Wiadomości Statystyczne. The Polish Statistician, 2025, *70*(9), 1–33 https://doi.org/10.59139/ws.2025.09.1
Studia metodologiczne / Methodological studies
Zgłoszony/submitted: 17.10.2022, poprawiony/revised: 25.05.2025, zaakceptowany/accepted: 5.06.2025

Propozycja nowej metody testowania ciągów losowych

Marek Leśniewicz[a]

Streszczenie. W 2022 r. amerykański National Institute of Standards and Technology wycofał swoje rekomendacje dla zestawu testów do badania losowości ciągów (*A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*), co pozbawiło środowisko badaczy i użytkowników formalnie uznawanego narzędzia weryfikacji tych ciągów. Celem artykułu jest przedstawienie propozycji nowej, autorskiej metody testowania ciągów losowych, obiektywnej i wolnej od specyficznych właściwości i ograniczeń innych znanych zestawów testów. Metoda została oparta na rzeczywistych właściwościach ciągów losowych o dyskretnym, równomiernym rozkładzie prawdopodobieństwa i wykorzystuje elementarne parametry tych ciągów opisywane przez probabilistykę, statystykę i teorię informacji. W przeciwieństwie do dotychczas stosowanych zestawów testów proponowana metoda jest jednoparametryczna, teoretycznie nie zależy od liczebności próby, a wynik weryfikacji zaledwie jednej próby ciągu losowego o odpowiedniej liczebności jednoznacznie kwalifikuje bądź dyskwalifikuje go jako realizację ciągu zmiennych losowych o domyślnie referencyjnych właściwościach i parametrach probabilistycznych. Metoda została opracowana jako narzędziowe wsparcie projektu kwantowego mikrofalowego generatora ciągów losowych o wydajności 1 Gbit/s, skonstruowanego w ramach własnych prac statutowych w Wojskowym Instytucie Łączności.

Słowa kluczowe: ciąg losowy, rozkład prawdopodobieństwa, test chi-kwadrat

JEL: C12, C13, C15

A proposal for a new method of testing random sequences

Abstract. The withdrawal of the recommendation for a test suite to the randomness of sequences by the US National Institute of Standards and Technology in 2022 (*A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*) deprived researchers and users of this tool of a formally recognised and thus commonly accepted instrument for verifying such sequences. The aim of this paper is to present the author's novel method of random sequence testing, which is objective and free from the specific properties and limitations of other well-known test sets. The method was based on real properties of random sequences with a discrete uniform probability distribution. It uses the elementary parameters of these sequences, described by the probability theory, statistics and the information theory. In contrast to the sets of tests used so far, the proposed method is single-parameter. Theoretically, it does not depend on the sample size, and the result of the verification of only one test of a random sequence of an appropriate size unambiguously qualifies or disqualifies it as the implementation of a sequence of random variables with default reference properties and probabilistic parameters. The method was developed as a tool to support the project of a quantum microwave random sequence generator with the capacity of 1 Gbit/s. It was constructed as part of the statutory work of the Military Communications Institute.

Keywords: random sequence, probability distribution, chi-square test

[a] Wojskowy Instytut Łączności – Państwowy Instytut Badawczy, Polska / Military Communications Institute – National Research Institute, Poland. ORCID: https://orcid.org/0000-0001-5505-1268.
E-mail: m.lesniewicz@wil.waw.pl.

1. Wprowadzenie

Ciągi losowe są obecne w wielu dziedzinach nauki i techniki, np. kryptografii, telekomunikacji, cyfrowym przetwarzaniu sygnałów, technice algorytmów randomizowanych, statystyce, obliczeniach numerycznych czy symulacjach stochastycznych. We wszystkich tych zastosowaniach powinno się używać ciągów doskonale losowych (ang. *perfectly random sequences*), tzn. bezwarunkowo spełniających ściśle określone teoretyczne wymagania probabilistyczne (nie statystyczne), oraz prawdziwie losowych, tzn. charakteryzujących się nieprzewidywalnością (ang. *unpredictability*) wartości kolejnych elementów ciągu i niepowtarzalnością (ang. *unrepeatability*) w sensie prawdopodobieństwa ponownej realizacji, równego 2^{-m}, gdzie m jest liczebnością dowolnego podciągu w danej próbie ciągu.

Z tych względów w nauce i technice w zasadzie nie stosuje się już generowanych algorytmicznie ciągów pseudolosowych. W przypadku kryptografii powód jest oczywisty, ujęty w ścisłe normy prawne, ale nawet w zastosowaniach niewymagających nieprzewidywalności stwierdza się, że wyniki np. symulacji stochastycznych są niedokładne z powodu niskiej jakości statystyk ciągów generowanych algorytmicznie. Autorzy metod generowania ciągów pseudolosowych mają tego pełną świadomość i przyznają, że „zbudowanie generatora ciągu, który przejdzie wszystkie testy statystyczne, jest nierealizowalnym marzeniem" (L'Ecuyer, 2012, s. 1). Wynika to z tego, że funkcje generujące ciągi zwykle nie są intencjonalnie konstruowane zgodnie z zaawansowanymi zasadami algebry i teorii liczb, nie mają dowiedzionych właściwości i parametrów statystycznych, bywają natomiast skrajnie upraszczane, co – zwiększając potencjalną szybkość generowania ciągów – powoduje ich wynikową degradację statystyczną. Z tych powodów duża część generatorów liczb losowych w komputerowych pakietach matematycznych jest oceniana jako nienadająca się do zastosowań dużej wagi.

