..54,$
○ $M = 31, d = 1..16,$	○ $M = 109, d = 1..55,$
○ $M = 32, d = 1..16,$	○ $M = 113, d = 1..57,$
○ $M = 37, d = 1..19,$	○ $M = 127, d = 1..64,$
○ $M = 41, d = 1..21,$	○ $M = 128, d = 1..64,$
○ $M = 43, d = 1..22,$	○ $M = 131, d = 1..66,$
○ $M = 47, d = 1..24,$	○ $M = 137, d = 1..69,$
○ $M = 53, d = 1..27,$	○ $M = 139, d = 1..70,$
○ $M = 59, d = 1..30,$	○ $M = 149, d = 1..75,$
○ $M = 61, d = 1..31,$	○ $M = 151, d = 1..76,$

- $M = 157, d = 1..79,$
- $M = 163, d = 1..82,$
- rozkładu skoków w profilu liniowym,
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320$ bitowe,
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 6..8,$
- wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- wypełniania drzewa binarnego o wysokości $h = 3..10$ na podstawie ciągu bitów
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

2. Procedury testowe przydatne do badania pojedynczych ciągów

2.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32,$
- autokorelacji dla ciągu bloków
 - $d = 2, t = 1..2047,$
 - $d = 3, t = 1..1364,$
 - $d = 4, t = 1..1023,$
 - $d = 5, t = 1..818,$
 - $d = 6, t = 1..682,$
 - $d = 7, t = 1..584,$
 - $d = 8, t = 1..511,$
 - $d = 9, t = 1..454,$
 - $d = 10, t = 1..409,$
 - $d = 11, t = 1..371,$
 - $d = 12, t = 1..340,$
 - $d = 13, t = 1..314,$
 - $d = 14, t = 1..292,$
 - $d = 15, t = 1..272,$
 - $d = 16, t = 1..251,$
 - $d = 17, t = 1..240,$
 - $d = 18, t = 1..227,$
 - $d = 19, t = 1..215,$
 - $d = 20, t = 1..204,$
 - $d = 21, t = 1..194,$
 - $d = 22, t = 1..185,$
 - $d = 23, t = 1..177,$
 - $d = 24, t = 1..170,$
 - $d = 25, t = 1..163,$
 - $d = 26, t = 1..157,$
 - $d = 27, t = 1..151,$
 - $d = 28, t = 1..145,$
 - $d = 29, t = 1..140,$
 - $d = 30, t = 1..136,$
 - $d = 31, t = 1..131,$
 - $d = 32, t = 1..127,$
- Corona – $d = 3..5,$
- Cramera-von Misesa – $d = 16..32,$
- gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,

- kostka ośmiościenna – zgodności dla liczby zwycięstw,
- kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
- kostka sześcienna – zgodności dla liczby zwycięstw,
- kolekcjonera – $d = 2..4$,
- Kołmogorowa – $d = 16..32$,
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- Kuipera – $d = 16..32$,
- pochodnej ciągu bloków
 - pierwsza pochodna – $d = 2..5, 8, 11$,
 - druga pochodna – $d = 2..5, 8, 10, 11, 16$,
- porządku książek

○ $d = 4, k = 2..3$, ○ $d = 5, k = 2..4$, ○ $d = 6, k = 2..5$, ○ $d = 7, k = 2..6$, ○ $d = 8, k = 2..7$, ○ $d = 9, k = 2..8$, ○ $d = 10, k = 2..9$,	○ $d = 11, k = 2..10$, ○ $d = 12, k = 2..11$, ○ $d = 13, k = 2..12$, ○ $d = 14..17, k = 2..13$, ○ $d = 18, k = 2..11$, ○ $d = 19..26, k = 2..10$,
--	--
- stosu książek

○ $d = 4, 9, 12, 14, 20, 24, 26$, $k = 2..3$, ○ $d = 5..7, 10, 11, 13, 22, 23$, $k = 2..4$,	○ $d = 8, k = 2..7$, ○ $d = 15, 17, k = 2..5$, ○ $d = 16, k = 2..9$, ○ $d = 18, 19, 21, 25, k = 2$,
--	--
- serii malejących z pomijaniem w przód – $d = 4..32$,
- serii malejących z pomijaniem w tył – $d = 4..32$,
- serii rosnących z pomijaniem w przód – $d = 4..32$,
- serii rosnących z pomijaniem w tył – $d = 4..32$,
- sumy kwadratów odstępów $m = 1, 2, 3, 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32$ elementowych bloków $d = 17..32$ bitowych, bez zapętlania,
- sumy kwadratów odstępów m elementowych bloków d bitowych, z zapętlaniem

○ $m = 1, d = 17..27$, ○ $m = 2..4, d = 17..32$, ○ $m = 5, 7, 8, d = 17..31$, ○ $m = 11, d = 17..27$,	○ $m = 13, 16, 17, d = 17..24$, ○ $m = 19, d = 17..21$, ○ $m = 23, d = 17$,
--	--

- sumy logarytmów odstępów m elementowych bloków d bitowych, bez zapętlenia
 - $m = 1, d = 22..32,$
 - $m = 2, d = 17..32,$
 - $m = 3, d = 17..24,$
 - $m = 4, d = 17,$
- sumy logarytmów odstępów m elementowych bloków d bitowych, z zapętleniem
 - $m = 1, d = 24..27,$
 - $m = 2..4, d = 17..32,$
 - $m = 3, d = 17..24,$
 - $m = 4, d = 17,$
- ściskania – $d = 10..16,$
- uniwersalny Maurera – $d = 1..5,$
- wag Hamminga w ciągu bloków
 - $d = 2, t = 2..4, k = 3,$
 - $d = 3$
 - $t = 2..3, k = 2, 4,$
 - $t = 4..8, k = 2,$
 - $d = 4, 6$
 - $t = 2, k = 3, 5,$
 - $t = 3..5, k = 3,$
 - $d = 5, 7, 9, 11$
 - $t = 2, k = 2, 4, 6,$
 - $t = 3..4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 8$
 - $t = 2..3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 10, 12$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
 - $t = 4, k = 3,$
 - $d = 13, 15$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4..8, k = 2,$
 - $d = 14, 16$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
 - $d = 20$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $d = 32, t = 2, k = 3, 5, 7, 9,$
- wypełniania drzewa binarnego – zgodności dla pozycji kolizji
 - $h = 3, d = 4..32,$
 - $h = 4, d = 4..24,$
 - $h = 5, d = 4..20,$
 - $h = 6, d = 4..14,$
 - $h = 7, d = 4..10,$
 - $h = 8, d = 4..5,$
 - $h = 9, d = 4,$
- wypełniania drzewa binarnego – zgodności dla wypełnień
 - $h = 3, d = 4..17,$
 - $h = 4, d = 4..26,$
 - $h = 5, d = 4..17,$
 - $h = 6, d = 4..15,$
 - $h = 7, d = 4..9,$
 - $h = 8, d = 4..6,$
 - $h = 9, d = 4.$

2.2. Procedury dla ciągów wektorów nienakładających się

- iloczynów
 - $t = 2..4, d = 8..32,$
 - $t = 5, 13, 16, d = 10..32,$
 - $t = 6..12, 14..15, d = 9..32,$
- kolizji indeksów maksymalnych elementów
 - $t = 3, d = 9, 11, 12, 15..19,$
 - $t = 4, d = 7..28,$
 - $t = 5, d = 7..30,$
 - $t = 6, d = 7..22,$
 - $t = 7, d = 7..16,$
 - $t = 8, d = 7..12,$
 - $t = 9, d = 7..10,$
 - $t = 10, d = 7..8,$
- liczby komórek z dokładnie $b = 1$ punktem dla wektorów:
 - $t = 1, d = 10,$
 - $t = 2, d = 5,$
 - $t = 5, d = 2,$
 - $t = 10, d = 1,$
- logarytm prawdopodobieństwa dla wektorów:
 - $t = 1, d = 4, 8, 12,$
 - $t = 2, d = 2, 4, 6,$
 - $t = 3, d = 4,$
 - $t = 4, d = 1, 2, 3,$
 - $t = 6, d = 2,$
 - $t = 8, d = 1,$
 - $t = 12, d = 1,$
- m najbliższych par – metryka (E)uklidesowa i (N)ieskończoność:
 - $m = 2$
 - $t = 8, d = 29..31,$
 - $t = 9, d = 26..31,$
 $d = 32 - E,$
 - $t = 10, d = 23..31,$
 $d = 32 - E,$
 - $t = 11, d = 21..31,$
 $d = 32 - E,$
 - $t = 12, d = 19..31,$
 $d = 32 - E,$
 - $t = 13, d = 18..30,$
 $d = 31 - N,$
 - $t = 14, d = 17..20,$
 $d = 21..31 - N,$
 - $t = 15, d = 16..17,$
 $d = 18..31 - N,$
 - $t = 16, d = 16..31 - N,$
 - $m = 3$
 - $t = 5, d = 26..31,$
 $d = 32 - E,$
 - $t = 6, d = 22..31,$
 $d = 32 - E,$
 - $t = 7, d = 19..31,$
 $d = 32 - E,$
 - $t = 8, d = 17..28,$
 - $t = 9, d = 16..25,$
 - $t = 10, d = 16..22,$
 - $t = 11, d = 16..20,$
 - $t = 12, d = 16..18,$
 - $t = 13, d = 16..17,$
 - $t = 14, d = 16,$
 - $m = 4$
 - $t = 3, d = 28..32,$
 - $t = 4, d = 21..32,$
 - $t = 5, d = 17..25,$

- $t = 6, d = 16..21,$
 - $t = 7, d = 16..18,$
 - $t = 8, d = 16,$
 - $m = 5,$
 - $t = 2, d = 29..32,$
 - $t = 3, d = 19..27,$
 - $t = 4, d = 16..20,$
 - $t = 5, d = 16,$
 - $m = 6,$
 - $t = 2, d = 21..28,$
 - $t = 3, d = 16..18,$
 - $m = 7, t = 2, d = 17..20,$
 - $m = 8, t = 2, d = 16 - E,$
- maksimum z t
 - $t = 2..4, d = 12..32,$
 - $t = 5..6, 8 d = 13..32,$
 - $t = 7, 9..16, d = 14..32,$
 - momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 1..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
 - momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
 - permutacji
 - $t = 2, d = 3..6,$
 - $t = 3..4, d = 2..32,$
 - $t = 5, d = 2..30,$
 - $t = 6, d = 2..26,$
 - $t = 7, d = 2..22,$
 - $t = 8, d = 2..19,$
 - $t = 9, d = 2..17,$
 - $t = 10, d = 2..16,$
 - podziałów
 - $t = 3, d = 2..3,$
 - $t = 4..5, d = 2..4,$
 - $t = 6..8, d = 2..5,$
 - $t = 9..14, d = 3..6,$
 - $t = 15..16, d = 3..7,$
 - pokerowy
 - $t = 3, d = 2..3,$
 - $t = 4..5, d = 2..4,$
 - $t = 6..8, d = 2..5,$
 - $t = 9..14, d = 2..6,$
 - $t = 15..16, d = 2..7,$

- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 1, d = 2..10,$
 - $t = 2, d = 1..5,$
 - $t = 3, d = 1..4,$
 - $t = 4..5, d = 2,$
 - $t = 6..10, d = 1,$
- potęgowo-dywergencyjny – $\delta = 2$
 - $t = 1, d = 3..7,$
 - $t = 2, d = 2..3,$
 - $t = 3, d = 1..2,$
 - $t = 4..7, d = 1,$
- potęgowo-dywergencyjny – $\delta = 3$
 - $t = 1, d = 4..7,$
 - $t = 2, d = 2..3,$
 - $t = 3, d = 2,$
 - $t = 4..7, d = 1,$
- potęgowo-dywergencyjny – $\delta = 4$
 - $t = 1, d = 4..5,$
 - $t = 2, d = 2,$
 - $t = 4..5, d = 1,$
- potęgowo-dywergencyjny – $\delta = 5..7$
 - $t = 1, d = 4,$
 - $t = 2, d = 2,$
 - $t = 4, d = 1,$
- zgodności Pearsona
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1.$

2.3. Procedury dla ciągów wektorów nakładających się

- częstości dla nakładających się wektorów
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$
 - $t = 4, d = 2,$
- nakładających się permutacji
 - $t = 3..5, d = 2..32,$
 - $t = 6, d = 4..32,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$
 - $t = 4, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego – $t = 2, d = 5.$

2.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..4096$,
- całkowitej liczby serii,
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- częstości serii 0 i 1 łącznie,
- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- liczby skoków w profilu liniowym,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146, 158, 166, 178$:
 - liczby powrotów do 0,
 - liczby przejść przez 0 ze zmianą strony,
 - łącznego „czasu” odchylenia w prawo,
 - maksymalnego odchylenia ścieżki,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64, 128$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64, 128$ bitowym,
- nakładających się t bitowych wzorców 0, 1 w bloku M bitowym
 - $t = 4, M = 16, 32, 64, 128$,
 - $t = 5, M = 32, 64, 128$,
 - $t = 6, M = 64, 128$,
- nienakładających się d bitowych wzorców w bloku M bitowym (wzorce w zapisie heksadecymalnym, ograniczone do początkowych 21):
 - $d = 2, M = 256$, wzorce: 1, 2,
 - $d = 3, M = 32, 64, 128, 256$, wzorce: 1, 3, 4, 6,
 - $d = 4, M = 256$, wzorce: 1, 3, 7, 8, C, E,
 - $d = 5, M = 256$, wzorce: 1, 3, 5, 7, B, F, 10, 14, 18, 1A, 1C, 1E,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89$ bitowego,
- niezależności wag Hamminga – wariant 2

○ $M = 4, d = 2$, ○ $M = 5, d = 2..3$, ○ $M = 7, d = 3..4$, ○ $M = 8, d = 3..4$, ○ $M = 11, d = 5..6$, ○ $M = 13, d = 6..7$, ○ $M = 16, d = 7..8$,	○ $M = 17, d = 7..8$, ○ $M = 19, d = 9$, ○ $M = 23, d = 10..11$, ○ $M = 29, d = 13..14$, ○ $M = 31, d = 14..15$, ○ $M = 32, d = 15..16$, ○ $M = 37, d = 18$,
---	---

- $M = 41, d = 20,$
- $M = 43, d = 21,$
- $M = 47, d = 22..23,$
- $M = 53, d = 26,$
- $M = 61, d = 29..30,$
- $M = 64, d = 31..32,$
- $M = 67, d = 33,$
- $M = 71, d = 34..35,$
- $M = 73, d = 36,$
- $M = 83, d = 41,$
- $M = 89, d = 44,$
- $M = 103, d = 51,$
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty $m = 64$ bitowe,
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 6..8,$
- wag Hamminga dla bloku $M = 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- wypełniania drzewa binarnego na podstawie ciągu bitów
 - zgodności dla pozycji kolizji – $h = 3..6,$
 - zgodności dla wypełnień – $h = 3..7.$

3. Procedury testowe przydatne do badania generatorów przy małej liczbie powtórzeń

3.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32,$
- autokorelacji dla ciągu bloków
 - $d = 2, t = 1..2047,$
 - $d = 3, t = 1..1364,$
 - $d = 4, t = 1..1023,$
 - $d = 5, t = 1..818,$
 - $d = 6, t = 1..682,$
 - $d = 7, t = 1..584,$
 - $d = 8, t = 1..511,$
 - $d = 9, t = 1..454,$
 - $d = 10, t = 1..409,$
 - $d = 11, t = 1..371,$
 - $d = 12, t = 1..340,$
 - $d = 13, t = 1..314,$
 - $d = 14, t = 1..292,$
 - $d = 15, t = 1..272,$
 - $d = 16, t = 1..251,$
 - $d = 17, t = 1..240,$
 - $d = 18, t = 1..227,$
 - $d = 19, t = 1..215,$
 - $d = 20, t = 1..204,$
 - $d = 21, t = 1..194,$
 - $d = 22, t = 1..185,$
 - $d = 23, t = 1..177,$
 - $d = 24, t = 1..170,$
 - $d = 25, t = 1..163,$
 - $d = 26, t = 1..157,$
 - $d = 27, t = 1..151,$
 - $d = 28, t = 1..145,$
 - $d = 29, t = 1..140,$
 - $d = 30, t = 1..136,$
 - $d = 31, t = 1..131,$
 - $d = 32, t = 1..127,$

- Corona – $d = 3..5$,
- Cramera-von Misesa – $d = 16..32$,
- gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka ośmiościenna – zgodności dla liczby zwycięstw,
 - kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
- kolekcjonera – $d = 2..3$,
- Kołmogorowa – $d = 16..32$,
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- porządku książek

○ $d = 4, k = 2..3$, ○ $d = 5, k = 2..4$, ○ $d = 6, k = 2..5$, ○ $d = 7, k = 2..6$, ○ $d = 8, k = 2..7$, ○ $d = 9, k = 2..8$, ○ $d = 10, k = 2..9$,	○ $d = 11, k = 2..10$, ○ $d = 12, k = 2..11$, ○ $d = 13, k = 2..12$, ○ $d = 14..17, k = 2..13$, ○ $d = 18, k = 2..11$, ○ $d = 19..26, k = 2..10$,
--	--
- stosu książek

○ $d = 13, 23, 24, 9,$ $k = 2,$ ○ $d = 4, 5, 11, 12, 14, 15, 20,$ $k = 2..3,$	○ $d = 6, 7, 10, 17,$ $k = 2..4,$ ○ $d = 8, k = 2..7,$ ○ $d = 16, k = 2..9,$
--	--
- serii malejących z pomijaniem w przód – $d = 5..12, 15..17$,
- serii malejących z pomijaniem w tył – $d = 5..12, 15..17$,
- serii rosnących z pomijaniem w przód – $d = 5..12, 15..17$,
- serii rosnących z pomijaniem w tył – $d = 5..12, 15..17$,
- sumy kwadratów odstępów m elementowych bloków d bitowych, bez zapętlenia,

○ $m = 17, d = 27,$ ○ $m = 19, d = 20, 27,$	○ $m = 23, 29, 31, 32, d = 17..32,$
--	--
- ściskania – $d = 10..14$,
- uniwersalny Maurera – $d = 1..5$,
- wag Hamminga w ciągu bloków

○ $d = 2, t = 2..4, k = 3,$ ○ $d = 3$	▪ $t = 2, k = 2, 4,$ ▪ $t = 3..8, k = 2,$
--	--

- $d = 4, 6$
 - $t = 2, k = 3, 5,$
 - $t = 3..5, k = 3,$
- $d = 5$
 - $t = 2..3, k = 2, 4,$
 - $t = 4..8, k = 2,$
- $d = 7, 9$
 - $t = 2, k = 2, 4, 6,$
 - $t = 3..4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 8$
 - $t = 2, k = 3, 5,$
 - $t = 3..4, k = 3,$
- $d = 10, 12, t = 2..3, k = 3, 5,$
- $d = 11$
- $d = 13$
 - $t = 2, k = 2, 4, 6,$
 - $t = 3, k = 2, 4,$
 - $t = 4..7, k = 2,$
- $d = 14, 16$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
- $d = 15$
 - $t = 2..3, k = 2, 4, 6,$
 - $t = 4..7, k = 2,$
- $d = 20, t = 2, k = 3, 5, 7,$
- $d = 32, t = 2, k = 3, 7, 9,$
- wypełniania drzewa binarnego – zgodności dla pozycji kolizji
 - $h = 3, d = 4..17,$
 - $h = 4, d = 4..14,$
 - $h = 5, d = 4..20,$
 - $h = 6, d = 4..14,$
 - $h = 7, d = 4..10,$
 - $h = 8, d = 4..5,$
 - $h = 9, d = 4,$
- wypełniania drzewa binarnego – zgodności dla wypełnień
 - $h = 3, 6, d = 4..15,$
 - $h = 4..5, d = 4..17,$
 - $h = 7, d = 4..9,$
 - $h = 8, d = 4..6,$
 - $h = 9, d = 4.$

3.2. Procedury dla ciągów wektorów nienakładających się

- kolizji indeksów maksymalnych elementów
 - $t = 4, d = 9,$
 - $t = 5, d = 8, 9, 11,$
 - $t = 6, d = 10,$
 - $t = 7, d = 8,$
- logarytm prawdopodobieństwa dla wektorów:
 - $t = 1, d = 4, 8,$
 - $t = 2, d = 2, 4,$
 - $t = 4, d = 1, 2,$
 - $t = 8, d = 1,$
- m najbliższych par – metryka (E)uklidesowa i (N)ieskończoność:
 - $m = 3$
 - $t = 5, d = 27..31,$
 - $d = 26, 32 - E,$
 - $t = 6, d = 22..30,$
 - $d = 31 - N,$
 - $t = 7, d = 19..31,$
 - $d = 32 - E,$

- $t = 8, d = 17..28,$
 - $t = 9, d = 16..25,$
 - $t = 10, d = 16..22,$
 - $t = 11, d = 16..20,$
 - $t = 12, d = 16..18,$
 - $t = 13, d = 16..17,$
 - $t = 14, d = 16 - N,$
 - $m = 4$
 - $t = 3, d = 28..32,$
 - $t = 4, d = 21..32,$
 - $t = 5, d = 17..25,$
 - $t = 6, d = 16..21,$
 - $t = 7, d = 16..18,$
 - $t = 8, d = 16,$
 - $m = 5,$
 - $t = 2, d = 29..32,$
 - $t = 3, d = 19..27,$
 - $t = 4, d = 17..20,$
 - $d = 16 - E,$
 - $t = 5, d = 16,$
 - $m = 6,$
 - $t = 2, d = 21..28,$
 - $t = 3, d = 17..18,$
 - $d = 16 - E,$
 - $m = 7,$
 - $t = 2, d = 17..18 - E,$
 - $t = 2, d = 19..20,$
 - $m = 8, t = 2, d = 16 - E,$
- maksimum z t
 - $t = 2, d = 13..32,$
 - $t = 3, 6, d = 14..32,$
 - $t = 4, 5, 7..9, 11..14, 16, d = 15..32,$
 - $t = 10, 15, d = 16..32,$
 - momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 1..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
 - momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 2..29,$
 - $t = 2, d = 1..14,$
 - $t = 3, d = 1..9,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..5,$
 - $t = 6..7, d = 1..4,$
 - $t = 8..9, d = 1..3,$
 - $t = 10..14, d = 1..2,$
 - $t = 15..29, d = 1,$
 - permutacji
 - $t = 3, d = 2..20,$
 - $t = 4, d = 2..17,$
 - $t = 5, d = 2..25,$
 - $t = 6, d = 2..22,$
 - $t = 7, d = 2..18,$
 - $t = 8, d = 2..14,$
 - $t = 9, d = 2..12,$
 - $t = 10, d = 2..11,$

- podziałów
 - $t = 3, d = 2,$
 - $t = 4..5, d = 2..3,$
 - $t = 6, d = 2..4,$
 - $t = 7, 9, d = 3..4,$
 - $t = 8, 10..13, d = 3..5,$
 - $t = 14, d = 3..6,$
 - $t = 15, d = 4..6,$
 - $t = 16, d = 4..5,$
- pokerowy
 - $t = 3, d = 2,$
 - $t = 4, d = 2..3,$
 - $t = 5..7, d = 2..4,$
 - $t = 8..12, d = 2..5,$
 - $t = 13..16, d = 2..6,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 1, d = 2..7,$
 - $t = 2, d = 1..3,$
 - $t = 3, d = 1..2,$
 - $t = 4..7, d = 1,$
- potęgowo-dywergencyjny – $\delta = 2$
 - $t = 1, d = 4..6,$
 - $t = 2, d = 2..3,$
 - $t = 3, d = 2,$
 - $t = 4..6, d = 1,$
- potęgowo-dywergencyjny – $\delta = 3..5$
 - $t = 1, d = 4,$
 - $t = 2, d = 2,$
 - $t = 4, d = 1,$
- zgodności Pearsona
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1.$

3.3. Procedury dla ciągów wektorów nakładających się

- częstości dla nakładających się wektorów
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$
 - $t = 4, d = 2,$
- nakładających się permutacji
 - $t = 3..5, d = 2..32,$
 - $t = 6, d = 5..15,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$

- $t = 4, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego – $t = 2, d = 5.$

3.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..4096,$
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127$ bitowego,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146:$
 - liczby powrotów do 0,
 - łącznego „czasu” odchylenia w prawo,
 - maksymalnego odchylenia ścieżki,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64$ bitowym,
- nakładających się t bitowych wzorców 0, 1 w bloku M bitowym
 - $t = 4, M = 32, 64, 128,$
 - $t = 5, M = 32, 64,$
 - $t = 6, M = 64,$
- nienakładających się d bitowych wzorców w bloku M bitowym – $d = 3, M = 64, 128, 256,$ wzorce: 1, 3, 4, 6,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67$ bitowego,
- niezależności wag Hamminga – wariant 2
 - $M = 5, d = 3,$
 - $M = 7, 8, d = 4,$
 - $M = 19, d = 9,$
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 6..8,$
- wag Hamminga dla bloku $M = 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- wypełniania drzewa binarnego na podstawie ciągu bitów
 - zgodności dla pozycji kolizji – $h = 3..6,$
 - zgodności dla wypełnień – $h = 3..7.$

4. Procedury testowe przydatne do badania generatorów przy dużej liczbie powtórzeń

4.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32$,
- autokorelacji dla ciągu bloków
 - $d = 2, t = 1..2047$,
 - $d = 3, t = 1..1364$,
 - $d = 4, t = 1..1023$,
 - $d = 5, t = 1..818$,
 - $d = 6, t = 1..682$,
 - $d = 7, t = 1..584$,
 - $d = 8, t = 1..511$,
 - $d = 9, t = 1..454$,
 - $d = 10, t = 1..409$,
 - $d = 11, t = 1..371$,
 - $d = 12, t = 1..340$,
 - $d = 13, t = 1..314$,
 - $d = 14, t = 1..292$,
 - $d = 15, t = 1..272$,
 - $d = 16, t = 1..251$,
 - $d = 17, t = 1..240$,
 - $d = 18, t = 1..227$,
 - $d = 19, t = 1..215$,
 - $d = 20, t = 1..204$,
 - $d = 21, t = 1..194$,
 - $d = 22, t = 1..185$,
 - $d = 23, t = 1..177$,
 - $d = 24, t = 1..170$,
 - $d = 25, t = 1..163$,
 - $d = 26, t = 1..157$,
 - $d = 27, t = 1..151$,
 - $d = 28, t = 1..145$,
 - $d = 29, t = 1..140$,
 - $d = 30, t = 1..136$,
 - $d = 31, t = 1..131$,
 - $d = 32, t = 1..127$,
- Cramera-von Misesa – $d = 16..32$,
- gra w kości – kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
- kolekcjonera – $d = 2$,
- Kołmogorowa – $d = 16..32$,
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- porządku książek
 - $d = 4..5, k = 2..3$,
 - $d = 6..7, k = 2..4$,
 - $d = 8, k = 2..5$,
 - $d = 9..10, k = 3..4$,
 - $d = 11..17, k = 3$,
- stosu książek
 - $d = 4, k = 2..3$,
 - $d = 5..6, k = 2$,
 - $d = 7, 11, 16, k = 3$,
 - $d = 8, k = 2..7$,

- serii malejących z pomijaniem w przód – $d = 5..6$,
- serii malejących z pomijaniem w tył – $d = 5..6$,
- serii rosnących z pomijaniem w przód – $d = 5..6$,
- serii rosnących z pomijaniem w tył – $d = 5..6$,
- ściskania – $d = 11$,
- uniwersalny Maurera – $d = 2$,
- wag Hamminga w ciągu bloków
 - $d = 2, t = 2..3, k = 3$,
 - $d = 3, t = 2, k = 2, 4$,
 - $d = 5, 7, 9, 11, t = 2..3, k = 2$,
 - $d = 6, 8, 10, 12, 14, 16, 20, t = 2, k = 3$,
 - $d = 13, 15, t = 2, k = 2$,
 - $d = 32, t = 2, k = 9$,
- wypełniania drzewa binarnego – zgodności dla pozycji kolizji
 - $h = 3, d = 4..9$,
 - $h = 4, d = 4..7$,
 - $h = 5, d = 4..5$,
- wypełniania drzewa binarnego – zgodności dla wypełnień
 - $h = 3, d = 4..7$,
 - $h = 4, d = 4..8$,
 - $h = 5..6, d = 4$.

4.2. Procedury dla ciągów wektorów nienakładających się

- logarytm prawdopodobieństwa dla wektorów:
 - $t = 1, d = 4, 8$,
 - $t = 2, d = 2, 4$,
 - $t = 4, d = 1, 2$,
 - $t = 8, d = 1$,
- m najbliższych par – metryka (E)uklidesowa i (N)ieskończoność:
 - $m = 4$
 - $t = 5, d = 18 - N$,
 - $t = 6, d = 17$,
 - $d = 19 - N$,
 - $m = 5$
 - $t = 2, d = 29, 32 - N$,
 - $t = 3, d = 19..20$,
 - $d = 21..23 - N$,
 - $t = 4, d = 19..20$,
 - $m = 6$
 - $t = 5, d = 16 - E$,
 - $t = 2, d = 21..28$,
 - $t = 3, d = 16..18 - E$,
 - $m = 7$
 - $t = 2, d = 17..20 - E$,
 - $m = 8, t = 2, d = 16 - E$,

- maksimum z t
 - $t = 2..3, d = 15..32,$
 - $t = 4..10, d = 16..32,$
 - $t = 11..13, d = 17..32,$
 - $t = 14, 16 d = 18..32,$
 - $t = 15, d = 17..32,$
- momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 3..30,$
 - $t = 2, d = 2..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 2..29,$
 - $t = 2, d = 1..14,$
 - $t = 3, d = 1..9,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..5,$
 - $t = 6..7, d = 1..4,$
 - $t = 8..9, d = 1..3,$
 - $t = 10..14, d = 1..2,$
 - $t = 15..29, d = 1,$
- permutacji
 - $t = 3, d = 2..12,$
 - $t = 4, d = 2..10,$
 - $t = 5, d = 2..6,$
 - $t = 6..8, d = 2..4,$
 - $t = 9..10, d = 2,$
- podziałów
 - $t = 5..6, d = 3,$
 - $t = 7, 9, 11, 13, d = 4,$
- pokerowy
 - $t = 4..6, d = 2..3,$
 - $t = 7..10, 12, 13, d = 2..4,$
 - $t = 11, 14, d = 2..5,$
 - $t = 15..16, d = 2..3,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 1, d = 2..6,$
 - $t = 2, d = 1..3,$
 - $t = 3, d = 1..2,$
 - $t = 4..6, d = 1,$
- potęgowo-dywergencyjny – $\delta = 2$
 - $t = 1, d = 4,$
 - $t = 2, d = 2,$
 - $t = 4, d = 1,$
- zgodności Pearsona
 - $t = 1, d = 2..7,$
 - $t = 2, d = 1..3,$
 - $t = 3, d = 1..2,$
 - $t = 4..7, d = 1.$

4.3. Procedury dla ciągów wektorów nakładających się

- nakładających się permutacji
 - $t = 3..4, d = 2..14, 16,$
 - $t = 5, d = 3..12,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$
 - $t = 4, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego – $t = 2, d = 5.$

4.4. Procedury dla ciągów bitów

- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41$ bitowego,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64$ bitowym,
- nienakładających się d bitowych wzorców w bloku M bitowym – $d = 3, M = 256$, wzorce: 1, 3, 4, 6,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19$ bitowego,
- wag Hamminga dla bloku $M = 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- wypełniania drzewa binarnego na podstawie ciągu bitów
 - zgodności dla pozycji kolizji – $h = 3..4,$
 - zgodności dla wypełnień – $h = 3..4.$

5. Rekomendowany pakiet procedur dla ciągów długości 1 kiB

5.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32,$
- autokorelacji dla ciągu bloków

○ $d = 2, t = 1..2047,$	○ $d = 10, t = 1..409,$
○ $d = 3, t = 1..1364,$	○ $d = 11, t = 1..371,$
○ $d = 4, t = 1..1023,$	○ $d = 12, t = 1..340,$
○ $d = 5, t = 1..818,$	○ $d = 13, t = 1..314,$
○ $d = 6, t = 1..682,$	○ $d = 14, t = 1..292,$
○ $d = 7, t = 1..584,$	○ $d = 15, t = 1..272,$
○ $d = 8, t = 1..511,$	○ $d = 16, t = 1..251,$
○ $d = 9, t = 1..454,$	○ $d = 17, t = 1..240,$

- $d = 18, t = 1..227,$
- $d = 19, t = 1..215,$
- $d = 20, t = 1..204,$
- $d = 21, t = 1..194,$
- $d = 22, t = 1..185,$
- $d = 23, t = 1..177,$
- $d = 24, t = 1..170,$
- $d = 25, t = 1..163,$
- $d = 26, t = 1..157,$
- $d = 27, t = 1..151,$
- $d = 28, t = 1..145,$
- $d = 29, t = 1..140,$
- $d = 30, t = 1..136,$
- $d = 31, t = 1..131,$
- $d = 32, t = 1..127,$
- Cramera-von Misesa – $d = 16..32,$
- gra w kości – kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
- kolekcjonera – $d = 2,$
- Kołmogorowa – $d = 16..32,$
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- porządku książek
 - $d = 4..5, k = 2..3,$
 - $d = 6..7, k = 2..4,$
 - $d = 8, k = 2..5,$
 - $d = 9..10, k = 3..4,$
 - $d = 11..15, 17, k = 3,$
- stosu książek
 - $d = 4, k = 2..3,$
 - $d = 5..6, k = 2,$
 - $d = 7, 11, 16, k = 3,$
 - $d = 8, k = 2..6,$
- serii malejących z pomijaniem w przód – $d = 5..6,$
- serii malejących z pomijaniem w tył – $d = 5..6,$
- serii rosnących z pomijaniem w przód – $d = 5..6,$
- serii rosnących z pomijaniem w tył – $d = 5..6,$
- ściskania – $d = 11,$
- uniwersalny Maurera – $d = 2,$
- wag Hamminga w ciągu bloków
 - $d = 2, t = 2..3, k = 3,$
 - $d = 3, t = 2, k = 2,$
 - $d = 5, 7, 9, 11, t = 2..3, k = 2,$
 - $d = 6, 8, 10, 12, 14, 16, 20, t = 2, k = 3,$
 - $d = 13, 15, t = 2, k = 2,$
 - $d = 32, t = 2, k = 9,$

- wypełniania drzewa binarnego – zgodności dla pozycji kolizji
 - $h = 3, d = 4..9,$
 - $h = 4, d = 4..7,$
 - $h = 5, d = 4..5,$
- wypełniania drzewa binarnego – zgodności dla wypełnień
 - $h = 3, d = 4..7,$
 - $h = 4, d = 4..8,$
 - $h = 5..6, d = 4.$

5.2. Procedury dla ciągów wektorów nienakładających się

- logarytm prawdopodobieństwa dla wektorów – $t = 1, d = 4, 8,$
- m najbliższych par – metryka (E)uklidesowa i (N)ieskończoność:
 - $m = 4$
 - $t = 5, d = 18 - N,$
 - $t = 6, d = 17,$
 - $d = 19 - N,$
 - $m = 5,$
 - $t = 2, d = 29, 32 - N,$
 - $t = 3, d = 19..20,$
 - $d = 21..23 - N,$
 - $t = 4, d = 19..20,$
 - $d = 16 - E,$
 - $d = 18 - N,$
 - $t = 5, d = 16 - E,$
 - $m = 6,$
 - $t = 2, d = 21..28,$
 - $t = 3, d = 16..18 - E,$
 - $m = 7,$
 - $t = 2, d = 17..20 - E,$
 - $m = 8, t = 2, d = 16 - E,$
- maksimum z t
 - $t = 2..3, d = 15..32,$
 - $t = 4..10, d = 16..32,$
 - $t = 11..13, d = 17..32,$
 - $t = 14, 16 d = 18..32,$
 - $t = 15, d = 17..32,$
- momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów – $t = 1, d = 3..30,$
- momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów – $t = 1, d = 2..29,$
- permutacji
 - $t = 3, d = 2..12,$
 - $t = 4, d = 2..10,$
 - $t = 5, d = 2..6,$
 - $t = 6..8, d = 2..4,$
 - $t = 9..10, d = 2,$

- podziałów
 - $t = 5..6, d = 3,$
 - $t = 7, 9, 11, 13, d = 4,$
 - pokerowy
 - $t = 4..6, d = 2..3,$
 - $t = 7..10, 12, 13, d = 2..4,$
 - $t = 11, 14, d = 2..5,$
 - $t = 15..16, d = 2..3,$
 - potęgowo-dywergencyjny – $\delta = 2, t = 1, d = 4,$
 - zgodności Pearsona – $t = 1, d = 2..7,$
- przypadek $d = 4$ można pominąć obliczając:

$$U^Z = 0,503384U^P + 0,497344U^L - 0,000234,$$

gdzie U^Z , U^P i U^L są odpowiadającymi p_{val} testów zgodności Pearsona, potęgowo-dywergencyjnego i logarytmu prawdopodobieństwa dla $d = 4$.

5.3. Procedury dla ciągów wektorów nakładających się

- nakładających się permutacji
 - $t = 3..4, d = 2..14, 16,$
 - $t = 5, d = 3..12,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 2, d = 4,$
 - $t = 3, d = 3,$
 - $t = 4, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego – $t = 2, d = 5.$

5.4. Procedury dla ciągów bitów

- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41$ bitowego,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64$ bitowym,
- nienakładających się d bitowych wzorców w bloku M bitowym – $d = 3, M = 256$, wzorce: 1, 3, 4, 6,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19$ bitowego,
- wag Hamminga dla bloku $M = 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163$ bitowego,
- wypełniania drzewa binarnego na podstawie ciągu bitów
 - zgodności dla pozycji kolizji – $h = 3..4,$
 - zgodności dla wypełnień – $h = 3..4.$

Załącznik B

Zestawy procedur dla ciągów długości 1 Mb

1. Zestaw początkowy

1.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32$,
- autokorelacji dla ciągu bloków
 - $d = 2, t = 1..8192, 241809..249999$,
 - $d = 3, t = 1..8192, 158475..166666$,
 - $d = 4, t = 1..8192, 116809..124999$,
 - $d = 5, t = 1..8192, 91809..99999$,
 - $d = 6, t = 1..8192, 75142..83332$,
 - $d = 7, t = 1..8192, 63237..71428$,
 - $d = 8, t = 1..8192, 54309..62499$,
 - $d = 9, t = 1..8192, 47364..55555$,
 - $d = 10, t = 1..8192, 41809..49999$,
 - $d = 11, t = 1..8192, 37263..45454$,
 - $d = 12, t = 1..8192, 33475..41666$,
 - $d = 13, t = 1..8192, 30270..38461$,
 - $d = 14, t = 1..8192, 27523..35713$,
 - $d = 15, t = 1..8192, 25142..33332$,
 - $d = 16, t = 1..8192, 23059..31249$,
 - $d = 17, t = 1..8192, 21220..29411$,
 - $d = 18, t = 1..8192, 19586..27777$,
 - $d = 19, t = 1..8192, 18124..26315$,
 - $d = 20, t = 1..8192, 16809..24999$,
 - $d = 21, t = 1..8192, 15618..23809$,
 - $d = 22, t = 1..8192, 14536..22726$,
 - $d = 23, t = 1..8192, 13548..21738$,
 - $d = 24, t = 1..8192, 12642..20832$,
 - $d = 25, t = 1..8192, 11809..19999$,
 - $d = 26, t = 1..8192, 11039..19230$,
 - $d = 27, t = 1..8192, 10327..18518$,
 - $d = 28, t = 1..8192, 9666..17856$,
 - $d = 29, t = 1..8192, 9050..17240$,
 - $d = 30, t = 1..8192, 8475..16666$,
 - $d = 31, t = 1..16128$,
 - $d = 32, t = 1..15624$,
- Corona – $d = 3..11$,
- Cramera-von Misesa – $d = 16..32$,
- częstości dla podprzedziału
 - $d = 16..32, [1/128; 1/4 + 1/128), M = 32, 256, 2048$,
 - $d = 16..32, [1/64; 1/2 + 1/128), M = 32, 256, 2048$,
 - $d = 16..32, [1/2 - 1/128; 3/4), M = 32, 256, 2048$,
- gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka ośmiościenna – zgodności dla liczby zwycięstw,
 - kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka sześcienna – zgodności dla liczby zwycięstw,

- kolekcjonera – $d = 2..9$,
- Kołmogorowa – $d = 16..32$,
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- Kuipera – $d = 16..32$,
- największej liczby elementów w przedziale – podciągi 1 kiB, $D = 1/2, 0,4, 0,3, 1/4, 0,2, 1/8, 1/16, 1/32, 1/64, 1/128, d = 4..32$,
- odstępów dni urodzin
 - $d = 11..12, n = 16, 32$,
 - $d = 13..15, n = 32, 64$,
 - $d = 16..18, n = 64, 128$,
 - $d = 19..21, n = 128, 256$,
 - $d = 22..24, n = 256, 512$,
- pochodnej ciągu bloków – $d = 2..32$
 - pierwsza pochodna,
 - druga pochodna,
- porządku książek

○ $d = 4, k = 1..3$, ○ $d = 5, k = 1..4$, ○ $d = 6, k = 1..5$, ○ $d = 7, k = 1..6$, ○ $d = 8, k = 1..7$, ○ $d = 9, k = 1..8$, ○ $d = 10, k = 1..9$,	○ $d = 11, k = 1..10$, ○ $d = 12, k = 1..11$, ○ $d = 13, k = 1..12$, ○ $d = 14..17, k = 1..13$, ○ $d = 18, k = 1..11$, ○ $d = 19..26, k = 1..10$,
--	--
- serii malejących z macierzą kowariancji – $d = 2..32$,
- serii rosnących z macierzą kowariancji – $d = 2..32$,
- serii malejących z pomijaniem w przód – $d = 2..32$,
- serii malejących z pomijaniem w tył – $d = 2..32$,
- serii rosnących z pomijaniem w przód – $d = 2..32$,
- serii rosnących z pomijaniem w tył – $d = 2..32$,
- stosu książek

○ $d = 4, k = 1..3$, ○ $d = 5, k = 1..4$, ○ $d = 6, k = 1..5$,	○ $d = 7, k = 1..6$, ○ $d = 8, k = 1..7$, ○ $d = 9, k = 1..8$,
---	---