Nie oznacza to, że wszystkie ciągi pseudolosowe są złe – zaawansowane funkcje kryptograficzne generują ciągi w zasadzie nieodróżnialne od doskonale losowych (ang. *looks like random*)[1]. Problemem pozostaje jednak niepowtarzalność w sensie unikatowości prób używanych np. do obliczeń numerycznych – jeśli wielokrotnie używa się tych samych prób, to pomimo ich dobrych statystyk wyniki wielu obliczeń będą dotyczyły jednej realizacji ciągu i w procesie optymalizacji algorytmów obliczeń mogą się do niej dopasować lub z nią skorelować, a w przypadku innej realizacji mogą być już inne. Ten sam problem dotyczy losowego wyboru reprezentacji w badaniach statystycznych i błędu losowania (ang. *sampling error*), ponieważ „w zastosowaniach wnioskowania statystycznego coraz częściej nie zostaje spełnione fundamentalne założenie o posiadaniu przez badacza próby losowej (probabilistycznej)", a sama „duża

[1] Jest to pokazane w części 8.

liczebność próby, której doboru dokonano za pomocą techniki nieprobabilistycznej, nie może stanowić alternatywy dla wyboru losowego", zatem próby „powinny być generowane przez mechanizm losowy" (Szreder, 2022, s. 1; zob. też Szreder i Kozłowski, 2024).

Możliwe jest jednak uzyskiwanie prawdziwie losowych i bliskich matematycznej doskonałości ciągów z generatorów sprzętowych TRNG (ang. *true random number generator*), które wytwarzają intencjonalnie nieprzewidywalne i niepowtarzalne ciągi o referencyjnych statystykach, wynikających z modelowania stochastycznych właściwości i parametrów zjawisk losowych, które są ich źródłami. Takie generatory uznaje się za wzorcowe źródła ciągów doskonale losowych, ponieważ doskonałość wytwarzanych przez nie ciągów wynika nie tyle z testowania, ile z dowiedzionego bezpieczeństwa kryptograficznego. Polega ono na analitycznym wykazaniu, że w racjonalnym czasie analizy statystycznej żaden ciąg wytwarzany przez generator nie okaże nawet oznaki nielosowości, a ponieważ jego ewentualne usterki techniczne, zgodnie z obecnie obowiązującymi zaleceniami (International Telecommunication Union, 2019), są identyfikowane za pomocą mechanizmów autotestowania, to nigdy nie wygeneruje on ciągu nielosowego. Ów racjonalny czas jest przyjmowany jako dotychczasowy czas istnienia Wszechświata.

Generatory TRNG są trudne do skonstruowania i kosztowne w produkcji, ale w przypadku kryptograficznego bezpieczeństwa państwa nikt nie pyta o problemy technologiczne i związane z nimi koszty, tylko o to, czy dany generator ma certyfikat bezpieczeństwa służby ochrony państwa, który formalnie gwarantuje bezwarunkowe spełnienie wszystkich wymagań matematycznych w określonym czasie i dla dowolnej liczebności wygenerowanych ciągów w założonych, klimatyczno-mechanicznych i elektromagnetycznych warunkach pracy (Leśniewicz, 2009).

Jednak niezależnie od tego, czy dany generator otrzymał dowód bezpieczeństwa i potwierdzający go certyfikat, użytkownik generowanych przezeń ciągów powinien mieć możliwość weryfikacji ich losowości. Celem niniejszej pracy jest zatem przedstawienie propozycji nowej, autorskiej metody testowania ciągów losowych, obiektywnej i wolnej od specyficznych właściwości i ograniczeń dotychczas stosowanych testów losowości.

2. Probabilistyczny model ciągu losowego

Generowanie ciągów losowych sprowadza się do wytwarzania ciągów binarnych o dyskretnym, równomiernym rozkładzie prawdopodobieństwa. Nie ma sprzętowych generatorów ciągów o innych rozkładach (Jacak i in., 2021), a bardziej złożone rozkłady uzyskuje się przez numeryczne przetwarzanie ciągów binarnych w ciągi o rozkładzie normalnym i innych (Sulewski, 2019).

Probabilistyczny model ciągu doskonale losowego o takim rozkładzie wygląda tak. Dana jest przestrzeń zdarzeń elementarnych $\Omega = \{0, 1\}$ na σ-ciele $\mathcal{F} = 2^{\Omega} = \{\emptyset, \Omega, \{0\}, \{1\}\}$ i prawdopodobieństwach $P(0) = P(1) = 1/2$. Modelem ciągu doskonale losowego jest zatem ciąg N-wymiarowych, niezależnych zmiennych losowych $X_1, ..., X_N$, określonych na przestrzeni $\Omega = \{0, 1\}^N$ i przyjmujących wartości ze zbioru $\{0, 2^N - 1\}$ o dyskretnym, równomiernym rozkładzie prawdopodobieństwa $P(X_1, ..., X_N) = 1/2^N$.