- $d = 10, k = 1..9,$
- $d = 11, k = 1..10,$
- $d = 12, k = 1..11,$
- $d = 13, k = 1..12,$
- $d = 14..17, k = 1..13,$
- $d = 18, k = 1..11,$
- $d = 19..26, k = 1..10,$
- sum do G
 - $d = 16..32, G = 16, 32, 64, 128, 256, 512,$
 - $d = 16..23, G = 1024,$
- sumy kwadratów odstępów $m = 1, 2, 3, 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024, 8192, 15625$ elementowych bloków $d = 17..32$ bitowych,
 - bez zapętlania,
 - z zapętlaniem,
- sumy logarytmów odstępów $m = 1, 2, 3, 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024, 8192, 15625$ elementowych bloków $d = 17..32$ bitowych,
 - bez zapętlania,
 - z zapętlaniem,
- ściskania – $d = 10..20,$
- uniwersalny Maurera – $d = 1..11,$

- wag Hamminga w ciągu bloków
 - $d = 2, t = 2..8, k = 3,$
 - $d = 3, t = 2..8, k = 2, 4,$
 - $d = 4, t = 2..8, k = 3, 5,$
 - $d = 5, t = 2..8, k = 2, 4, 6,$
 - $d = 6, t = 2..8, k = 3, 5, 7,$
 - $d = 7, t = 2..8, k = 2, 4, 6, 8,$
 - $d = 8, t = 2..8, k = 3, 5, 7, 9,$
 - $d = 9, t = 2..8, k = 2, 4, 6, 8, 10,$
 - $d = 10$
 - $t = 2..7, k = 3, 5, 7, 9, 11,$
 - $t = 8, k = 3, 5, 7, 9,$
 - $d = 11$
 - $t = 2..7, k = 2, 4, 6, 8, 10, 12,$
 - $t = 8, k = 2, 4, 6, 8, 10,$
 - $d = 12$
 - $t = 2..7, k = 3, 5, 7, 9, 11, 13,$
 - $t = 8, k = 3, 5, 7, 9,$
 - $d = 13$
 - $t = 2..7, k = 2, 4, 6, 8, 10, 12, 14,$
 - $t = 8, k = 2, 4, 6, 8, 10,$
 - $d = 14$
 - $t = 2..6, k = 3, 5, 7, 9, 11, 13, 15,$
 - $t = 7, k = 3, 5, 7, 9, 11, 13,$
 - $t = 8, k = 3, 5, 7, 9,$
 - $d = 15$
 - $t = 2..6, k = 2, 4, 6, 8, 10, 12, 14, 16,$
 - $t = 7, k = 2, 4, 6, 8, 10, 12, 14,$
 - $t = 8, k = 2, 4, 6, 8, 10,$
 - $d = 16$
 - $t = 2..6, k = 3, 5, 7, 9, 11, 13, 15, 17,$
 - $t = 7, k = 3, 5, 7, 9, 11, 13,$
 - $t = 8, k = 3, 5, 7, 9,$
 - $d = 20$
 - $t = 2..6, k = 3, 5, 7, 9, 11, 13, 15, 17, 19, 21,$
 - $t = 7, k = 3, 5, 7, 9, 11, 13,$

- $t = 8, k = 3, 5, 7, 9,$
- $d = 32$
 - $t = 2..5, k = 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33,$
 - $t = 6, k = 3, 5, 7, 9, 11, 13, 15, 17, 19, 21,$
 - $t = 7, k = 3, 5, 7, 9, 11, 13,$
 - $t = 8, k = 3, 5, 7, 9,$
- wypełniania drzewa binarnego o wysokości $h = 3..11$ na podstawie ciągu bloków $d = 4..32$ bitowych
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

1.2. Procedury dla ciągów wektorów nienakładających się

- dwupoziomowy Bickela-Breimana – $t = 2..16, d = 16..32,$
- iloczynów – $t = 2..16, d = 2..32,$
- kolizji
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- kolizji indeksów maksymalnych elementów – $t = 2..16, d = 2..32,$
- liczby komórek z co najmniej $b = 1..31$ punktami dla wektorów:
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- liczby komórek z dokładnie $b = 0..31$ punktami dla wektorów:
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- logarytm prawdopodobieństwa
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$

- $t = 7, d = 1..4,$
- $t = 8..10, d = 1..3,$
- $t = 11..15, d = 1..2,$
- $t = 16..30, d = 1,$

- m najbliższych par – metryka euklidesowa i nieskończoność

Tabela B.1. Wartości m dla poszczególnych par t i d w teście m najbliższych par

$d \setminus t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16	88	72	62	55	51	47	44	41	39	37	36	34	33	32	31
17	85	70	60	54	49	45	42	40	38	36	35	33	32	31	30
18	83	68	58	52	48	44	41	39	37	35	34	32	31	30	29
19	81	66	57	51	46	43	40	38	36	34	33	31	30	29	28
20	79	64	55	50	45	42	39	37	35	33	32	31	29	28	27
21	77	62	54	48	44	41	38	36	34	32	31	30	29	28	27
22	75	61	53	47	43	40	37	35	33	32	30	29	28	27	26
23	73	60	52	46	42	39	36	34	32	31	30	28	27	26	26
24	72	58	51	45	41	38	36	34	32	30	29	28	27	26	25
25	70	57	50	44	40	37	35	33	31	30	28	27	26	25	25
26	69	56	49	43	40	37	34	32	31	29	28	27	26	25	24
27	68	55	48	43	39	36	34	32	30	29	27	26	25	24	24
28	66	54	47	42	38	35	33	31	29	28	27	26	25	24	23
29	65	53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24	23	22	22

- maksimum z $t - t = 2..16, d = 2..32,$
- momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 1..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- permutacji – $t = 2..10, d = 2..32,$
- podziałów – $t = 2..16, d = 2..32,$
- pokerowy – $t = 2..16, d = 2..32,$
- potęgowo-dywergencyjny – $\delta = 1..7$
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$

- $t = 5, d = 1..6,$
- $t = 6, d = 1..5,$
- $t = 7, d = 1..4,$
- zgodności bitowej – $t = 2..16, d = 16..32,$
- zgodności Pearsona
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1.$

1.3. Procedury dla ciągów wektorów nakładających się

- częstości dla nakładających się wektorów
 - $t = 2, d = 6..7,$
 - $t = 3, d = 4..5,$
 - $t = 5, d = 3,$
 - $t = 8, d = 2,$
- logarytm prawdopodobieństwa
 - $t = 2, d = 6..7,$
 - $t = 3, d = 4..5,$
 - $t = 5, d = 3,$
 - $t = 8, d = 2,$
- nakładających się permutacji – $t = 3..6, d = 2..32,$
- potęgowo-dywergencyjny – $\delta = 1..7$
 - $t = 2, d = 6, 7,$
 - $t = 3, d = 4, 5,$
 - $t = 5, d = 3,$
 - $t = 8, d = 2,$
- wolnych cel Marsagli
 - $t = 7..9, d = 2,$
 - $t = 4..6, d = 3,$
 - $t = 2..4, d = 4,$
 - $t = 2..3, d = 5,$
 - $t = 2, d = 6..8,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego
 - $t = 2, d = 6..8,$
 - $t = 3, d = 5,$
 - $t = 4, d = 4,$
 - $t = 6, d = 3,$
 - $t = 9, d = 2.$

1.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..8192, 491809..500000,$
- całkowitej liczby serii,
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- częstości serii 0 i 1 łącznie,
- częstości serii 0 i 1 osobno zliczanych,

- częstości w bloku – $M = 512, 1024, 2048, 4096$,
- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024$ bitowego,
- długości ciągu potrzebnego do zaobserwowania po M serii 0 oraz 1 – $M = 512, 1000, 1024, 2048, 4096, 8192, 10000, 16384, 20000, 32768$,
- korelacji wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024$ bitowego,
- liczby skoków w profilu liniowym,
- liczby skoków w profilu liniowym dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768$ bitowe,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146, 158, 166, 178, 194, 202, 206, 214, 218, 226, 254, 256, 262, 274, 278, 298, 302, 314, 326, 334, 346, 358, 362, 382, 386, 394, 398, 422, 446, 454, 458, 466, 478, 482, 502, 512, 514, 526, 538, 542, 554, 562, 566, 586,$

614, 622, 626, 634, 662, 674, 694, 698, 706, 718, 734, 746, 758, 766, 778, 794, 802, 818, 838, 842, 862, 866, 878, 886, 898, 914, 922, 926, 934, 958, 974, 982, 998, 1006, 1018, 1024, 1042, 1046, 1082, 1094, 1114, 1126, 1138, 1142, 1154, 1174, 1186, 1198, 1202, 1214, 1226, 1234, 1238, 1262, 1282, 1286, 1294, 1306, 1318, 1322, 1346, 1354, 1366, 1382, 1402, 1418, 1438, 1454, 1466, 1478, 1486, 1502, 1514, 1522, 1538, 1546, 1574, 1594, 1618, 1622, 1642, 1646, 1654, 1658, 1678, 1706, 1714, 1718, 1726, 1754, 1762, 1766, 1774, 1814, 1822, 1838, 1858, 1874, 1882, 1894, 1906, 1934, 1942, 1954, 1966, 1982, 1994, 2018, 2026, 2038, 2042, 2048, 16384, 31250

- liczby powrotów do 0,
- liczby przejść przez 0 ze zmianą strony,
- łącznego „czasu” odchylenia w prawo,
- maksymalnego odchylenia ścieżki,
- momentu pierwszego osiągnięcia wartości $y = 1..100$,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192, 10000, 16384, 20000, 32768$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192, 10000, 16384, 20000, 32768$ bitowym,
- nakładających się t bitowych wzorców 0 oraz 1 w bloku M bitowym
 - $t = 4, M = 16, 32, 64, 128, 256, 512, 1000, 1024$,
 - $t = 5, M = 32, 64, 128, 256, 512, 1000, 1024$,
 - $t = 6, M = 64, 128, 256, 512, 1000, 1024$,
 - $t = 7, M = 128, 256, 512, 1000, 1024$,
 - $t = 8, M = 256, 512, 1000, 1024$,
 - $t = 9, M = 512, 1000, 1024$,
 - $t = 10, M = 1000, 1024$,
- nienakładających się d bitowych wzorców w bloku M bitowym (wzorce w zapisie heksadecymalnym, ograniczone do początkowych 21):
 - $d = 2, M = 16, 32, 64, 128, 192, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192,$

- 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 2,
- $d = 3$, $M = 16, 32, 64, 128, 192, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 4, 6,$
 - $d = 4$, $M = 16, 32, 64, 128, 192, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152,$

- 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 7, 8, C, E,
- $d = 5$, $M = 32, 64, 128, 192, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, B, F, 10, 14, 18, 1A, 1C, 1E,$
 - $d = 6$, $M = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776,$

- 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, B, D, F, 13, 17, 1F, 20, 28, 2C, 30, 32, 34, 38, 3A, 3C, 3E,
- $d = 7$, $M = 128$, 192, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 7, 9, B, D, F, 13, 15, 17, 1B, 1D, 1F, 23, 27, 2B, 2F, 37, 3F, 40,
 - $d = 8$, $M = 192$, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904,

- 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 13, 15, 17, 19, 1B, 1D, 1F, 23, 25, 27, 2B, 2D, 2F,
- $d = 9$, $M = 320$, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 23, 25, 27, 29, 2B,
 - $d = 10$, $M = 704$, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056,

- 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 23, 25, 27, 29, 2B,
- $d = 11$, $M = 1088$, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
 - $d = 12$, $M = 2368$, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072,

- 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
- $d = 13$, $M = 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,$
 - $d = 14$, $M = 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,$
 - $d = 15$, $M = 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096,$

- 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
- $d = 16$, $M = 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
 - $d = 17$, $M = 131072, 262144, 524288$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
 - $d = 18$, $M = 262144, 524288$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
 - $d = 19$, $M = 524288$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
 - niezależności wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024$ bitowego,

• niezależności wag Hamminga – wariant 2

- $M = 4, d = 1..2,$
- $M = 5, d = 1..3,$
- $M = 7, d = 1..4,$
- $M = 8, d = 1..4,$
- $M = 11, d = 1..6,$
- $M = 13, d = 1..7,$
- $M = 16, d = 1..8,$
- $M = 17, d = 1..9,$
- $M = 19, d = 1..10,$
- $M = 23, d = 1..12,$
- $M = 29, d = 1..15,$
- $M = 31, d = 1..16,$
- $M = 32, d = 1..16,$
- $M = 37, d = 1..19,$
- $M = 41, d = 1..21,$
- $M = 43, d = 1..22,$
- $M = 47, d = 1..24,$
- $M = 53, d = 1..27,$
- $M = 59, d = 1..30,$
- $M = 61, d = 1..31,$
- $M = 64, d = 1..32,$
- $M = 67, d = 1..34,$
- $M = 71, d = 1..36,$
- $M = 73, d = 1..37,$
- $M = 79, d = 1..40,$
- $M = 83, d = 1..42,$
- $M = 89, d = 1..45,$
- $M = 97, d = 1..49,$
- $M = 101, d = 1..51,$
- $M = 103, d = 1..52,$
- $M = 107, d = 1..54,$
- $M = 109, d = 1..55,$
- $M = 113, d = 1..57,$
- $M = 127, d = 1..64,$
- $M = 128, d = 1..64,$
- $M = 131, d = 1..66,$
- $M = 137, d = 1..69,$
- $M = 139, d = 1..70,$
- $M = 149, d = 1..75,$
- $M = 151, d = 1..76,$
- $M = 157, d = 1..79,$
- $M = 163, d = 1..82,$
- $M = 167, d = 1..84,$
- $M = 173, d = 1..87,$
- $M = 179, d = 1..90,$
- $M = 181, d = 1..91,$
- $M = 191, d = 1..96,$
- $M = 193, d = 1..97,$
- $M = 197, d = 1..99,$
- $M = 199, d = 1..100,$
- $M = 211, d = 1..106,$
- $M = 223, d = 1..112,$
- $M = 227, d = 1..114,$
- $M = 229, d = 1..115,$
- $M = 233, d = 1..117,$
- $M = 239, d = 1..120,$
- $M = 241, d = 1..121,$
- $M = 251, d = 1..126,$
- $M = 256, d = 1..128,$
- $M = 257, d = 1..129,$
- $M = 263, d = 1..132,$
- $M = 269, d = 1..135,$
- $M = 271, d = 1..136,$
- $M = 277, d = 1..139,$
- $M = 281, d = 1..141,$
- $M = 283, d = 1..142,$
- $M = 293, d = 1..147,$
- $M = 307, d = 1..154,$
- $M = 311, d = 1..156,$
- $M = 313, d = 1..157,$

- $M = 317, d = 1..159,$
- $M = 331, d = 1..166,$
- $M = 337, d = 1..169,$
- $M = 347, d = 1..174,$
- $M = 349, d = 1..175,$
- $M = 353, d = 1..177,$
- $M = 359, d = 1..180,$
- $M = 367, d = 1..184,$
- $M = 373, d = 1..187,$
- $M = 379, d = 1..190,$
- $M = 383, d = 1..192,$
- $M = 389, d = 1..195,$
- $M = 397, d = 1..199,$
- $M = 401, d = 1..201,$
- $M = 409, d = 1..205,$
- $M = 419, d = 1..210,$
- $M = 421, d = 1..211,$
- $M = 431, d = 1..216,$
- $M = 433, d = 1..217,$
- $M = 439, d = 1..220,$
- $M = 443, d = 1..222,$
- $M = 449, d = 1..225,$
- $M = 457, d = 1..229,$
- $M = 461, d = 1..231,$
- $M = 463, d = 1..232,$
- $M = 467, d = 1..234,$
- $M = 479, d = 1..240,$
- $M = 487, d = 1..244,$
- $M = 491, d = 1..246,$
- $M = 499, d = 1..250,$
- $M = 503, d = 1..252,$
- $M = 509, d = 1..255,$
- $M = 512, d = 1..256,$
- $M = 521, d = 1..261,$
- $M = 523, d = 1..262,$
- $M = 541, d = 1..271,$
- $M = 547, d = 1..274,$
- $M = 557, d = 1..279,$
- $M = 563, d = 1..282,$
- $M = 569, d = 1..285,$
- $M = 571, d = 1..286,$
- $M = 577, d = 1..289,$
- $M = 587, d = 1..294,$
- $M = 593, d = 1..297,$
- $M = 599, d = 1..300,$
- $M = 601, d = 1..301,$
- $M = 607, d = 1..304,$
- $M = 613, d = 1..307,$
- $M = 617, d = 1..309,$
- $M = 619, d = 1..310,$
- $M = 631, d = 1..316,$
- $M = 641, d = 1..321,$
- $M = 643, d = 1..322,$
- $M = 647, d = 1..324,$
- $M = 653, d = 1..327,$
- $M = 659, d = 1..330,$
- $M = 661, d = 1..331,$
- $M = 673, d = 1..337,$
- $M = 677, d = 1..339,$
- $M = 683, d = 1..342,$
- $M = 691, d = 1..346,$
- $M = 701, d = 1..351,$
- $M = 709, d = 1..355,$
- $M = 719, d = 1..360,$
- $M = 727, d = 1..364,$
- $M = 733, d = 1..367,$
- $M = 739, d = 1..370,$
- $M = 743, d = 1..372,$
- $M = 751, d = 1..376,$
- $M = 757, d = 1..379,$
- $M = 761, d = 1..381,$
- $M = 769, d = 1..385,$

- $M = 773, d = 1..387,$
- $M = 787, d = 1..394,$
- $M = 797, d = 1..399,$
- $M = 809, d = 1..405,$
- $M = 811, d = 1..406,$
- $M = 821, d = 1..411,$
- $M = 823, d = 1..412,$
- $M = 827, d = 1..414,$
- $M = 829, d = 1..415,$
- $M = 839, d = 1..420,$
- $M = 853, d = 1..427,$
- $M = 857, d = 1..429,$
- $M = 859, d = 1..430,$
- $M = 863, d = 1..432,$
- $M = 877, d = 1..439,$
- $M = 881, d = 1..441,$
- $M = 883, d = 1..442,$
- $M = 887, d = 1..444,$
- $M = 907, d = 1..454,$
- $M = 911, d = 1..456,$
- $M = 919, d = 1..460,$
- $M = 929, d = 1..465,$
- $M = 937, d = 1..469,$
- $M = 941, d = 1..471,$
- $M = 947, d = 1..474,$
- $M = 953, d = 1..477,$
- $M = 967, d = 1..484,$
- $M = 971, d = 1..486,$
- $M = 977, d = 1..489,$
- $M = 983, d = 1..492,$
- $M = 991, d = 1..496,$
- $M = 997, d = 1..499,$
- $M = 1009, d = 1..505,$
- $M = 1013, d = 1..507,$
- $M = 1019, d = 1..510,$
- $M = 1021, d = 1..511,$
- $M = 1024, d = 1..512,$
- rozkładu skoków w profilu liniowym,
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768$ bitowe,
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 6..14,$
- wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599,$

601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024 bitowego,

- wypełniania drzewa binarnego o wysokości $h = 3..10$ na podstawie ciągu bitów
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

2. Procedury testowe przydatne do badania pojedynczych ciągów

2.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32$,
- autokorelacji dla ciągu bloków

○ $d = 2, t = 1..249999$,	○ $d = 18, t = 1..27777$,
○ $d = 3, t = 1..166666$,	○ $d = 19, t = 1..26315$,
○ $d = 4, t = 1..124999$,	○ $d = 20, t = 1..24999$,
○ $d = 5, t = 1..99999$,	○ $d = 21, t = 1..23809$,
○ $d = 6, t = 1..83332$,	○ $d = 22, t = 1..22726$,
○ $d = 7, t = 1..71428$,	○ $d = 23, t = 1..21738$,
○ $d = 8, t = 1..62499$,	○ $d = 24, t = 1..20832$,
○ $d = 9, t = 1..55555$,	○ $d = 25, t = 1..19999$,
○ $d = 10, t = 1..49999$,	○ $d = 26, t = 1..19230$,
○ $d = 11, t = 1..45454$,	○ $d = 27, t = 1..18518$,
○ $d = 12, t = 1..41666$,	○ $d = 28, t = 1..17856$,
○ $d = 13, t = 1..38461$,	○ $d = 29, t = 1..17240$,
○ $d = 14, t = 1..35713$,	○ $d = 30, t = 1..16666$,
○ $d = 15, t = 1..33332$,	○ $d = 31, t = 1..16128$,
○ $d = 16, t = 1..31249$,	○ $d = 32, t = 1..15624$,
○ $d = 17, t = 1..29411$,	
- Corona – $d = 3..11$,
- Cramera-von Misesa – $d = 16..32$,
- częstości dla podprzedziału
 - $d = 16..22, [1/128; 1/4 + 1/128), M = 32, 256$,
 - $d = 16..22, [1/64; 1/2 + 1/128), M = 32, 256$,
 - $d = 16..22, [1/2 - 1/128; 3/4), M = 32, 256$,
 - $d = 23..32, [1/128; 1/4 + 1/128), M = 32$,

- $d = 23..32, [1/64; 1/2 + 1/128), M = 32,$
- $d = 23..32, [1/2 - 1/128; 3/4), M = 32,$
- gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka ośmiościenna – zgodności dla liczby zwycięstw,
 - kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka sześcienna – zgodności dla liczby zwycięstw,
- kolekcjonera – $d = 2..8,$
- Kołmogorowa – $d = 16..32,$
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- Kuipera – $d = 16..32,$
- największej liczby elementów w przedziale – podciągi 1 kiB
 - $D = 1/2, d = 4..7, 9, 10, 12..15,$
 - $D = 0,4, d = 4..10, 12, 13,$
 - $D = 0,3, d = 4..10,$
 - $D = 1/4, d = 4..8,$
 - $D = 0,2, d = 4..6,$
 - $D = 1/8, d = 4..5,$
 - $D = 1/16, d = 4,$
- odstępów dni urodzin
 - $d = 20, n = 128,$
 - $d = 21..23, n = 256,$
 - $d = 23..24, n = 512,$
- pochodnej ciągu bloków – $d = 2..32$
 - pierwsza pochodna,
 - druga pochodna,
- porządku książek

○ $d = 4, k = 1..3,$ ○ $d = 5, k = 1..4,$ ○ $d = 6, k = 1..5,$ ○ $d = 7, k = 1..6,$ ○ $d = 8, k = 1..7,$ ○ $d = 9, k = 1..8,$	○ $d = 10, k = 2..9,$ ○ $d = 11, k = 1..10,$ ○ $d = 12, k = 1..11,$ ○ $d = 13, k = 1..12,$ ○ $d = 14..17, k = 1..13,$ ○ $d = 18, k = 1..11,$
--	---

- $d = 19, k = 1..10,$
- $d = 20..22, k = 2..10,$
- $d = 23..25, k = 2..9,$
- $d = 26, k = 1..9,$
- serii malejących z pomijaniem w przód – $d = 6..32,$
- serii malejących z pomijaniem w tył – $d = 6..32,$
- serii rosnących z pomijaniem w przód – $d = 6..32,$
- serii rosnących z pomijaniem w tył – $d = 6..32,$
- stosu książek
 - $d = 4, k = 1..3,$
 - $d = 5, k = 1..4,$
 - $d = 6, k = 1..5,$
 - $d = 7, k = 1..6,$
 - $d = 8, k = 1..7,$
 - $d = 9, k = 1..8,$
 - $d = 10, k = 1..9,$
 - $d = 11..12, k = 1..8,$
 - $d = 13, k = 1..10,$
 - $d = 14, k = 1..8,$
 - $d = 15, k = 2..9,$
 - $d = 16, k = 1..10,$
 - $d = 17, k = 1..7,$
 - $d = 18, k = 1..8,$
 - $d = 19, k = 1..7,$
 - $d = 20, k = 2..8,$
 - $d = 21, k = 1..9,$
 - $d = 22..24, k = 2..7,$
 - $d = 25, k = 2..10,$
 - $d = 26, k = 1..9,$
- sum do G
 - $d = 16..32, G = 16, 32, 64, 128, 256,$
 - $d = 16..23, G = 512,$
- sumy kwadratów odstępów $m = 1, 2, 3, 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024, 8192, 15625$ elementowych bloków $d = 17..32$ bitowych,
 - bez zapętlenia – za wyjątkiem:
 - $m = 1..2, d = 17..18,$
 - $m = 3..8, d = 17,$
 - z zapętleniem – za wyjątkiem:
 - $m = 1, 8192, 15625, d = 17..32,$

- $m = 2, d = 17..18,$
- $m = 3..8, d = 17,$

- sumy logarytmów odstępów bez zapętlania

- $m = 1, d = 31..32,$
- $m = 2, d = 23..32,$
- $m = 3, 4, 5, d = 20..32,$
- $m = 7, 8, 11, d = 19..32,$
- $m = 13, 16, 17, 19, 23, d = 18..32,$
- $m = 29, 31, 32, d = 17..32,$
- $m = 37, d = 17..26,$
- $m = 41, 43, d = 17..25,$
- $m = 47, d = 17..20,$
- $m = 53, d = 18,$

- sumy logarytmów odstępów z zapętlaniem

- $m = 2, d = 23..32,$
- $m = 3, 4, 5, d = 20..32,$
- $m = 7, 8, 11, d = 19..32,$
- $m = 13, 16, 17, 19, 23, d = 18..32,$
- $m = 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, d = 17..32,$
- $m = 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, d = 17..31,$
- $m = 229, 233, 239, 241, 251, 256, 257, d = 17..29, 31,$
- $m = 263, 269, 271, 277, 281, 283, 293, d = 17..28, 31,$
- $m = 307, 311, 313, 317, d = 17..23, 25, 28, 31,$
- $m = 331, 337, 347, 349, 353, 359, d = 17..19, 21, 22, 25,$
- $m = 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, d = 17..19, 22,$
- $m = 431, d = 17..19,$
- $m = 433, 439, 443, 449, 457, d = 17..18,$
- $m = 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, d = 18,$

- ściskania – $d = 10..20,$
- uniwersalny Maurera – $d = 1..11,$
- wag Hamminga w ciągu bloków

- | | |
|---|---|
| <ul style="list-style-type: none"> ○ $d = 2, t = 2..5, k = 3,$ ○ $d = 3$ <ul style="list-style-type: none"> ▪ $t = 2..4, k = 2, 4,$ ▪ $t = 5..8, k = 2,$ | <ul style="list-style-type: none"> ○ $d = 4$ <ul style="list-style-type: none"> ▪ $t = 2..3, k = 3, 5,$ ▪ $t = 4..5, k = 3,$ |
|---|---|

- $d = 5$
 - $t = 2..3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 6$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 7$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 8$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 9$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 10$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 11$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 12$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 13$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 14$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 15$
 - $t = 2, k = 2, 4, 6, 8, 10,$
12,
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 16$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 20$
 - $t = 2, k = 3, 5, 7, 9, 11,$
13,
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 32$
 - $t = 2, k = 3, 5, 7, 9, 11,$
13,
 - $t = 3, k = 3,$
- wypełniania drzewa binarnego o wysokości $h = 3..11$ na podstawie ciągu bloków $d = 4..32$ bitowych
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

2.2. Procedury dla ciągów wektorów nienakładających się

- iloczynów

- $t = 2, d = 11..32,$
- $t = 3..9, d = 12..32,$
- $t = 10..16, d = 13..32,$

- kolizji

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 15..21,$ | ○ $t = 5, d = 3..4,$ |
| ○ $t = 2, d = 8..10,$ | ○ $t = 6..7, d = 3,$ |
| ○ $t = 3, d = 5..7,$ | ○ $t = 8..10, d = 2,$ |
| ○ $t = 4, d = 4..5,$ | ○ $t = 15..21, d = 1,$ |

- kolizji indeksów maksymalnych elementów

- $t = 2, d = 13..22,$
- $t = 3..16, d = 10..22,$

- liczby komórek z co najmniej $b = 1$ punktem dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 15..21,$ | ○ $t = 5, d = 3..4,$ |
| ○ $t = 2, d = 8..10,$ | ○ $t = 6..7, d = 3,$ |
| ○ $t = 3, d = 5..7,$ | ○ $t = 8..10, d = 2,$ |
| ○ $t = 4, d = 4..5,$ | ○ $t = 15..21, d = 1,$ |

liczby komórek z co najmniej $b = 2$ punktami dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 14..20,$ | ○ $t = 5, d = 3..4,$ |
| ○ $t = 2, d = 7..10,$ | ○ $t = 6, d = 3,$ |
| ○ $t = 3, d = 5..6,$ | ○ $t = 7..10, d = 2,$ |
| ○ $t = 4, d = 4..5,$ | ○ $t = 14..20, d = 1,$ |

liczby komórek z co najmniej $b = 3$ punktami dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 14..17,$ | ○ $t = 5, d = 3,$ |
| ○ $t = 2, d = 7..8,$ | ○ $t = 7..8, d = 2,$ |
| ○ $t = 3, d = 5,$ | ○ $t = 14..17, d = 1,$ |
| ○ $t = 4, d = 4,$ | |

liczby komórek z co najmniej $b = 4$ punktami dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 14..16,$ | ○ $t = 5, d = 3,$ |
| ○ $t = 2, d = 7..8,$ | ○ $t = 7..8, d = 2,$ |
| ○ $t = 3, d = 5,$ | ○ $t = 14..16, d = 1,$ |
| ○ $t = 4, d = 4,$ | |

liczby komórek z co najmniej $b = 5..6$ punktami dla wektorów:

- | | |
|------------------------|-------------------|
| ○ $t = 1, d = 14..15,$ | ○ $t = 3, d = 5,$ |
| ○ $t = 2, d = 7,$ | ○ $t = 5, d = 3,$ |

$$\circ \quad t = 7, d = 2,$$

$$\circ \quad t = 14..15, d = 1,$$

liczby komórek z co najmniej $b = 7..10$ punktami dla wektorów:

$$\circ \quad t = 1, d = 14,$$

$$\circ \quad t = 7, d = 2,$$

$$\circ \quad t = 2, d = 7,$$

$$\circ \quad t = 14, d = 1,$$

- liczby komórek z dokładnie $b = 0$ punktami dla wektorów:

$$\circ \quad t = 1, d = 15..21,$$

$$\circ \quad t = 5, d = 3..4,$$

$$\circ \quad t = 2, d = 8..10,$$

$$\circ \quad t = 6..7, d = 3,$$

$$\circ \quad t = 3, d = 5..7,$$

$$\circ \quad t = 8..10, d = 2,$$

$$\circ \quad t = 4, d = 4..5,$$

$$\circ \quad t = 15..21, d = 1,$$

liczby komórek z dokładnie $b = 1$ punktem dla wektorów:

$$\circ \quad t = 1, d = 15..21,$$

$$\circ \quad t = 5, d = 3..4,$$

$$\circ \quad t = 2, d = 8..10,$$

$$\circ \quad t = 6..7, d = 3,$$

$$\circ \quad t = 3, d = 5..7,$$

$$\circ \quad t = 8..10, d = 2,$$

$$\circ \quad t = 4, d = 4..5,$$

$$\circ \quad t = 15..21, d = 1,$$

liczby komórek z dokładnie $b = 2$ punktami dla wektorów:

$$\circ \quad t = 1, d = 14..20,$$

$$\circ \quad t = 6, d = 3,$$

$$\circ \quad t = 2, d = 7..10,$$

$$\circ \quad t = 7, d = 2,$$

$$\circ \quad t = 3, d = 5..6,$$

$$\circ \quad t = 8..10, d = 2,$$

$$\circ \quad t = 4, d = 4..5,$$

$$\circ \quad t = 14..20, d = 1,$$

$$\circ \quad t = 5, d = 3..4,$$

liczby komórek z dokładnie $b = 3$ punktami dla wektorów:

$$\circ \quad t = 1, d = 14..18,$$

$$\circ \quad t = 5..6, d = 3,$$

$$\circ \quad t = 2, d = 7..9,$$

$$\circ \quad t = 7..9, d = 2,$$

$$\circ \quad t = 3, d = 5..6,$$

$$\circ \quad t = 14..18, d = 1,$$

$$\circ \quad t = 4, d = 4,$$

liczby komórek z dokładnie $b = 4$ punktami dla wektorów:

$$\circ \quad t = 1, d = 14..16,$$

$$\circ \quad t = 5, d = 3,$$

$$\circ \quad t = 2, d = 7..8,$$

$$\circ \quad t = 7..8, d = 2,$$

$$\circ \quad t = 3, d = 5,$$

$$\circ \quad t = 14..16, d = 1,$$

$$\circ \quad t = 4, d = 4,$$

liczby komórek z dokładnie $b = 5$ punktami dla wektorów:

$$\circ \quad t = 1, d = 14..15,$$

$$\circ \quad t = 5, d = 3,$$

$$\circ \quad t = 2, d = 7,$$

$$\circ \quad t = 7, d = 2,$$

$$\circ \quad t = 3, d = 5,$$

$$\circ \quad t = 14..15, d = 1,$$

liczby komórek z dokładnie $b = 6..9$ punktami dla wektorów:

$$\circ \quad t = 1, d = 14,$$

$$\circ \quad t = 2, d = 7,$$

○ $t = 7, d = 2,$

○ $t = 14, d = 1,$

• logarytm prawdopodobieństwa

○ $t = 1, d = 2..22,$

○ $t = 5, d = 1..4,$

○ $t = 2, d = 1..11,$

○ $t = 6..7, d = 1..3,$

○ $t = 3, d = 1..7,$

○ $t = 8..11, d = 1..2,$

○ $t = 4, d = 1..5,$

○ $t = 12..22, d = 1,$

• m najbliższych par

Tabela B.2. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka euklidesowa

$d \backslash t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16		72	62	55	51	47	44	41	39	37	36	34	33	32	31
17		70	60	54	49	45	42	40	38	36	35	33	32	31	30
18		68	58	52	48	44	41	39	37	35	34	32	31	30	29
19		66	57	51	46	43	40	38	36	34	33	31	30	29	28
20		64	55	50	45	42	39	37	35	33	32	31	29	28	27
21	77	62	54	48	44	41	38	36	34	32	31	30	29	28	27
22	75	61	53	47	43	40	37	35	33	32	30	29	28	27	26
23	73	60	52	46	42	39	36	34	32	31	30	28	27	26	26
24	72	58	51	45	41	38	36	34	32	30	29	28	27	26	25
25	70	57	50	44	40	37	35	33	31	30	28	27	26	25	25
26	69	56	49	43	40	37	34	32	31	29	28	27	26	25	24
27	68	55	48	43	39	36	34	32	30	29	27	26	25	24	24
28	66	54	47	42	38	35	33	31	29	28	27	26	25	24	23
29	65	53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24	23	22	22

Tabela B.3. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka nieskończoność

$d \backslash t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
19									36	34	33	31	30	29	28
20						42	39	37	35	33	32	31	29	28	27
21				48	44	41	38	36	34	32	31	30	29	28	27
22			53	47	43	40	37	35	33	32	30	29	28	27	26
23			52	46	42	39	36	34	32	31	30	28	27	26	26
24		58	51	45	41	38	36	34	32	30	29	28	27	26	25
25		57	50	44	40	37	35	33	31	30	28	27	26	25	25
26		56	49	43	40	37	34	32	31	29	28	27	26	25	24
27	68	55	48	43	39	36	34	32	30	29	27	26	25	24	24
28	66	54	47	42	38	35	33	31	29	28	27	26	25	24	23
29	65	53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24			

- maksimum z t
 - $t = 2, d = 14..32,$
 - $t = 3..8, d = 15..32,$
 - $t = 9..16, d = 16..32,$
- momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów

○ $t = 1, d = 1..30,$ ○ $t = 2, d = 1..15,$ ○ $t = 3, d = 1..10,$ ○ $t = 4, d = 1..7,$ ○ $t = 5, d = 1..6,$	○ $t = 6, d = 1..5,$ ○ $t = 7, d = 1..4,$ ○ $t = 8..10, d = 1..3,$ ○ $t = 11..15, d = 1..2,$ ○ $t = 16..30, d = 1,$
--	--
- momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów

○ $t = 1, d = 2..30,$ ○ $t = 2, d = 1..15,$ ○ $t = 3, d = 1..10,$ ○ $t = 4, d = 1..7,$ ○ $t = 5, d = 1..6,$	○ $t = 6, d = 1..5,$ ○ $t = 7, d = 1..4,$ ○ $t = 8..10, d = 1..3,$ ○ $t = 11..15, d = 1..2,$ ○ $t = 16..30, d = 1,$
--	--
- permutacji – $t = 2..10, d = 2..32,$
- podziałów

○ $t = 2, d = 2..6,$ ○ $t = 3, d = 2..9,$ ○ $t = 4..7, d = 2..10,$ ○ $t = 8, d = 2..9,$ ○ $t = 9, d = 2..11,$	○ $t = 10, d = 2..10,$ ○ $t = 11, d = 2..9,$ ○ $t = 12..14, d = 2..11,$ ○ $t = 15, d = 2..12,$ ○ $t = 16, d = 3..10,$
--	--
- pokerowy

○ $t = 2, d = 2..6,$ ○ $t = 3, d = 2..9,$ ○ $t = 4..7, d = 2..10,$ ○ $t = 8, d = 2..9,$ ○ $t = 9, d = 2..11,$	○ $t = 10, d = 2..10,$ ○ $t = 11, d = 2..9,$ ○ $t = 12..14, d = 2..11,$ ○ $t = 15, d = 2..12,$ ○ $t = 16, d = 2..10,$
--	--
- potęgowo-dywergencyjny – $\delta = 1$

○ $t = 1, d = 2..23,$ ○ $t = 2, d = 1..11,$ ○ $t = 3, d = 1..7,$ ○ $t = 4, d = 1..5,$	○ $t = 5, d = 1..4,$ ○ $t = 6, 7, d = 1..3,$ ○ $t = 8..11, d = 1..2,$ ○ $t = 12..23, d = 1,$
--	---

potęgowo-dywergencyjny – $\delta = 2, 3$

- $t = 1, d = 4..23,$
- $t = 2, d = 2..11,$
- $t = 3, d = 2..7,$
- $t = 4, d = 1..5,$
- $t = 5, d = 1..4,$
- $t = 6, 7, d = 1..3,$
- $t = 8..11, d = 1..2,$
- $t = 12..23, d = 1,$

potęgowo-dywergencyjny – $\delta = 4$

- $t = 1, d = 4..21,$
- $t = 2, d = 2..10,$
- $t = 3, d = 2..7,$
- $t = 4, d = 1..5,$
- $t = 5, d = 1..4,$
- $t = 6, 7, d = 1..3,$
- $t = 8..10, d = 1..2,$
- $t = 11..21, d = 1,$

potęgowo-dywergencyjny – $\delta = 5$

- $t = 1, d = 4..15,$
- $t = 2, d = 2..7,$
- $t = 3, d = 2..5,$
- $t = 4, 5, d = 1..3,$
- $t = 6, 7, d = 1..2,$
- $t = 8..15, d = 1,$

potęgowo-dywergencyjny – $\delta = 6$

- $t = 1, d = 5..13,$
- $t = 2, d = 3..6,$
- $t = 3, d = 2..4,$
- $t = 4, d = 2..3,$
- $t = 5, d = 1..2,$
- $t = 6, d = 1..2,$
- $t = 7..13, d = 1,$

potęgowo-dywergencyjny – $\delta = 7$

- $t = 1, d = 5..12,$
- $t = 2, d = 3..6,$
- $t = 3, d = 2..4,$
- $t = 4, d = 2..3,$
- $t = 5, d = 1..2,$
- $t = 6, d = 1..2,$
- $t = 7..12, d = 1,$

• zgodności Pearsona

- $t = 1, d = 2..30,$
- $t = 2, d = 1..15,$
- $t = 3, d = 1..10,$
- $t = 4, d = 1..7,$
- $t = 5, d = 1..6,$
- $t = 6, d = 1..5,$
- $t = 7, d = 1..4,$
- $t = 8..10, d = 1..3,$
- $t = 11..15, d = 1..2,$
- $t = 16..30, d = 1,$

2.3. Procedury dla ciągów wektorów nakładających się

• częstości dla nakładających się wektorów

- $t = 2, d = 6..7,$
- $t = 3, d = 4..5,$
- $t = 5, d = 3,$
- $t = 8, d = 2,$

• nakładających się permutacji – $t = 3..6, d = 2..32,$

- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 2, d = 6..7,$
 - $t = 3, d = 4..5,$
 - $t = 5, d = 3,$
 - $t = 8, d = 2,$
- wolnych cel Marsagli
 - $t = 8, d = 2,$
 - $t = 9, d = 2,$
 - $t = 6, d = 3,$
 - $t = 4, d = 4,$
 - $t = 3, d = 5,$
 - $t = 2, d = 8,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego
 - $t = 2, d = 8,$
 - $t = 4, d = 4,$
 - $t = 6, d = 3,$
 - $t = 9, d = 2.$

2.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..500000,$
- całkowitej liczby serii,
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- częstości serii 0 i 1 łącznie,
- częstości w bloku – $M = 512, 1024, 2048, 4096,$
- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997$ bitowego,
- długości ciągu potrzebnego do zaobserwowania po M serii 0 oraz 1 – $M = 1024, 2048, 4096, 8192, 10000,$
- liczby skoków w profilu liniowym,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146, 158, 166, 178, 194, 202, 206, 214, 218, 226, 254, 256, 262, 274, 278, 298, 302, 314, 326, 334, 346, 358, 362, 382, 386, 394, 398, 422, 446, 454, 458, 466, 478, 482, 502, 512, 514, 526, 538, 542, 554, 562, 566, 586, 614, 622, 626, 634, 662, 674, 694, 698, 706, 718, 734, 746, 758, 766, 778, 794, 802, 818, 838, 842, 862, 866, 878, 886, 898, 914, 922, 926, 934, 958, 974, 982, 998, 1006, 1018, 1024, 1042, 1046, 1082, 1094, 1114, 1126, 1138, 1142, 1154, 1174, 1186, 1198, 1202, 1214, 1226, 1234,$

1238, 1262, 1282, 1286, 1294, 1306, 1318, 1322, 1346, 1354, 1366, 1382, 1402, 1418, 1438, 1454, 1466, 1478, 1486, 1502, 1514, 1522, 1538, 1546, 1574, 1594, 1618, 1622, 1642, 1646, 1654, 1658, 1678, 1706, 1714, 1718, 1726, 1754, 1762, 1766, 1774, 1814, 1822, 1838, 1858, 1874, 1882, 1894, 1906, 1934, 1942, 1954, 1966, 1982, 1994, 2018, 2026, 2038, 2042, 2048, 16384

- liczby powrotów do 0,
- liczby przejść przez 0 ze zmianą strony,
- łącznego „czasu” odchylenia w prawo,
- maksymalnego odchylenia ścieżki,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192, 10000, 16384$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192, 10000, 16384$ bitowym,
- nakładających się t bitowych wzorców 0 oraz 1 w bloku M bitowym
 - $t = 4, M = 16, 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 5, M = 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 6, M = 64, 128, 256, 512, 1000, 1024,$
 - $t = 7, M = 128, 256, 512, 1000, 1024,$
 - $t = 8, M = 256, 512, 1000, 1024,$
 - $t = 9, M = 512, 1000, 1024,$
 - $t = 10, M = 1000, 1024,$
- nienakładających się d bitowych wzorców w bloku M bitowym (wzorce w zapisie heksadecymalnym, ograniczone do początkowych 21):
 - $d = 2, M = 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056,$

- 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 2,
- $d = 3$, $M = 64$, 128, 192, 256, 320, 448, 512, 704, 832, 1000, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 4, 6,
 - $d = 4$, $M = 1472$, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 7, 8, C, E,

- $d = 5$, $M = 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, B, F, 10, 14, 18, 1A, 1C, 1E,$
- $d = 6$, $M = 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, B, D, F, 13, 17, 1F, 20, 28, 2C, 30, 32, 34, 38, 3A, 3C, 3E,$
- $d = 7$, $M = 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896,$

- 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 7, 9, B, D, F, 13, 15, 17, 1B, 1D, 1F, 23, 27, 2B, 2F, 37, 3F, 40,
- $d = 8$, $M = 3008$, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 13, 15, 17, 19, 1B, 1D, 1F, 23, 25, 27, 2B, 2D, 2F,
 - $d = 9$, $M = 4096$, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592,

- 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, 524288, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 23, 25, 27, 29, 2B,
- $d = 10$, $M =$ 8128, 8192, 8384, 8768, 8896, 9536, 9664, 10048, 10432, 10688, 11072, 11456, 11584, 12224, 12352, 12608, 12736, 13504, 14272, 14528, 14656, 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 23, 25, 27, 29, 2B,
 - $d = 11$, $M =$ 14912, 15296, 15424, 16064, 16384, 16448, 16832, 17216, 17344, 17728, 17984, 18112, 18752, 19648, 19904, 20032, 20288, 21184, 21568, 22208, 22336, 22592, 22976, 23488, 23872, 24256, 24512, 24896, 25408, 25664, 26176, 26816, 26944, 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304,

- 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
- $d = 12$, $M = 27584, 27712, 28096, 28352, 28736, 29248, 29504, 29632, 29888, 30656, 31168, 31424, 31936, 32192, 32576, 32768, 33344, 33472, 34624, 35008, 35648, 36032, 36416, 36544, 36928, 37568, 37952, 38336, 38464, 38848, 39232, 39488, 39616, 40384, 41024, 41152, 41408, 41792, 42176, 42304, 43072, 43328, 43712, 44224, 44864, 45376, 46016, 46528, 46912, 47296, 47552, 48064, 48448, 48704, 49216, 49472, 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
 - $d = 13$, $M = 50368, 51008, 51776, 51904, 52544, 52672, 52928, 53056, 53696, 54592, 54848, 54976, 55232, 56128, 56384, 56512, 56768, 58048, 58304, 58816, 59456, 59968, 60224, 60608, 60992, 61888, 62144, 62528, 62912, 63424, 63808, 64576, 64832, 65216, 65344, 65536, 131072, 262144$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
 - $d = 14$, $M = 131072$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7$ bitowego,
 - niezależności wag Hamminga – wariant 2
 - $M = 4, d = 1..2$,
 - $M = 5, d = 1..3$,
 - $M = 7, d = 2..4$,
 - $M = 8, d = 2..4$,
 - $M = 11, d = 3..6$,
 - $M = 13, d = 4..7$,
 - $M = 16, d = 5..8$,
 - $M = 17, d = 5..9$,
 - $M = 19, d = 6..10$,
 - $M = 23, d = 8..12$,
 - $M = 29, d = 10..15$,
 - $M = 31, d = 11..16$,
 - $M = 32, d = 12..16$,
 - $M = 37, d = 14..19$,
 - $M = 41, d = 16..21$,
 - $M = 43, d = 16..22$,
 - $M = 47, d = 18..24$,
 - $M = 53, d = 21..27$,
 - $M = 59, d = 23..30$,
 - $M = 61, d = 24..31$,
 - $M = 64, d = 26..32$,
 - $M = 67, d = 27..34$,
 - $M = 71, d = 28..36$,
 - $M = 73, d = 30..37$,
 - $M = 79, d = 32..40$,
 - $M = 83, d = 34..42$,
 - $M = 89, d = 37..45$,
 - $M = 97, d = 40..49$,

- $M = 101, d = 43..51,$
- $M = 103, d = 44..52,$
- $M = 107, d = 46..54,$
- $M = 109, d = 46..55,$
- $M = 113, d = 50..57,$
- $M = 127, d = 54..64,$
- $M = 128, d = 55..64,$
- $M = 131, d = 56..66,$
- $M = 137, d = 60..69,$
- $M = 139, d = 60..70,$
- $M = 149, d = 65..75,$
- $M = 151, d = 66..76,$
- $M = 157, d = 69..79,$
- $M = 163, d = 73..82,$
- $M = 167, d = 73..84,$
- $M = 173, d = 76..87,$
- $M = 179, d = 79..90,$
- $M = 181, d = 80..90,$
- $M = 191, d = 85..96,$
- $M = 193, d = 87..96,$
- $M = 197, d = 88..98,$
- $M = 199, d = 89..99,$
- $M = 211, d = 95..105,$
- $M = 223, d = 100..111,$
- $M = 227, d = 102..113,$
- $M = 229, d = 103..114,$
- $M = 233, d = 105..116,$
- $M = 239, d = 108..119,$
- $M = 241, d = 109..120,$
- $M = 251, d = 114..125,$
- $M = 256, d = 116..128,$
- $M = 257, d = 117..128,$
- $M = 263, d = 120..131,$
- $M = 269, d = 122..134,$
- $M = 271, d = 124..135,$
- $M = 277, d = 126..138,$
- $M = 281, d = 128..140,$
- $M = 283, d = 130..141,$
- $M = 293, d = 135..146,$
- $M = 307, d = 141..153,$
- $M = 311, d = 143..155,$
- $M = 313, d = 144..156,$
- $M = 317, d = 146..158,$
- $M = 331, d = 153..165,$
- $M = 337, d = 156..168,$
- $M = 347, d = 161..173,$
- $M = 349, d = 161..174,$
- $M = 353, d = 164..176,$
- $M = 359, d = 166..179,$
- $M = 367, d = 171..183,$
- $M = 373, d = 173..186,$
- $M = 379, d = 176..189,$
- $M = 383, d = 179..191,$
- $M = 389, d = 181..194,$
- $M = 397, d = 186..198,$
- $M = 401, d = 187..200,$
- $M = 409, d = 191..204,$
- $M = 419, d = 196..209,$
- $M = 421, d = 198..210,$
- $M = 431, d = 201..215,$
- $M = 433, d = 202..216,$
- $M = 439, d = 206..219,$
- $M = 443, d = 207..221,$
- $M = 449, d = 211..224,$
- $M = 457, d = 215..228,$
- $M = 461, d = 216..230,$
- $M = 463, d = 217..231,$
- $M = 467, d = 218..233 \setminus 219,$
- $M = 479, d = 224..239 \setminus 225,$
- $M = 487, d = 230..243,$
- $M = 491, d = 231..245 \setminus 232,$
- $M = 499, d = 236..249,$

- $M = 503, d = 237..251,$
- $M = 509, d = 241..254,$
- $M = 512, d = 242..256,$
- $M = 521, d = 245..260 \setminus 246,$
- $M = 523, d = 248..261,$
- $M = 541, d = 256..270,$
- $M = 547, d = 258..273,$
- $M = 557, d = 263..278 \setminus 264,$
- $M = 563, d = 267..281,$
- $M = 569, d = 271..284,$
- $M = 571, d = 272..285,$
- $M = 577, d = 274..288,$
- $M = 587, d = 278..293,$
- $M = 593, d = 283..296,$
- $M = 599, d = 283..299,$
- $M = 601, d = 284..300,$
- $M = 607, d = 289..303,$
- $M = 613, d = 290..306,$
- $M = 617, d = 292..308,$
- $M = 619, d = 294..309,$
- $M = 631, d = 300..315,$
- $M = 641, d = 306..320,$
- $M = 643, d = 307..321,$
- $M = 647, d = 308..323,$
- $M = 653, d = 310, 313..326,$
- $M = 659, d = 315..329,$
- $M = 661, d = 316..330,$
- $M = 673, d = 320..336 \setminus 321,$
- $M = 677, d = 323..338,$
- $M = 683, d = 326..341 \setminus 327,$
- $M = 691, d = 329..345 \setminus 330,$
- $M = 701, d = 333..350 \setminus 334,$
- $M = 709, d = 338..354 \setminus 339,$
- $M = 719, d = 344..359 \setminus 345,$
- $M = 727, d = 347..363 \setminus 348,$
- $M = 733, d = 352..366,$
- $M = 739, d = 354..369,$
- $M = 743, d = 356..371,$
- $M = 751, d = 358..375,$
- $M = 757, d = 363..378 \setminus 364,$
- $M = 761, d = 365..380,$
- $M = 769, d = 367..384 \setminus 369,$
- $M = 773, d = 369..386 \setminus 371,$
- $M = 787, d = 379..393,$
- $M = 797, d = 381..398,$
- $M = 809, d = 390..404,$
- $M = 811, d = 391..405,$
- $M = 821, d = 393..410,$
- $M = 823, d = 394..411,$
- $M = 827, d = 398..413,$
- $M = 829, d = 399..414,$
- $M = 839, d = 403..419 \setminus 404,$
- $M = 853, d = 409..426 \setminus 410,$
- $M = 857, d = 414..428,$
- $M = 859, d = 415..429,$
- $M = 863, d = 415..431 \setminus 416,$
- $M = 877, d = 420..438 \setminus 422,$
- $M = 881, d = 423..440 \setminus 424,$
- $M = 883, d = 424..441 \setminus 425,$
- $M = 887, d = 428..443,$
- $M = 907, d = 435..453 \setminus 439,$
- $M = 911, d = 437..455 \setminus 439,$
- $M = 919, d = 443..459 \setminus 444,$
- $M = 929, d = 448..464,$
- $M = 937, d = 450..468 \setminus 452,$
- $M = 941, d = 452..470 \setminus 454,$
- $M = 947, d = 458..473,$
- $M = 953, d = 460..476 \setminus 461,$
- $M = 967, d = 466..483,$
- $M = 971, d = 467..485 \setminus 469,$
- $M = 977, d = 469..488 \setminus 472,$
- $M = 983, d = 479..491,$

- $M = 991, d = 479..495,$
- $M = 997, d = 482..498,$
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208, 6464, 6592, 6848, 6976, 7232, 8128, 8192$ bitowe,
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 6..13,$
- wag Hamminga dla bloku $M = 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024$ bitowego,
- wypełniania drzewa binarnego o wysokości $h = 3..10$ na podstawie ciągu bitów
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień – oprócz $h = 10$.

3. Procedury testowe przydatne do badania generatorów przy małej liczbie powtórzeń

3.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32,$
- autokorelacji dla ciągu bloków

○ $d = 2, t = 1..249999,$	○ $d = 13, t = 1..38461,$
○ $d = 3, t = 1..166666,$	○ $d = 14, t = 1..35713,$
○ $d = 4, t = 1..124999,$	○ $d = 15, t = 1..33332,$
○ $d = 5, t = 1..99999,$	○ $d = 16, t = 1..31249,$
○ $d = 6, t = 1..83332,$	○ $d = 17, t = 1..29411,$
○ $d = 7, t = 1..71428,$	○ $d = 18, t = 1..27777,$
○ $d = 8, t = 1..62499,$	○ $d = 19, t = 1..26315,$
○ $d = 9, t = 1..55555,$	○ $d = 20, t = 1..24999,$
○ $d = 10, t = 1..49999,$	○ $d = 21, t = 1..23809,$
○ $d = 11, t = 1..45454,$	○ $d = 22, t = 1..22726,$
○ $d = 12, t = 1..41666,$	○ $d = 23, t = 1..21738,$

-
- $d = 24, t = 1..20832,$
 - $d = 25, t = 1..19999,$
 - $d = 26, t = 1..19230,$
 - $d = 27, t = 1..18518,$
 - $d = 28, t = 1..17856,$
 - $d = 29, t = 1..17240,$
 - $d = 30, t = 1..16666,$
 - $d = 31, t = 1..16128,$
 - $d = 32, t = 1..15624,$
 - Corona – $d = 3..11,$
 - Cramera-von Misesa – $d = 16..32,$
 - częstości dla podprzedziału
 - $d = 16..22, [1/128; 1/4 + 1/128), M = 32, 256,$
 - $d = 16..22, [1/64; 1/2 + 1/128), M = 32, 256,$
 - $d = 16..22, [1/2 - 1/128; 3/4), M = 32, 256,$
 - $d = 23..32, [1/128; 1/4 + 1/128), M = 32,$
 - $d = 23..32, [1/64; 1/2 + 1/128), M = 32,$
 - $d = 23..32, [1/2 - 1/128; 3/4), M = 32,$
 - gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka ośmiościenna – zgodności dla liczby zwycięstw,
 - kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka sześcienna – zgodności dla liczby zwycięstw,
 - kolekcjonera – $d = 2..8,$
 - Kołmogorowa – $d = 16..32,$
 - Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka $K^+,$
 - statystyka $K^-,$
 - Kuipera – $d = 16..32,$
 - największej liczby elementów w przedziale – podciągi 1 kiB
 - $D = 1/2, d = 4..7, 9, 10, 12..15,$
 - $D = 0,4, d = 4..7, 9, 10, 12, 13,$
 - $D = 0,3, d = 4..7, 9, 10,$
 - $D = 1/4, d = 4..7,$
 - $D = 0,2, d = 4..6,$
 - $D = 1/8, d = 4..5,$
 - $D = 1/16, d = 4,$

- odstępów dni urodzin
 - $d = 22, n = 256,$
 - $d = 23, n = 512,$
- pochodnej ciągu bloków – $d = 2..32$
 - pierwsza pochodna,
 - druga pochodna,
- porządku książek

○ $d = 4, k = 2..3,$ ○ $d = 5, k = 2..4,$ ○ $d = 6, k = 2..5,$ ○ $d = 7, k = 1..6,$ ○ $d = 8, k = 2..7,$ ○ $d = 9, k = 1..8,$ ○ $d = 10, k = 2..9,$ ○ $d = 11, k = 1..10,$ ○ $d = 12, k = 1..11,$	○ $d = 13, k = 1..12,$ ○ $d = 14..16, k = 2..13,$ ○ $d = 17, k = 1..13,$ ○ $d = 18, k = 1..11,$ ○ $d = 19, k = 1..10,$ ○ $d = 20..22, k = 2..10,$ ○ $d = 23..25, k = 2..9,$ ○ $d = 26, k = 1..9,$
--	--
- serii malejących z pomijaniem w przód – $d = 6..32,$
- serii malejących z pomijaniem w tył – $d = 6..32,$
- serii rosnących z pomijaniem w przód – $d = 6..32,$
- serii rosnących z pomijaniem w tył – $d = 6..32,$
- stosu książek

○ $d = 4, k = 2..3,$ ○ $d = 5, k = 2..4,$ ○ $d = 6, k = 2..5,$ ○ $d = 7, k = 1..6,$ ○ $d = 8, k = 2..7,$ ○ $d = 9, k = 1..8,$ ○ $d = 10, k = 2..7,$ ○ $d = 11, k = 1..8,$ ○ $d = 12, k = 1..7,$ ○ $d = 13, k = 1..6,$ ○ $d = 14, k = 2..8,$	○ $d = 15, k = 2..6,$ ○ $d = 16, k = 2..10,$ ○ $d = 17, k = 1..6,$ ○ $d = 18, k = 1..8,$ ○ $d = 19, k = 1..7,$ ○ $d = 20, k = 2..6,$ ○ $d = 21, k = 1..9,$ ○ $d = 22, k = 2..7,$ ○ $d = 23..25, k = 2..6,$ ○ $d = 26, k = 1..4,$
--	---
- sum do G
 - $d = 16..32, G = 16, 32, 64, 128, 256,$
 - $d = 16..18, G = 512,$

- sumy kwadratów odstępów bez zapętlenia

- | | |
|---|--|
| ○ $m = 1, d = 20..32,$ | ○ $m = 59, 61, d = 25..31,$ |
| ○ $m = 2, 3, d = 19..32,$ | ○ $m = 64, 67, 71, d = 31,$ |
| ○ $m = 4, 5, 7, 8, 11, 13, 29, 31,$
32, 37, 41, $d = 18..32$ | ○ $m = 919, 929, 937, 941, 947,$
953, 967, 971, 977, 983, 991,
997, 1009, 1013, 1019, 1021,
1024, $d = 26,$ |
| ○ $m = 16, 17, 19, 23, d = 17..32,$ | ○ $m = 8192, 15625, k = 17..32,$ |
| ○ $m = 43, 47, d = 22..32,$ | |
| ○ $m = 53, d = 24..31,$ | |

- sumy kwadratów odstępów z zapętleniem

- | | |
|--|--|
| ○ $m = 1, d = 21..30,$ | ○ $m = 47, d = 20..32,$ |
| ○ $m = 2, 3, d = 19..32,$ | ○ $m = 53, 59, 61, 64, 67, 71,$
$d = 31..32,$ |
| ○ $m = 4, 5, 7, 8, 11, 13, 32, 37,$
41, 43, $d = 18..32,$ | ○ $m = 73, 79, 83, 89, 97, 101,$
103, 107, 109, $d = 31,$ |
| ○ $m = 16, 17, 19, 23, 29, 31,$
$d = 17..32,$ | |

- sumy logarytmów odstępów bez zapętlenia

- $m = 2, d = 24..32,$
- $m = 3, 5, 7, d = 20..32,$
- $m = 4, d = 21..32,$
- $m = 8, 11, 13, 16, 17, d = 19..32,$

- sumy logarytmów odstępów z zapętleniem

- $m = 2, d = 24..32,$
- $m = 3, 5, 7, 8, d = 20..32,$
- $m = 4, d = 21..32,$
- $m = 11, 13, 16, 17, 19, d = 19..32,$

- ściskania – $d = 10..20,$

- uniwersalny Maurera – $d = 1..10,$

- wag Hamminga w ciągu bloków

- | | |
|-----------------------------|----------------------------|
| ○ $d = 2, t = 2..5, k = 3,$ | ○ $d = 5$ |
| ○ $d = 3$ | ▪ $t = 2..3, k = 2, 4, 6,$ |
| ▪ $t = 2..4, k = 2, 4,$ | ▪ $t = 4, k = 2, 4,$ |
| ▪ $t = 5..8, k = 2,$ | ▪ $t = 5..8, k = 2,$ |
| ○ $d = 4$ | ○ $d = 6$ |
| ▪ $t = 2..3, k = 3, 5,$ | ▪ $t = 2, k = 3, 5, 7,$ |
| ▪ $t = 4..5, k = 3,$ | ▪ $t = 3, k = 3, 5,$ |
| | ▪ $t = 4..5, k = 3,$ |

- $d = 7$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 8$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 9$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 10$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 11$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 12$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 13$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 14$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 15$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 16$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 20$
 - $t = 2, k = 3, 5, 7, 9, 11,$
13,
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 32$
 - $t = 2, k = 3, 5, 7, 9, 11,$
13,
 - $t = 3, k = 3,$
- wypełniania drzewa binarnego o wysokości $h = 3..11$ na podstawie ciągu bloków $d = 4..32$ bitowych
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

3.2. Procedury dla ciągów wektorów nienakładających się

- iloczynów
 - $t = 2..5, d = 12..32,$
 - $t = 6..16, d = 13..32,$

- kolizji

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 15..17,$ | ○ $t = 5, d = 3,$ |
| ○ $t = 2, d = 8,$ | ○ $t = 8, d = 2,$ |
| ○ $t = 3, d = 5,$ | ○ $t = 15..17, d = 1,$ |
| ○ $t = 4, d = 4,$ | |

- kolizji indeksów maksymalnych elementów

- $t = 2, d = 13..21,$
- $t = 3..16, d = 11..22,$

- liczby komórek z co najmniej $b = 1..2$ punktami dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 15..17,$ | ○ $t = 5, d = 3,$ |
| ○ $t = 2, d = 8,$ | ○ $t = 8, d = 2,$ |
| ○ $t = 3, d = 5,$ | ○ $t = 15..17, d = 1,$ |
| ○ $t = 4, d = 4,$ | |

liczby komórek z co najmniej $b = 3$ punktami dla wektorów:

- | | |
|--------------------|--------------------|
| ○ $t = 1, d = 16,$ | ○ $t = 8, d = 2,$ |
| ○ $t = 2, d = 8,$ | ○ $t = 16, d = 1,$ |
| ○ $t = 4, d = 4,$ | |

liczby komórek z co najmniej $b = 4$ punktami dla wektorów:

- | | |
|--------------------|--------------------|
| ○ $t = 1, d = 15,$ | ○ $t = 5, d = 3,$ |
| ○ $t = 3, d = 5,$ | ○ $t = 15, d = 1,$ |

- liczby komórek z dokładnie $b = 0$ punktami dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 15..17,$ | ○ $t = 5, d = 3,$ |
| ○ $t = 2, d = 8,$ | ○ $t = 8, d = 2,$ |
| ○ $t = 3, d = 5,$ | ○ $t = 15..17, d = 1,$ |
| ○ $t = 4, d = 4,$ | |

liczby komórek z dokładnie $b = 1$ punktem dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 15..20,$ | ○ $t = 5, d = 3..4,$ |
| ○ $t = 2, d = 8..10,$ | ○ $t = 6, d = 3,$ |
| ○ $t = 3, d = 5..6,$ | ○ $t = 8..10, d = 2,$ |
| ○ $t = 4, d = 4..5,$ | ○ $t = 15..20, d = 1,$ |

liczby komórek z dokładnie $b = 2$ punktami dla wektorów:

- | | |
|------------------------|------------------------|
| ○ $t = 1, d = 15..16,$ | ○ $t = 5, d = 3,$ |
| ○ $t = 2, d = 8,$ | ○ $t = 8, d = 2,$ |
| ○ $t = 3, d = 5,$ | ○ $t = 15..16, d = 1,$ |
| ○ $t = 4, d = 4,$ | |

liczby komórek z dokładnie $b = 3$ punktami dla wektorów:

- $t = 1, d = 15,$
- $t = 3, d = 5,$
- $t = 5, d = 3,$
- $t = 15, d = 1,$

liczby komórek z dokładnie $b = 5$ punktami dla wektorów:

- $t = 1, d = 14,$
- $t = 2, d = 7,$
- $t = 7, d = 2,$
- $t = 14, d = 1,$

• logarytm prawdopodobieństwa

- $t = 1, d = 2..21,$
- $t = 2, d = 1..10,$
- $t = 3, d = 1..7,$
- $t = 4, d = 1..5,$
- $t = 5, d = 1..4,$
- $t = 6..7, d = 1..3,$
- $t = 8..10, d = 1..2,$
- $t = 11..21, d = 1,$

• m najbliższych par

Tabela B.4. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka euklidesowa

$d \setminus t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16			62	55	51	47	44	41	39	37	36	34	33	32	31
17		70	60	54	49	45	42	40	38	36	35	33	32	31	30
18		68	58	52	48	44	41	39	37	35	34	32	31	30	29
19		66	57	51	46	43	40	38	36	34	33	31	30	29	28
20		64	55	50	45	42	39	37	35	33	32	31	29	28	27
21		62	54	48	44	41	38	36	34	32	31	30	29	28	27
22	75	61	53	47	43	40	37	35	33	32	30	29	28	27	26
23	73	60	52	46	42	39	36	34	32	31	30	28	27	26	26
24	72	58	51	45	41	38	36	34	32	30	29	28	27	26	25
25	70	57	50	44	40	37	35	33	31	30	28	27	26	25	25
26	69	56	49	43	40	37	34	32	31	29	28	27	26	25	24
27	68	55	48	43	39	36	34	32	30	29	27	26	25	24	24
28	66	54	47	42	38	35	33	31	29	28	27	26	25	24	23
29	65	53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24	23	22	22

Tabela B.5. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka nieskończoność

$d \setminus t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
20												31	29	28	27
21						41	38	36	34	32	31	30	29	28	27
22				47	43	40	37	35	33	32	30	29	28	27	26
23			52	46	42	39	36	34	32	31	30	28	27	26	26
24			51	45	41	38	36	34	32	30	29	28	27	26	25
25		57	50	44	40	37	35	33	31	30	28	27	26	25	25
26		56	49	43	40	37	34	32	31	29	28	27	26	25	24
27		55	48	43	39	36	34	32	30	29	27	26	25	24	24
28		54	47	42	38	35	33	31	29	28	27	26	25	24	23
29	65	53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24			

- maksimum z t
 - $t = 2, d = 15..32,$
 - $t = 3..4, d = 16..32,$
 - $t = 5..15, d = 17..32,$
 - $t = 16, d = 18..32,$
- momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 1..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$
 - $t = 11..15, d = 1..2,$
 - $t = 16..30, d = 1,$
- permutacji
 - $t = 2, d = 3..19,$
 - $t = 3..10, d = 2..32,$
- podziałów
 - $t = 2, d = 3..4,$
 - $t = 3, d = 2..5,$
 - $t = 4, d = 2..6,$
 - $t = 5, d = 2..7,$
 - $t = 6..9, d = 2..8,$
 - $t = 10..13, d = 2..9,$
 - $t = 14, d = 3..9,$
 - $t = 15..16, d = 3..10,$
- pokerowy
 - $t = 2, d = 3..4,$
 - $t = 3, d = 2..5,$
 - $t = 4, d = 2..6,$
 - $t = 5, d = 2..7,$
 - $t = 6..9, d = 2..8,$
 - $t = 10..14, d = 2..9,$
 - $t = 15..16, d = 2..10,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 1, d = 2..20,$
 - $t = 2, d = 1..10,$
 - $t = 3, d = 1..6,$
 - $t = 4, d = 1..5,$
 - $t = 5, d = 1..4,$
 - $t = 6, d = 1..3,$
 - $t = 7..10, d = 1..2,$
 - $t = 11..20, d = 1,$
- potęgowo-dywergencyjny – $\delta = 2$
 - $t = 1, d = 4..20,$
 - $t = 2, d = 2..10,$
 - $t = 3, d = 2..6,$
 - $t = 4, d = 1..5,$

○ $t = 5, d = 1..4,$

○ $t = 6, d = 1..3,$

potęgowo-dywergencyjny – $\delta = 3$

○ $t = 1, d = 5..21,$

○ $t = 2, d = 3..10,$

○ $t = 3, d = 2..7,$

○ $t = 4, d = 2..5,$

potęgowo-dywergencyjny – $\delta = 4$

○ $t = 1, d = 5..13,$

○ $t = 2, d = 3..6,$

○ $t = 3, d = 2..4,$

potęgowo-dywergencyjny – $\delta = 5$

○ $t = 1, d = 5..12,$

○ $t = 2, d = 3..6,$

○ $t = 3, d = 2..4,$

potęgowo-dywergencyjny – $\delta = 6$

○ $t = 1, d = 5..11,$

○ $t = 2, d = 3..5,$

○ $t = 3, d = 2..3,$

potęgowo-dywergencyjny – $\delta = 7$

○ $t = 1, d = 5..9,$

○ $t = 2, d = 3..4,$

○ $t = 3, d = 2..3,$

- zgodności Pearsona

○ $t = 1, d = 2..30,$

○ $t = 2, d = 1..15,$

○ $t = 3, d = 1..10,$

○ $t = 4, d = 1..7,$

○ $t = 5, d = 1..6,$

○ $t = 7..10, d = 1..2,$

○ $t = 11..20, d = 1,$

○ $t = 5, d = 1..4,$

○ $t = 6..7, d = 1..3,$

○ $t = 8..10, d = 1..2,$

○ $t = 11..21, d = 1,$

○ $t = 4, d = 2..3,$

○ $t = 5..6, d = 1..2,$

○ $t = 7..13, d = 1,$

○ $t = 4, d = 2..3,$

○ $t = 5..6, d = 1..2,$

○ $t = 7..12, d = 1,$

○ $t = 4, d = 2,$

○ $t = 5, d = 1..2,$

○ $t = 6..11, d = 1,$

○ $t = 4, d = 2,$

○ $t = 5..9, d = 1,$

○ $t = 6, d = 1..5,$

○ $t = 7, d = 1..4,$

○ $t = 8..10, d = 1..3,$

○ $t = 11..15, d = 1..2,$

○ $t = 16..30, d = 1.$

3.3. Procedury dla ciągów wektorów nakładających się

- częstości dla nakładających się wektorów

○ $t = 2, d = 6..7,$

○ $t = 3, d = 4..5,$

○ $t = 5, d = 3,$

○ $t = 8, d = 2,$

- nakładających się permutacji – $t = 3..6, d = 2..32,$

- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 2, d = 6..7,$
 - $t = 3, d = 4..5,$
 - $t = 5, d = 3,$
 - $t = 8, d = 2,$
- wolnych cel Marsagli
 - $t = 2, d = 8,$
 - $t = 4, d = 4,$
 - $t = 6, d = 3,$
 - $t = 8..9, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego
 - $t = 2, d = 8,$
 - $t = 4, d = 4,$
 - $t = 6, d = 3,$
 - $t = 9, d = 2.$

3.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..500000$,
- całkowitej liczby serii,
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- częstości serii 0 i 1 łącznie,
- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997$ bitowego,
- liczby skoków w profilu liniowym,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146, 158, 166, 178, 194, 202, 206, 214, 218, 226, 254, 256, 262, 274, 278, 298, 302, 314, 326, 334, 346, 358, 362, 382, 386, 394, 398, 422, 446, 454, 458, 466, 478, 482, 502, 512, 514, 526, 538, 542, 554, 562, 566, 586, 614, 622, 626, 634, 662, 674, 694, 698, 706, 718, 734, 746, 758, 766, 778, 794, 802, 818, 838, 842, 862, 866, 878, 886, 898, 914, 922, 926, 934, 958, 974, 982, 998, 1006, 1018, 1024, 1042, 1046, 1082, 1094, 1114, 1126, 1138, 1142, 1154, 1174, 1186, 1198, 1202, 1214, 1226, 1234, 1238, 1262, 1282, 1286, 1294, 1306, 1318, 1322, 1346, 1354, 1366, 1382, 1402, 1418, 1438, 1454, 1466, 1478, 1486, 1502, 1514, 1522, 1538, 1546, 1574, 1594, 1618, 1622, 1642, 1646, 1654, 1658, 1678, 1706, 1714, 1718, 1726, 1754, 1762, 1766, 1774, 1814, 1822, 1838, 1858, 1874, 1882, 1894, 1906, 1934, 1942, 1954, 1966, 1982, 1994, 2018, 2026, 2038, 2042, 2048$

- liczby powrotów do 0,
- liczby przejść przez 0 ze zmianą strony,
- łącznego „czasu” odchylenia w prawo,
- maksymalnego odchylenia ścieżki,

$M = 16384$

- liczby powrotów do 0,
- liczby przejść przez 0 ze zmianą strony,
- maksymalnego odchylenia ścieżki,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192, 10000$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192, 10000$ bitowym,
- nakładających się t bitowych wzorców 0 oraz 1 w bloku M bitowym
 - $t = 4, M = 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 5, M = 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 6, M = 64, 128, 256, 512, 1000, 1024,$
 - $t = 7, M = 128, 256, 512, 1000, 1024,$
 - $t = 8, M = 256, 512, 1000, 1024,$
 - $t = 9, M = 512, 1000, 1024,$
 - $t = 10, M = 1000, 1024,$
- nienakładających się d bitowych wzorców w bloku M bitowym (wzorce w zapisie heksadecymalnym, ograniczone do początkowych 21):
 - $d = 2, M = 524288$, wzorce: 1, 2,
 - $d = 3, M = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 524288$, wzorce: 1, 3, 4, 6,
 - $d = 4, M = 524288$, wzorce: 1, 3, 7, 8, C, E,
 - $d = 5, M = 524288$, wzorce: 1, 3, 5, 7, B, F, 10, 14, 18, 1A, 1C, 1E,
 - $d = 11..12, M = 262144$, wzorce: 1, 3, 5, 7, 9, B, D, F, 11, 13, 15, 17, 19, 1B, 1D, 1F, 21, 23, 25, 27, 29,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7$ bitowego,
- niezależności wag Hamminga – wariant 2

○ $M = 4, d = 1,$ ○ $M = 5, d = 1..3,$ ○ $M = 7, d = 2..4,$ ○ $M = 8, d = 2..4,$ ○ $M = 11, d = 3..6,$	○ $M = 13, d = 4..7,$ ○ $M = 16, d = 5..8,$ ○ $M = 17, d = 6..9,$ ○ $M = 19, d = 6..10,$ ○ $M = 23, d = 8..12,$
---	--

-
- $M = 29, d = 11..15,$
 - $M = 31, d = 12..16,$
 - $M = 32, d = 12..16,$
 - $M = 37, d = 15..19,$
 - $M = 41, d = 16..21,$
 - $M = 43, d = 17..22,$
 - $M = 47, d = 19..24,$
 - $M = 53, d = 22..27,$
 - $M = 59, d = 25..30,$
 - $M = 61, d = 25..31,$
 - $M = 64, d = 27..32,$
 - $M = 67, d = 29..34,$
 - $M = 71, d = 30..36,$
 - $M = 73, d = 31..37,$
 - $M = 79, d = 34..40,$
 - $M = 83, d = 36..42,$
 - $M = 89, d = 39..45,$
 - $M = 97, d = 42..49,$
 - $M = 101, d = 44..51,$
 - $M = 103, d = 45..51,$
 - $M = 107, d = 47..54,$
 - $M = 109, d = 48..54,$
 - $M = 113, d = 50..56,$
 - $M = 127, d = 57..63,$
 - $M = 128, d = 58..64,$
 - $M = 131, d = 59..66,$
 - $M = 137, d = 62..69,$
 - $M = 139, d = 62..70,$
 - $M = 149, d = 67..74,$
 - $M = 151, d = 69..75,$
 - $M = 157, d = 72..78,$
 - $M = 163, d = 75..81,$
 - $M = 167, d = 77..83,$
 - $M = 173, d = 79..86,$
 - $M = 179, d = 83..89,$
 - $M = 181, d = 83..90,$
 - $M = 191, d = 87..95,$
 - $M = 193, d = 90..96,$
 - $M = 197, d = 90..98,$
 - $M = 199, d = 92..99,$
 - $M = 211, d = 97..105,$
 - $M = 223, d = 102..111,$
 - $M = 227, d = 105..113,$
 - $M = 229, d = 106..114,$
 - $M = 233, d = 108..116,$
 - $M = 239, d = 111..119,$
 - $M = 241, d = 112..120,$
 - $M = 251, d = 117..125,$
 - $M = 256, d = 120..128,$
 - $M = 257, d = 123..128,$
 - $M = 263, d = 123..131,$
 - $M = 269, d = 126..134,$
 - $M = 271, d = 127..135,$
 - $M = 277, d = 131..138,$
 - $M = 281, d = 131..140,$
 - $M = 283, d = 134..141,$
 - $M = 293, d = 137..146,$
 - $M = 307, d = 147..153,$
 - $M = 311, d = 148..155,$
 - $M = 313, d = 149..156,$
 - $M = 317, d = 150..158,$
 - $M = 331, d = 159..165,$
 - $M = 337, d = 162..168,$
 - $M = 347, d = 167..173,$
 - $M = 349, d = 167..174,$
 - $M = 353, d = 170..176,$
 - $M = 359, d = 174..179,$
 - $M = 367, d = 178..183,$
 - $M = 373, d = 181..186,$
 - $M = 379, d = 183..189,$
 - $M = 383, d = 184..191,$
 - $M = 389, d = 187..194,$

- $M = 397, d = 192..198,$
- $M = 401, d = 191..200,$
- $M = 409, d = 199..204,$
- $M = 419, d = 203..209,$
- $M = 421, d = 205..210,$
- $M = 431, d = 211..215,$
- $M = 433, d = 211..216,$
- $M = 439, d = 210..219,$
- $M = 443, d = 211..221,$
- $M = 449, d = 216..224,$
- $M = 457, d = 220..228,$
- $M = 461, d = 220..230,$
- $M = 463, d = 222..231,$
- $M = 467, d = 231..233,$
- $M = 479, d = 235..239,$
- $M = 487, d = 240..243,$
- $M = 491, d = 241..245,$
- $M = 499, d = 239..249,$
- $M = 503, d = 243..251,$
- $M = 509, d = 244..254,$
- $M = 512, d = 248..256,$
- $M = 521, d = 253..260,$
- $M = 523, d = 252..261,$
- $M = 541, d = 264..270,$
- $M = 547, d = 271..273,$
- $M = 557, d = 269..278,$
- $M = 563, d = 273..281,$
- $M = 569, d = 275..284,$
- $M = 571, d = 276..285,$
- $M = 577, d = 278..288,$
- $M = 587, d = 285..293,$
- $M = 593, d = 287..296,$
- $M = 599, d = 291..299,$
- $M = 601, d = 294..300,$
- $M = 607, d = 296..303,$
- $M = 613, d = 301..306,$
- $M = 617, d = 302..308,$
- $M = 619, d = 302..309,$
- $M = 631, d = 310..315,$
- $M = 641, d = 316..320,$
- $M = 643, d = 314..321,$
- $M = 647, d = 318..323,$
- $M = 653, d = 317..326,$
- $M = 659, d = 323..329,$
- $M = 661, d = 321..330,$
- $M = 673, d = 328..336,$
- $M = 677, d = 327..338,$
- $M = 683, d = 333..341,$
- $M = 691, d = 337..345,$
- $M = 701, d = 339..350,$
- $M = 709, d = 346..354,$
- $M = 719, d = 354..359,$
- $M = 727, d = 356..363,$
- $M = 733, d = 359..366,$
- $M = 739, d = 363..369,$
- $M = 743, d = 362..371,$
- $M = 751, d = 368..375,$
- $M = 757, d = 370..378,$
- $M = 761, d = 374..380,$
- $M = 769, d = 377..384,$
- $M = 773, d = 377..386,$
- $M = 787, d = 388..393,$
- $M = 797, d = 389..398,$
- $M = 809, d = 396..404,$
- $M = 811, d = 399..405,$
- $M = 821, d = 402..410,$
- $M = 823, d = 403..411,$
- $M = 827, d = 405..413,$
- $M = 829, d = 406..414,$
- $M = 839, d = 410..419,$
- $M = 853, d = 419..426,$
- $M = 857, d = 422..428,$

- $M = 859, d = 423..428,$
- $M = 863, d = 423..431,$
- $M = 877, d = 431..438,$
- $M = 881, d = 437..440,$
- $M = 883, d = 436..441,$
- $M = 887, d = 439..443,$
- $M = 907, d = 447..453,$
- $M = 911, d = 451..455,$
- $M = 919, d = 452..459,$
- $M = 929, d = 456..464,$
- $M = 937, d = 461..468,$
- $M = 941, d = 462..470,$
- $M = 947, d = 466..473,$
- $M = 953, d = 470..475,$
- $M = 967, d = 475..482,$
- $M = 971, d = 480..485,$
- $M = 977, d = 482..488,$
- $M = 983, d = 483..491,$
- $M = 991, d = 490..495,$
- $M = 997, d = 491..498,$
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008, 3392, 3776, 3904, 4096, 4288, 4544, 4672, 5056, 5312, 5696, 6208$ bitowe,
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 6..13,$
- wag Hamminga dla bloku $M = 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283$ bitowego,
- wypełniania drzewa binarnego o wysokości $h = 3..10$ na podstawie ciągu bitów
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień – oprócz $h = 10$.

4. Procedury testowe przydatne do badania generatorów przy dużej liczbie powtórzeń

4.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32,$
- autokorelacji dla ciągu bloków
 - $d = 2, t = 1..249999,$
 - $d = 3, t = 1..166666,$
 - $d = 4, t = 1..124999,$
 - $d = 5, t = 1..99999,$
 - $d = 6, t = 1..83332,$
 - $d = 7, t = 1..71428,$
 - $d = 8, t = 1..62499,$
 - $d = 9, t = 1..55555,$
 - $d = 10, t = 1..49999,$
 - $d = 11, t = 1..45454,$
 - $d = 12, t = 1..41666,$
 - $d = 13, t = 1..38461,$
 - $d = 14, t = 1..35713,$
 - $d = 15, t = 1..33332,$
 - $d = 16, t = 1..31249,$
 - $d = 17, t = 1..29411,$
 - $d = 18, t = 1..27777,$
 - $d = 19, t = 1..26315,$

- $d = 20, t = 1..24999,$
- $d = 21, t = 1..23809,$
- $d = 22, t = 1..22726,$
- $d = 23, t = 1..21738,$
- $d = 24, t = 1..20832,$
- $d = 25, t = 1..19999,$
- $d = 26, t = 1..19230,$
- $d = 27, t = 1..18518,$
- $d = 28, t = 1..17856,$
- $d = 29, t = 1..17240,$
- $d = 30, t = 1..16666,$
- $d = 31, t = 1..16128,$
- $d = 32, t = 1..15624,$
- Corona – $d = 3..10,$
- Cramera-von Misesa – $d = 16..32,$
- częstości dla podprzedziału
 - $d = 16..32, [1/128; 1/4 + 1/128), M = 32,$
 - $d = 16..32, [1/64; 1/2 + 1/128), M = 32,$
 - $d = 16..32, [1/2 - 1/128; 3/4), M = 32,$
- gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka ośmiościenna – zgodności dla liczby zwycięstw,
 - kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
 - kostka sześcienna – zgodności dla liczby zwycięstw,
- kolekcjonera – $d = 2..6,$
- Kołmogorowa – $d = 16..32,$
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- Kuipera – $d = 16..32,$
- pochodnej ciągu bloków
 - pierwsza pochodna – $d = 2..8,$
 - druga pochodna – $d = 3..8,$
- porządku książek
 - $d = 4, k = 2..3,$
 - $d = 5, k = 2..4,$
 - $d = 6, k = 2..5,$
 - $d = 7, k = 2..6,$
 - $d = 8, k = 2..7,$
 - $d = 9, k = 2..8,$
 - $d = 10, k = 2..9,$
 - $d = 11, k = 2..10,$
 - $d = 12, k = 2..11,$
 - $d = 13, k = 2..12,$
 - $d = 14..17, k = 2..13,$
 - $d = 18, k = 2..11,$
 - $d = 19..22, k = 2..10,$
 - $d = 23..26, k = 2..9,$

- serii malejących z pomijaniem w przód – $d = 7..32$,
- serii malejących z pomijaniem w tył – $d = 7..32$,
- serii rosnących z pomijaniem w przód – $d = 7..32$,
- serii rosnących z pomijaniem w tył – $d = 7..32$,
- stosu książek
 - $d = 4, k = 2..3$,
 - $d = 5, k = 2..4$,
 - $d = 6, k = 2..5$,
 - $d = 7..8, k = 2..6$,
 - $d = 9, k = 2..8$,
 - $d = 10, k = 2..7$,
 - $d = 11, k = 2..8$,
 - $d = 12, k = 2..7$,
 - $d = 13, k = 2..5$,
 - $d = 14, k = 2..8$,
 - $d = 15, k = 2..6$,
 - $d = 16, k = 2..7$,
 - $d = 17, k = 2..6$,
 - $d = 18, k = 2..8$,
 - $d = 19, k = 2..7$,
 - $d = 20, k = 2..6$,
 - $d = 21, k = 2..9$,
 - $d = 22, k = 2..7$,
 - $d = 23..25, k = 2..6$,
 - $d = 26, k = 2..4$,
- sum do G
 - $d = 16..32, G = 16, 32, 64$,
 - $d = 16..18, G = 128$,
- sumy kwadratów odstępów bez zapętlania
 - $m = 1, d = 20..32$,
 - $m = 2, 3, 4, 5, d = 19..32$,
 - $m = 7, 8, 11, 13, 16, 17, 29, 31, 32, 37, 41, d = 18..32$,
 - $m = 19, 23, d = 17..32$,
 - $m = 43, 47, d = 22..32$,
 - $m = 53, d = 24..31$,
 - $m = 59, 61, d = 25..31$,
 - $m = 64, 67, 71, d = 31$,
 - $m = 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997, 1009, 1013, 1019, 1021, 1024, d = 26$,
 - $m = 8192, 15625, k = 17..32$,
- sumy kwadratów odstępów z zapętlaniem
 - $m = 2, 3, d = 19..32$,
 - $m = 4, 5, 7, 8, 11, 13, 32, 37, 41, 43, d = 18..32$,
 - $m = 16, 17, 19, 23, 29, 31, d = 17..32$,
 - $m = 47, d = 20..32$,
 - $m = 53, 59, 61, 64, 67, 71, d = 31..32$,
 - $m = 73, 79, 83, 89, 97, 101, 103, 107, 109, d = 31$,
- sumy logarytmów odstępów bez zapętlania
 - $m = 2, d = 25..32$,
 - $m = 3, d = 22..32$,

- sumy logarytmów odstępów z zapętlaniem
 - $m = 2, d = 25..32,$
 - $m = 3, d = 23..32,$
- ściskania – $d = 11..20,$
- uniwersalny Maurera – $d = 1..10,$
- wag Hamminga w ciągu bloków
 - $d = 2, t = 2..5, k = 3,$
 - $d = 3$
 - $t = 2..4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 4$
 - $t = 2..3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 5$
 - $t = 2, k = 2, 4, 6,$
 - $t = 3..4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 6$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 7$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 8$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 9$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 10$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 11$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 12$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 13$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 14$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 15$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 16$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$

- $t = 4..5, k = 3,$
- $d = 20$
- $t = 2, k = 3, 5, 7, 9, 11,$
- $t = 3, k = 3, 5,$
- $t = 4..5, k = 3,$
- $d = 32$
- $t = 2, k = 3, 5, 7, 9, 11, 13,$
- $t = 3, k = 3,$
- wypełniania drzewa binarnego o wysokości $h = 3..11$ na podstawie ciągu bloków $d = 4..32$ bitowych
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

4.2. Procedury dla ciągów wektorów nienakładających się

- iloczynów
 - $t = 2..3, d = 16..32,$
 - $t = 4..6, d = 17..32,$
 - $t = 7..16, d = 18..32,$
- kolizji indeksów maksymalnych elementów
 - $t = 3..7, d = 11..22,$
 - $t = 8..16, d = 12..22,$
- logarytm prawdopodobieństwa
 - $t = 1, d = 2..20,$
 - $t = 2, d = 1..10,$
 - $t = 3, d = 1..6,$
 - $t = 4, d = 1..5,$
 - $t = 5, d = 1..4,$
 - $t = 6, d = 1..3,$
 - $t = 7..10, d = 1..2,$
 - $t = 11..20, d = 1,$
- m najbliższych par

Tabela B.6. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka euklidesowa

$d \setminus t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16			62	55	51	47	44	41	39	37	36	34	33	32	31
17			60	54	49	45	42	40	38	36	35	33	32	31	30
18		68	58	52	48	44	41	39	37	35	34	32	31	30	29
19		66	57	51	46	43	40	38	36	34	33	31	30	29	28
20		64	55	50	45	42	39	37	35	33	32	31	29	28	27
21		62	54	48	44	41	38	36	34	32	31	30	29	28	27
22		61	53	47	43	40	37	35	33	32	30	29	28	27	26
23	73	60	52	46	42	39	36	34	32	31	30	28	27	26	26
24	72	58	51	45	41	38	36	34	32	30	29	28	27	26	25
25	70	57	50	44	40	37	35	33	31	30	28	27	26	25	25
26	69	56	49	43	40	37	34	32	31	29	28	27	26	25	24
27	68	55	48	43	39	36	34	32	30	29	27	26	25	24	24
28	66	54	47	42	38	35	33	31	29	28	27	26	25	24	23
29	65	53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24	23	22	22

Tabela B.7. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka nieskończoność

$d \backslash t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
21															27
22								35	33	32	30	29	28	27	26
23					42	39	36	34	32	31	30	28	27	26	26
24				45	41	38	36	34	32	30	29	28	27	26	25
25			50	44	40	37	35	33	31	30	28	27	26	25	25
26		56	49	43	40	37	34	32	31	29	28	27	26	25	24
27		55	48	43	39	36	34	32	30	29	27	26	25	24	24
28		54	47	42	38	35	33	31	29	28	27	26	25	24	23
29		53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24			

- maksimum z t
 - $t = 2, d = 16..32,$
 - $t = 3..7, d = 17..32,$
 - $t = 8..16, d = 18..32,$
- momentu pierwszego rzędu – ze względów praktycznych wykonywany dla wektorów

○ $t = 1, d = 1..30,$ ○ $t = 2, d = 1..15,$ ○ $t = 3, d = 1..10,$ ○ $t = 4, d = 1..7,$ ○ $t = 5, d = 1..6,$	○ $t = 6, d = 1..5,$ ○ $t = 7, d = 1..4,$ ○ $t = 8..10, d = 1..3,$ ○ $t = 11..15, d = 1..2,$ ○ $t = 16..30, d = 1,$
--	--
- momentu drugiego rzędu – ze względów praktycznych wykonywany dla wektorów

○ $t = 1, d = 2..30,$ ○ $t = 2, d = 1..15,$ ○ $t = 3, d = 1..10,$ ○ $t = 4, d = 1..7,$ ○ $t = 5, d = 1..6,$	○ $t = 6, d = 1..5,$ ○ $t = 7, d = 1..4,$ ○ $t = 8..10, d = 1..3,$ ○ $t = 11..15, d = 1..2,$ ○ $t = 16..30, d = 1,$
--	--
- permutacji
 - $t = 2, d = 3..4,$
 - $t = 3..10, d = 2..32,$
- podziałów

○ $t = 3, d = 2..4,$ ○ $t = 4..6, d = 2..6,$ ○ $t = 7, d = 3..7,$	○ $t = 8, d = 2..7,$ ○ $t = 9..10, d = 2..8,$ ○ $t = 11..16, d = 3..8,$
--	--

- pokerowy
 - $t = 3, d = 2..4,$
 - $t = 4..5, d = 2..6,$
 - $t = 6..9, d = 2..7,$
 - $t = 10..12, d = 2..8,$
 - $t = 13..14, d = 2..3,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 1, d = 2..17,$
 - $t = 2, d = 1..8,$
 - $t = 3, d = 1..5,$
 - $t = 4, d = 1..4,$
 - $t = 5, d = 1..3,$
 - $t = 6..8, d = 1..2,$
 - $t = 9..17, d = 1,$
- potęgowo-dywergencyjny – $\delta = 2$
 - $t = 1, d = 5..13,$
 - $t = 2, d = 3..6,$
 - $t = 3, d = 2..4,$
 - $t = 4, d = 1..3,$
 - $t = 5..6, d = 1..2,$
 - $t = 7..13, d = 1,$
- potęgowo-dywergencyjny – $\delta = 3$
 - $t = 1, d = 5..12,$
 - $t = 2, d = 3..6,$
 - $t = 3, d = 2..4,$
 - $t = 4, d = 2..3,$
 - $t = 5..6, d = 1..2,$
 - $t = 7..12, d = 1,$
- potęgowo-dywergencyjny – $\delta = 4$
 - $t = 1, d = 6..10,$
 - $t = 2, d = 3..5,$
 - $t = 3, d = 2..3,$
 - $t = 4..5, d = 2,$
 - $t = 6..10, d = 1,$
- potęgowo-dywergencyjny – $\delta = 5$
 - $t = 1, d = 6..9,$
 - $t = 2, d = 3..4,$
 - $t = 3, d = 2..3,$
 - $t = 4, d = 2,$
 - $t = 6..9, d = 1,$
- potęgowo-dywergencyjny – $\delta = 6$
 - $t = 1, d = 6..8,$
 - $t = 2, d = 3..4,$
 - $t = 3..4, d = 2,$
 - $t = 6..8, d = 1,$
- potęgowo-dywergencyjny – $\delta = 7$
 - $t = 1, d = 6..8,$
 - $t = 2, d = 3..4,$
 - $t = 3..4, d = 2,$
 - $t = 6..8, d = 1,$
- zgodności Pearsona
 - $t = 1, d = 2..30,$
 - $t = 2, d = 1..15,$
 - $t = 3, d = 1..10,$
 - $t = 4, d = 1..7,$
 - $t = 5, d = 1..6,$
 - $t = 6, d = 1..5,$
 - $t = 7, d = 1..4,$
 - $t = 8..10, d = 1..3,$

$$\circ \quad t = 11..15, d = 1..2,$$

$$\circ \quad t = 16..30, d = 1.$$

4.3. Procedury dla ciągów wektorów nakładających się

- nakładających się permutacji – $t = 3..6, d = 2..32$,
- potęgowo-dywergencyjny – $\delta = 1$
 - $\circ \quad t = 2, d = 6..7,$
 - $\circ \quad t = 3, d = 4..5,$
 - $\circ \quad t = 5, d = 3,$
 - $\circ \quad t = 8, d = 2,$
- wolnych cel Marsagli
 - $\circ \quad t = 6, d = 3,$
 - $\circ \quad t = 9, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego
 - $\circ \quad t = 2, d = 8,$
 - $\circ \quad t = 4, d = 4,$
 - $\circ \quad t = 6, d = 3,$
 - $\circ \quad t = 9, d = 2.$

4.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..500000$,
- całkowitej liczby serii,
- całkowitej liczby serii z uwzględnieniem proporcji 0 i 1,
- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997$ bitowego,
- liczby skoków w profilu liniowym,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146, 158, 166, 178, 194, 202, 206, 214, 218, 226, 254, 256, 262, 274, 278, 298, 302, 314, 326, 334, 346, 358, 362, 382, 386, 394, 398, 422, 446, 454, 458, 466, 478, 482, 502, 512, 514, 526, 538, 542, 554, 562, 566, 586, 614, 622, 626, 634, 662, 674, 694, 698, 706, 718, 734, 746, 758, 766, 778, 794, 802, 818, 838, 842, 862, 866, 878, 886, 898, 914, 922, 926, 934, 958, 974, 982, 998, 1006, 1018, 1024, 1042, 1046, 1082, 1094, 1114, 1126, 1138, 1142, 1154, 1174, 1186, 1198, 1202, 1214, 1226, 1234, 1238, 1262, 1282, 1286, 1294, 1306, 1318, 1322, 1346, 1354, 1366, 1382, 1402, 1418, 1438, 1454, 1466, 1478, 1486, 1502, 1514, 1522, 1538, 1546, 1574, 1594, 1618, 1622, 1642, 1646, 1654, 1658, 1678, 1706, 1714, 1718, 1726, 1754, 1762, 1766, 1774, 1814, 1822, 1838, 1858, 1874, 1882, 1894, 1906, 1934, 1942, 1954, 1966, 1982, 1994, 2018, 2026, 2038, 2042, 2048$

- liczby powrotów do 0,
- liczby przejść przez 0 ze zmianą strony,
- łącznego „czasu” odchylenia w prawo,
- maksymalnego odchylenia ścieżki,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192$ bitowym,
- nakładających się t bitowych wzorców 0 oraz 1 w bloku M bitowym
 - $t = 4, M = 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 5, M = 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 6, M = 64, 128, 256, 512, 1000, 1024,$
 - $t = 7, M = 128, 256, 512, 1000, 1024,$
 - $t = 8, M = 256, 512, 1000, 1024,$
 - $t = 9, M = 512, 1000, 1024,$
 - $t = 10, M = 1000, 1024,$
- nienakładających się d bitowych wzorców w bloku M bitowym (wzorce w zapisie heksadecymalnym):
 - $d = 2, M = 524288,$ wzorce: 1, 2,
 - $d = 3, M = 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 524288,$ wzorce: 1, 3, 4, 6,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7$ bitowego,
- niezależności wag Hamminga – wariant 2

○ $M = 4, d = 1,$ ○ $M = 7, 8, d = 3..4,$ ○ $M = 11, d = 4..6,$ ○ $M = 13, d = 5..7,$ ○ $M = 16, d = 6..8,$ ○ $M = 17, d = 6..9,$ ○ $M = 19, d = 7..10,$ ○ $M = 23, d = 9..12,$ ○ $M = 29, d = 12..15,$ ○ $M = 31..32, d = 13..16,$ ○ $M = 37, d = 16..19,$ ○ $M = 41, d = 17..21,$ ○ $M = 43, d = 18..22,$	○ $M = 47, d = 21..24,$ ○ $M = 53, d = 23..27,$ ○ $M = 59, d = 26..29,$ ○ $M = 61, d = 27..30,$ ○ $M = 64, d = 29..32,$ ○ $M = 67, d = 30..33,$ ○ $M = 71, d = 32..35,$ ○ $M = 73, d = 33..36,$ ○ $M = 79, d = 36..39,$ ○ $M = 83, d = 38..41,$ ○ $M = 89, d = 41..44,$ ○ $M = 97, d = 45..48,$ ○ $M = 101, d = 48..50,$
--	---

- $M = 103, d = 48..51,$
- $M = 107, d = 50..53,$
- $M = 109, d = 51..54,$
- $M = 113, d = 52..56,$
- $M = 127, d = 60..63,$
- $M = 128, d = 60..64,$
- $M = 131, d = 61..65,$
- $M = 137, d = 64..68,$
- $M = 139, d = 65..69,$
- $M = 149, d = 70..74,$
- $M = 151, d = 71..75,$
- $M = 157, d = 75..78,$
- $M = 163, d = 78..81,$
- $M = 167, d = 81..83,$
- $M = 173, d = 82..86,$
- $M = 179, d = 86..89,$
- $M = 181, d = 88..90,$
- $M = 191, 193, d = 92..95,$
- $M = 197, d = 93..98,$
- $M = 199, d = 95..99,$
- $M = 211, d = 102..105,$
- $M = 223, d = 110..111,$
- $M = 227, d = 112..113,$
- $M = 229, d = 113..114,$
- $M = 233, d = 114..116,$
- $M = 239, d = 115..119,$
- $M = 241, d = 118..120,$
- $M = 251, d = 124..125,$
- $M = 256, d = 125..127,$
- $M = 257, d = 127..128,$
- $M = 263, d = 128..131,$
- $M = 269, d = 131..134,$
- $M = 271, d = 131..135,$
- $M = 277, d = 136..137,$
- $M = 281, d = 136..140,$
- $M = 283, d = 139..140,$
- $M = 293, d = 143..146,$
- $M = 307, d = 149..153,$
- $M = 311, d = 152..155,$
- $M = 313, d = 153..156,$
- $M = 317, d = 155..157,$
- $M = 331, d = 163..165,$
- $M = 337, d = 166..168,$
- $M = 347, d = 172..173,$
- $M = 349, d = 172..174,$
- $M = 353, d = 174..176,$
- $M = 359, d = 178,$
- $M = 367, d = 181..183,$
- $M = 373, d = 182..183,$
- $M = 379, d = 185,$
- $M = 383, d = 188,$
- $M = 389, d = 191,$
- $M = 397, d = 195,$
- $M = 401, d = 196..198,$
- $M = 409, d = 201..202,$
- $M = 419, d = 206..208,$
- $M = 421, d = 207..209,$
- $M = 431, d = 213,$
- $M = 433, d = 214,$
- $M = 443, d = 218..219,$
- $M = 449, d = 222..223,$
- $M = 457, d = 225..226,$
- $M = 461, d = 227,$
- $M = 463, d = 229..231,$
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008$ bitowe,
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 7..12,$

- wag Hamminga dla bloku $M = 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283$ bitowego,
- wypełniania drzewa binarnego o wysokości $h = 3..10$ na podstawie ciągu bitów
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień – oprócz $h = 10$.

5. Rekomendowany pakiet procedur dla ciągów długości 1 Mb

5.1. Procedury dla ciągów bloków bitów

- Andersona-Darlinga – $d = 16..32$,
- autokorelacji dla ciągu bloków

○ $d = 2, t = 1..249999$,	○ $d = 18, t = 1..27777$,
○ $d = 3, t = 1..166666$,	○ $d = 19, t = 1..26315$,
○ $d = 4, t = 1..124999$,	○ $d = 20, t = 1..24999$,
○ $d = 5, t = 1..99999$,	○ $d = 21, t = 1..23809$,
○ $d = 6, t = 1..83332$,	○ $d = 22, t = 1..22726$,
○ $d = 7, t = 1..71428$,	○ $d = 23, t = 1..21738$,
○ $d = 8, t = 1..62499$,	○ $d = 24, t = 1..20832$,
○ $d = 9, t = 1..55555$,	○ $d = 25, t = 1..19999$,
○ $d = 10, t = 1..49999$,	○ $d = 26, t = 1..19230$,
○ $d = 11, t = 1..45454$,	○ $d = 27, t = 1..18518$,
○ $d = 12, t = 1..41666$,	○ $d = 28, t = 1..17856$,
○ $d = 13, t = 1..38461$,	○ $d = 29, t = 1..17240$,
○ $d = 14, t = 1..35713$,	○ $d = 30, t = 1..16666$,
○ $d = 15, t = 1..33332$,	○ $d = 31, t = 1..16128$,
○ $d = 16, t = 1..31249$,	○ $d = 32, t = 1..15624$,
○ $d = 17, t = 1..29411$,	
- Corona – $d = 3..10$,
- Cramera-von Misesa – $d = 16..32$,
- częstości dla podprzedziału
 - $d = 16..32, [1/128; 1/4 + 1/128), M = 32$,
 - $d = 16..32, [1/64; 1/2 + 1/128), M = 32$,
 - $d = 16..32, [1/2 - 1/128; 3/4), M = 32$,
- gra w kości
 - kostka ośmiościenna – zgodności dla liczby kolejek do rozstrzygnięcia,

- kostka ośmiościenna – zgodności dla liczby zwycięstw,
- kostka sześcienna – zgodności dla liczby kolejek do rozstrzygnięcia,
- kostka sześcienna – zgodności dla liczby zwycięstw,
- kolekcjonera – $d = 2..6$,
- Kołmogorowa – $d = 16..32$,
- Kołmogorowa-Smirnowa – $d = 16..32$
 - statystyka K^+ ,
 - statystyka K^- ,
- Kuipera – $d = 16..32$,
- pochodnej ciągu bloków
 - pierwsza pochodna – $d = 2..8$,
 - druga pochodna – $d = 3..8$,
- porządku książek

○ $d = 4, k = 2..3$, ○ $d = 5, k = 2..4$, ○ $d = 6, k = 2..5$, ○ $d = 7, k = 2..6$, ○ $d = 8, k = 2..7$, ○ $d = 9, k = 2..8$, ○ $d = 10, k = 2..9$, ○ $d = 11, k = 2..10$, ○ $d = 12, k = 2..11$,	○ $d = 13, k = 2..12$, ○ $d = 14..17, k = 2..12$, ○ $d = 18, k = 3..11$, ○ $d = 19, k = 4..10$, ○ $d = 20, k = 5..10$, ○ $d = 21, k = 6..10$, ○ $d = 22, k = 7..10$, ○ $d = 23..25, k = 7..9$, ○ $d = 26, k = 5..9$,
--	---
- serii malejących z pomijaniem w przód – $d = 7..32$,
- serii malejących z pomijaniem w tył – $d = 7..32$,
- serii rosnących z pomijaniem w przód – $d = 7..32$,
- serii rosnących z pomijaniem w tył – $d = 7..32$,
- stosu książek

○ $d = 4, k = 2..3$, ○ $d = 5, k = 2..4$, ○ $d = 6, k = 2..5$, ○ $d = 7..8, k = 2..6$, ○ $d = 9, k = 2..8$, ○ $d = 10, k = 2..7$, ○ $d = 11, k = 2..8$, ○ $d = 12, k = 2..7$, ○ $d = 13, k = 2..5$,	○ $d = 14, k = 2..8$, ○ $d = 15, k = 2..6$, ○ $d = 16, k = 2..7$, ○ $d = 17, k = 2..6$, ○ $d = 18, k = 2..8$, ○ $d = 19, k = 2..7$, ○ $d = 20, k = 2..6$, ○ $d = 21, k = 2..9$, ○ $d = 22, k = 2..7$,
--	--

- $d = 23..25, k = 2..6,$
- $d = 26, k = 2..4,$
- sum do G
 - $d = 16..32, G = 16, 32, 64,$
 - $d = 16..18, G = 128,$
- sumy kwadratów odstępów bez zapętlania
 - $m = 1, d = 20..32,$
 - $m = 2, 3, 4, 5, d = 19..32,$
 - $m = 7, 8, 11, 13, 16, 17, 29,$
 $31, 32, 37, 41, d = 18..32,$
 - $m = 19, 23, d = 17..32,$
 - $m = 43, 47, d = 22..32,$
 - $m = 53, d = 24..31,$
 - $m = 59, 61, d = 25..30,$
 - $m = 59, 64, 67, 71, d = 31,$
 - $m = 919, 947, 977, 1021,$
 $d = 26,$
 - $m = 8192, 15625, k = 17..32,$
- sumy kwadratów odstępów z zapętlaniem
 - $m = 2, 3, d = 19..32,$
 - $m = 4, 5, 7, 8, 11, 13, 32, 37,$
 $41, 43, d = 18..32,$
 - $m = 16, 17, 19, 23, 29, 31,$
 $d = 17..32,$
 - $m = 47, d = 20..32,$
 - $m = 53, 59, 64, 67, 71,$
 $d = 31..32,$
 - $m = 71, d = 32,$
 - $m = 73, 79, 83, 89, 97, 103,$
 $109, d = 31,$
- sumy logarytmów odstępów bez zapętlania
 - $m = 2, d = 25..32,$
 - $m = 3, d = 22..32,$
- sumy logarytmów odstępów z zapętlaniem
 - $m = 2, d = 26..32,$
 - $m = 3, d = 23..24,$
- ściskania – $d = 11..20,$
- uniwersalny Maurera – $d = 1..10,$
- wag Hamminga w ciągu bloków
 - $d = 2, t = 2..5, k = 3,$
 - $d = 3$
 - $t = 2..4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 4$
 - $t = 2..3, k = 3, 5,$
 - $t = 4..5, k = 3,$
 - $d = 5$
 - $t = 2, k = 2, 4, 6,$
 - $t = 3..4, k = 2, 4,$
 - $t = 5..8, k = 2,$
 - $d = 6$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$

- $d = 7$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 8$
 - $t = 2, k = 3, 5, 7,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 9$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 10$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 11$
 - $t = 2, k = 2, 4, 6, 8,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 12$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 13$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 14$
 - $t = 2, k = 3, 5, 7, 9,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 15$
 - $t = 2, k = 2, 4, 6, 8, 10,$
 - $t = 3, k = 2, 4, 6,$
 - $t = 4, k = 2, 4,$
 - $t = 5..8, k = 2,$
- $d = 16$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 20$
 - $t = 2, k = 3, 5, 7, 9, 11,$
 - $t = 3, k = 3, 5,$
 - $t = 4..5, k = 3,$
- $d = 32$
 - $t = 2, k = 3, 5, 7, 9, 11,$
13,
 - $t = 3, k = 3,$

- wypełniania drzewa binarnego o wysokości $h = 3..11$ na podstawie ciągu bloków $d = 4..32$ bitowych
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień.

5.2. Procedury dla ciągów wektorów nienakładających się

- iloczynów
 - $t = 2..3, d = 16..32,$
 - $t = 4..6, d = 17..32,$
 - $t = 7..16, d = 18..32,$

- kolizji indeksów maksymalnych elementów
 - $t = 3..7, d = 11..22,$
 - $t = 8..16, d = 12..22,$
- logarytm prawdopodobieństwa – $t = 1, d = 6..20,$
- m najbliższych par

Tabela B.8. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka euklidesowa

$d \setminus t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16			62	55	51	47	44	41	39	37	36	34	33	32	31
17			60	54	49	45	42	40	38	36	35	33	32	31	30
18		68	58	52	48	44	41	39	37	35	34	32	31	30	29
19		66	57	51	46	43	40	38	36	34	33	31	30	29	28
20		64	55	50	45	42	39	37	35	33	32	31	29	28	27
21		62	54	48	44	41	38	36	34	32	31	30	29	28	27
22		61	53	47	43	40	37	35	33	32	30	29	28	27	26
23	73	60	52	46	42	39	36	34	32	31	30	28	27	26	26
24	72	58	51	45	41	38	36	34	32	30	29	28	27	26	25
25	70	57	50	44	40	37	35	33	31	30	28	27	26	25	25
26	69	56	49	43	40	37	34	32	31	29	28	27	26	25	24
27	68	55	48	43	39	36	34	32	30	29	27	26	25	24	24
28	66	54	47	42	38	35	33	31	29	28	27	26	25	24	23
29	65	53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24	23	22	22

Tabela B.9. Wartości m dla poszczególnych par t i d w teście m najbliższych par – metryka nieskończoność

$d \setminus t$	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
21															27
22								35	33	32	30	29	28	27	26
23					42	39	36	34	32	31	30	28	27	26	26
24				45	41	38	36	34	32	30	29	28	27	26	25
25			50	44	40	37	35	33	31	30	28	27	26	25	25
26		56	49	43	40	37	34	32	31	29	28	27	26	25	24
27		55	48	43	39	36	34	32	30	29	27	26	25	24	24
28		54	47	42	38	35	33	31	29	28	27	26	25	24	23
29		53	46	41	37	35	32	30	29	27	26	25	24	23	23
30	64	52	45	40	37	34	32	30	28	27	26	25	24	23	22
31	63	51	44	40	36	33	31	29	28	27	25	24	24	23	22
32	62	51	44	39	36	33	31	29	27	26	25	24			

- maksimum z t
 - $t = 2, d = 16..32,$
 - $t = 3..7, d = 17..32,$
 - $t = 8..16, d = 18..32,$
- momentu pierwszego rzędu – $t = 1, d = 1..30,$
- momentu drugiego rzędu – $t = 1, d = 2..30,$

- permutacji
 - $t = 2, d = 3..4,$
 - $t = 3..10, d = 2..32,$
- podziałów
 - $t = 4..6, d = 2..6,$
 - $t = 7, d = 3..7,$
 - $t = 8, d = 2..7,$
 - $t = 9..10, d = 2..8,$
 - $t = 11..16, d = 3..8,$
- pokerowy
 - $t = 3, d = 2..4,$
 - $t = 4..5, d = 2..6,$
 - $t = 6..9, d = 2..7,$
 - $t = 10..12, d = 2..8,$
 - $t = 13..14, d = 2..3,$
- potęgowo-dywergencyjny
 - $\delta = 1, t = 1, d = 13..17,$
 - $\delta = 2, t = 1, d = 9..10, 13,$
 - $\delta = 3, t = 1, d = 11..12,$
 - $\delta = 4, t = 1, d = 7..8, 10,$
 - $\delta = 5, t = 1, d = 7..9,$
 - $\delta = 6, t = 1, d = 6,$
- zgodności Pearsona
 - $t = 1, d = 2..5, 12..30,$
 - $t = 2, d = 7..15,$
 - $t = 3, d = 5..10,$
 - $t = 4, d = 4..7,$
 - $t = 5, d = 3..6,$
 - $t = 6, d = 3..5,$
 - $t = 7, d = 2..4,$
 - $t = 8, 10, d = 2..3,$
 - $t = 9, d = 2..3,$
 - $t = 11..12, d = 2,$
 - $t = 13..15, d = 2.$

5.3. Procedury dla ciągów wektorów nakładających się

- nakładających się permutacji – $t = 3..6, d = 2..32,$
- potęgowo-dywergencyjny – $\delta = 1$
 - $t = 2, d = 6..7,$
 - $t = 3, d = 4..5,$
 - $t = 5, d = 3,$
 - $t = 8, d = 2,$
- wolnych cel Marsagli
 - $t = 6, d = 3,$
 - $t = 9, d = 2,$
- zmodyfikowany test Pearsona dla przypadku rzadkiego
 - $t = 2, d = 8,$
 - $t = 4, d = 4,$
 - $t = 6, d = 3,$
 - $t = 9, d = 2.$

5.4. Procedury dla ciągów bitów

- autokorelacji dla ciągu binarnego – $t = 1..500000,$

- częstości wag Hamminga dla bloku $M = 4, 5, 7, 8, 11, 13, 16, 17, 19, 23, 29, 31, 32, 37, 41, 43, 47, 53, 59, 61, 64, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283, 293, 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397, 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499, 503, 509, 512, 521, 523, 541, 547, 557, 563, 569, 571, 577, 587, 593, 599, 601, 607, 613, 617, 619, 631, 641, 643, 647, 653, 659, 661, 673, 677, 683, 691, 701, 709, 719, 727, 733, 739, 743, 751, 757, 761, 769, 773, 787, 797, 809, 811, 821, 823, 827, 829, 839, 853, 857, 859, 863, 877, 881, 883, 887, 907, 911, 919, 929, 937, 941, 947, 953, 967, 971, 977, 983, 991, 997$ bitowego,
- liczby skoków w profilu liniowym,
- losowych wycieczek dla podciągów długości $M = 128, 134, 142, 146, 158, 166, 178, 194, 202, 206, 214, 218, 226, 254, 256, 262, 274, 278, 298, 302, 314, 326, 334, 346, 358, 362, 382, 386, 394, 398, 422, 446, 454, 458, 466, 478, 482, 502, 512, 514, 526, 538, 542, 554, 562, 566, 586, 614, 622, 626, 634, 662, 674, 694, 698, 706, 718, 734, 746, 758, 766, 778, 794, 802, 818, 838, 842, 862, 866, 878, 886, 898, 914, 922, 926, 934, 958, 974, 982, 998, 1006, 1018, 1024, 1042, 1046, 1082, 1094, 1114, 1126, 1138, 1142, 1154, 1174, 1186, 1198, 1202, 1214, 1226, 1234, 1238, 1262, 1282, 1286, 1294, 1306, 1318, 1322, 1346, 1354, 1366, 1382, 1402, 1418, 1438, 1454, 1466, 1478, 1486, 1502, 1514, 1522, 1538, 1546, 1574, 1594, 1618, 1622, 1642, 1646, 1654, 1658, 1678, 1706, 1714, 1718, 1726, 1754, 1762, 1766, 1774, 1814, 1822, 1838, 1858, 1874, 1882, 1894, 1906, 1934, 1942, 1954, 1966, 1982, 1994, 2018, 2026, 2038, 2042, 2048$
 - liczby powrotów do 0,
 - liczby przejść przez 0 ze zmianą strony,
 - łącznego „czasu” odchylenia w prawo,
 - maksymalnego odchylenia ścieżki,
- najdłuższej serii 0 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192$ bitowym,
- najdłuższej serii 1 w bloku $M = 4, 8, 16, 32, 64, 128, 256, 512, 1000, 1024, 2048, 4096, 8192$ bitowym,
- nakładających się t bitowych wzorców 0 oraz 1 w bloku M bitowym
 - $t = 4, M = 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 5, M = 32, 64, 128, 256, 512, 1000, 1024,$
 - $t = 6, M = 64, 128, 256, 512, 1000, 1024,$
 - $t = 7, M = 128, 256, 512, 1000, 1024,$
 - $t = 8, M = 256, 512, 1000, 1024,$
 - $t = 9, M = 512, 1000, 1024,$

- $t = 10, M = 1000, 1024,$
- nienakładających się d bitowych wzorców w bloku M bitowym (wzorce w zapisie heksadecymalnym):
 - $d = 2, M = 524288,$ wzorce: 1, 2,
 - $d = 3, M = 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 524288,$ wzorce: 1, 3, 4, 6,
- niezależności wag Hamminga dla bloku $M = 4, 5, 7$ bitowego,
- niezależności wag Hamminga – wariant 2

○ $M = 4, d = 1,$ ○ $M = 7, 8, d = 3..4,$ ○ $M = 11, d = 4..6,$ ○ $M = 13, d = 5..7,$ ○ $M = 16, d = 6..8,$ ○ $M = 17, d = 6..9,$ ○ $M = 19, d = 7..10,$ ○ $M = 23, d = 9..12,$ ○ $M = 29, d = 12..15,$ ○ $M = 31..32, d = 13..16,$ ○ $M = 37, d = 16..19,$ ○ $M = 41, d = 17..21,$ ○ $M = 43, d = 18..22,$ ○ $M = 47, d = 21..24,$ ○ $M = 53, d = 23..27,$ ○ $M = 59, d = 26..29,$ ○ $M = 61, d = 27..30,$ ○ $M = 64, d = 29..32,$ ○ $M = 67, d = 30..33,$ ○ $M = 71, d = 32..35,$ ○ $M = 73, d = 33..36,$ ○ $M = 79, d = 36..39,$ ○ $M = 83, d = 38..41,$ ○ $M = 89, d = 41..44,$ ○ $M = 97, d = 45..48,$ ○ $M = 101, d = 48..50,$ ○ $M = 103, d = 48..51,$ ○ $M = 107, d = 50..53,$	○ $M = 109, d = 51..54,$ ○ $M = 113, d = 52..56,$ ○ $M = 127, d = 60..63,$ ○ $M = 128, d = 60..64,$ ○ $M = 131, d = 61..65,$ ○ $M = 137, d = 64..68,$ ○ $M = 139, d = 65..69,$ ○ $M = 149, d = 70..74,$ ○ $M = 151, d = 71..75,$ ○ $M = 157, d = 75..78,$ ○ $M = 163, d = 78..81,$ ○ $M = 167, d = 81..83,$ ○ $M = 173, d = 82..86,$ ○ $M = 179, d = 86..89,$ ○ $M = 181, d = 88..90,$ ○ $M = 191, 193, d = 92..95,$ ○ $M = 197, d = 93..98,$ ○ $M = 199, d = 95..99,$ ○ $M = 211, d = 102..105,$ ○ $M = 223, d = 110..111,$ ○ $M = 227, d = 112..113,$ ○ $M = 229, d = 113..114,$ ○ $M = 233, d = 114..116,$ ○ $M = 239, d = 115..119,$ ○ $M = 241, d = 118..120,$ ○ $M = 251, d = 124..125,$ ○ $M = 256, d = 125..127,$ ○ $M = 257, d = 127..128,$
---	---

- $M = 263, d = 128..131,$
- $M = 269, d = 131..134,$
- $M = 271, d = 131..135,$
- $M = 277, d = 136..137,$
- $M = 281, d = 136..140,$
- $M = 283, d = 139..140,$
- $M = 293, d = 143..146,$
- $M = 307, d = 149..153,$
- $M = 311, d = 152..155,$
- $M = 313, d = 153..156,$
- $M = 317, d = 155..157,$
- $M = 331, d = 163..165,$
- $M = 337, d = 166..168,$
- $M = 347, d = 172..173,$
- $M = 349, d = 172..174,$
- $M = 353, d = 174..176,$
- $M = 359, d = 178,$
- $M = 367, d = 181..183,$
- $M = 373, d = 182..183,$
- $M = 379, d = 185,$
- $M = 383, d = 188,$
- $M = 389, d = 191,$
- $M = 397, d = 195,$
- $M = 401, d = 196..198,$
- $M = 409, d = 201..202,$
- $M = 419, d = 206..208,$
- $M = 421, d = 207..209,$
- $M = 431, d = 213,$
- $M = 433, d = 214,$
- $M = 443, d = 218..219,$
- $M = 449, d = 222..223,$
- $M = 457, d = 225..226,$
- $M = 461, d = 227,$
- $M = 463, d = 229..231,$
- rozkładu złożoności liniowych dla ciągu dzielonego na fragmenty $m = 64, 128, 192, 256, 320, 448, 512, 704, 832, 1024, 1088, 1216, 1472, 1856, 1984, 2048, 2368, 2624, 2752, 3008$ bitowe,
- sumy kwadratów modułów DFT dla pociągów 2^d bitowych – $d = 7..12,$
- wag Hamminga dla bloku $M = 101, 103, 107, 109, 113, 127, 128, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199, 211, 223, 227, 229, 233, 239, 241, 251, 256, 257, 263, 269, 271, 277, 281, 283$ bitowego,
- wypełniania drzewa binarnego o wysokości $h = 3..10$ na podstawie ciągu bitów
 - zgodności dla pozycji kolizji,
 - zgodności dla wypełnień – oprócz $h = 10$.

5.5. Procedury realizowane poprzez kombinację liniową p_{val} innych procedur

5.5.1. Test potęgowo-dywergencyjny

- $d = 6$
 - $U^{6PD.2} = 0,334099U^{6PD.6} + 0,666423U^{6LP} - 0,000250,$
 - $U^{6PD.3} = 0,501058U^{6PD.6} + 0,499542U^{6LP} - 0,000304,$
 - $U^{6PD.4} = 0,667768U^{6PD.6} + 0,332771U^{6LP} - 0,000280,$
- $d = 7$
 - $U^{7P} = 0,250621U^{7PD.4} + 0,749861U^{7LP} - 0,000249,$
 - $U^{7PD.2} = 0,501064U^{7PD.4} + 0,499586U^{7LP} - 0,000348,$

- $U^{7PD.3} = 1,502790U^{7PD.4} - 0,602514U^{7PD.5} + 0,0997217U^{7LP} - 0,000003,$
- $U^{7PD.6} = -0,997838U^{7PD.4} + 1,997510U^{7PD.5} + 0,000181,$
- $U^{7PD.7} = -1,992110U^{7PD.4} + 2,991120U^{7PD.5} + 0,000543,$

- $d = 8$

- $U^{8P} = 0,251315U^{8PD.4} + 0,749800U^{8LP} - 0,000595,$
- $U^{8PD.2} = 0,502248U^{8PD.4} + 0,499239U^{8LP} - 0,000797,$
- $U^{8PD.3} = 1,505600U^{8PD.4} - 0,605084U^{8PD.5} + 0,0994773U^{8LP} - 0,000003,$
- $U^{8PD.6} = -0,996343U^{8PD.4} + 1,995600U^{8PD.5} + 0,000402,$
- $U^{8PD.7} = -1,986020U^{8PD.4} + 2,983790U^{8PD.5} + 0,001205,$

- $d = 9$

- $U^{9P} = 0,502199U^{9PD.2} + 0,498640U^{9LP} - 0,000441,$
- $U^{9PD.3} = 0,666323U^{9PD.2} + 0,335322U^{9PD.5} - 0,000869,$
- $U^{9PD.4} = 0,332286U^{9PD.2} + 0,669354U^{9PD.5} - 0,000868,$
- $U^{8PD.6} = -0,996343U^{8PD.4} + 1,995600U^{8PD.5} + 0,000402,$

- $d = 10$

- $U^{10P} = 0,504825U^{10PD.2} + 0,497033U^{10LP} - 0,000966,$
- $U^{10PD.3} = 0,498536U^{10PD.2} + 0,503243U^{10PD.4} - 0,000929,$

- $d = 11$

- $U^{11P} = 0,345815U^{11PD.3} + 0,662009U^{11LP} - 0,004030,$
- $U^{11PD.2} = 0,681915U^{11PD.3} + 0,325688U^{11LP} - 0,003918,$

- $d = 12$

- $U^{12PD.2} = 0,491293U^{12P} + 0,516266U^{12PD.3} - 0,003873,$

gdzie $U^{dPD.\delta}$ oznacza p_{val} dla procedury testu potęgowo-dywergencyjnego dla δ , a U^{dLP}

– testu logarytm-prawdopodobieństwa, U^{dP} – zgodności Pearsona, dla bloku d bitowego,

5.5.2. Test sumy kwadratów odstępów bez zapętlania

- $d = 26$

- $U^{929} = 0,640908U^{919} + 0,359437U^{947} - 0,000172,$
- $U^{937} = 0,352599U^{919} + 0,647746U^{947} - 0,000172,$
- $U^{941} = 0,210016U^{919} + 0,790237U^{947} - 0,000126,$
- $U^{953} = -0,110103U^{919} + 1,048380U^{947} + 0,061751U^{997} - 0,000014,$
- $U^{967} = -0,165127U^{919} + 0,769463U^{947} + 0,579658U^{997} - 0,184133U^{1021} - 0,000071,$

- $U^{971} = -0,154657U^{919} + 0,659619U^{947} + 0,707944U^{997} - 0,213047U^{1021} - 0,000073,$
- $U^{977} = 0,129280U^{919} + 1,306520U^{997} - 0,435563U^{1021} - 0,000115,$
- $U^{983} = -0,162823U^{947} + 1,093560U^{997} - 0,256259U^{1021} - 0,000058,$
- $U^{991} = 0,052710U^{947} + 1,092920U^{997} - 0,145606U^{1021} - 0,000013$
- $U^{1009} = 0,497078U^{997} + 0,503159U^{1021} - 0,000118,$
- $U^{1013} = 0,329575U^{997} + 0,670634U^{1021} - 0,000105,$

gdzie U^j oznacza p_{val} dla procedury odpowiadającej odstępowi równemu j ,

- $d = 31$

- $U^{61} = -0,053259U^{53} + 0,622073U^{59} + 0,588851U^{64} - 0,158165U^{67} + 0,000249,$

gdzie U^j oznacza p_{val} dla procedury odpowiadającej odstępowi równemu j ,

5.5.3. Test sumy kwadratów odstępów z zapętnieniem

- $d = 31$

- $U^{61} = -0,052396U^{53} + 0,619383U^{59} + 0,592723U^{64} - 0,160118U^{67} - 0,000206,$
- $U^{71} = -0,131845U^{64} + 0,446600U^{67} + 0,773307U^{73} - 0,088446U^{79} - 0,000194,$
- $U^{101} = 0,324348U^{97} + 0,676807U^{103} - 0,000579,$
- $U^{107} = 0,324922U^{103} + 0,676118U^{109} - 0,000521,$

gdzie U^j oznacza p_{val} dla procedury odpowiadającej odstępowi równemu j ,

- $d = 32$

- $U^{61} = -0,052753U^{53} + 0,620134U^{59} + 0,592372U^{64} - 0,160159U^{67} + 0,000205,$

gdzie U^j oznacza p_{val} dla procedury odpowiadającej odstępowi równemu j .