Ciąg zmiennych losowych nie jest jednak ciągiem losowym w sensie arytmetycznego ciągu liczb. Tak rozumiany ciąg jest realizacją ciągu zmiennych losowych i jego opis jest pochodną opisu probabilistycznego, zasadniczo w dziedzinie statystyki matematycznej. Można zatem powiedzieć, że probabilistyka opisuje ciąg losowy *a priori* i taki ciąg jest „niemierzalny", ponieważ jest opisany prawdopodobieństwami zmiennych losowych $P(X_1, ..., X_N)$ jako miarami przewidywalności. Natomiast arytmetyczny ciąg liczb jest realizacją ciągu losowego *a posteriori*, tzn. po zapisaniu próby o danej liczebności, charakteryzującej się statystykami z próby, po obliczeniu średnich częstości realizacji N-wymiarowych zmiennych losowych $X_1, ..., X_N$, jako $p(X_1, ..., X_N)$. Wartości $P(X_1, ..., X_N)$ i $p(X_1, ..., X_N)$ są zbieżne, zgodnie z prawami wielkich liczb, ale nigdy nie są sobie równe (Feller, 2006).

Chcąc zatem skonstruować generator ciągów losowych spełniających wymagania probabilistyczne, a następnie mierzyć i analizować wytwarzane przezeń ciągi, trzeba rozdzielić probabilistykę i statystykę, ponieważ właściwości i parametry tych ciągów *a priori* i *a posteriori* znacząco się różnią.

Probabilistyczna analiza prawdopodobieństw jest względnie prosta w przypadku liczby wymiarów zmiennych losowych $N = 1$ lub co najwyżej $N = 2$ (dalej w artykule używana będzie forma „zmiennych losowych N-wymiarowych"). W sytuacji kiedy niezbędna jest analiza rozkładów dla np. $N = 16$, opisanych $2^N = 65\,536$ różnymi prawdopodobieństwami, staje się ona nie tyle niemożliwa, ile bardzo pracochłonna i żmudna. Aby obejść tę trudność, Shannon (1949) wprowadził pojęcie *entropii informacyjnej* ciągu jako globalnej miary jego losowości[2], która sprowadza nawet najbardziej złożony N-wymiarowy rozkład zmiennej losowej $X_1, ..., X_N$ do jednoparametrycznej, więc prostej i wygodnej, ale już nieodwracalnej wartości

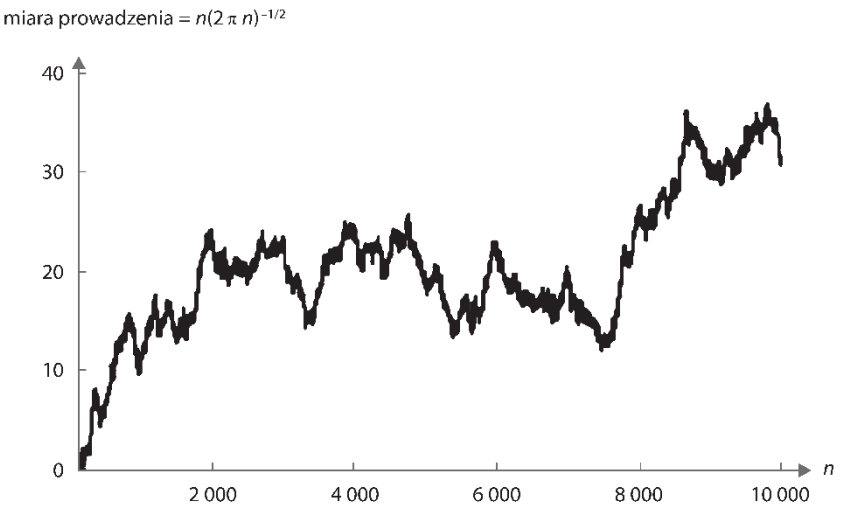
$$H(X_1, \ldots, X_N) = -\frac{1}{N}\sum_{i=1}^{k} P(X_i)\log_2 P(X_i), \tag{1}$$

[2] Owo wprowadzenie miało charakter heurystyczny i 75 lat temu nikt się nie spodziewał, że intuicyjnie zaproponowana zależność stanie się ważnym narzędziem w tak różnych dziedzinach nauki i techniki, jak teoria układów dynamicznych czy teoria kodowania.

gdzie $P(X_i)$ jest prawdopodobieństwem i-tej zmiennej losowej X_i, a $k = 2^N$ stanowi liczbę potencjalnych realizacji tej zmiennej, którą można utożsamiać z liczbą stopni swobody *df*.

Te dwa podejścia – probabilistyczne i teorioinformacyjne – są w zasadzie wystarczające do pełnej analizy losowości ciągów dowolnie wymiarowych zmiennych losowych. Problem zaczyna się jednak w przypadku analizowania ich realizacji, które okazują się pełne pozornie najdziwniejszych właściwości i parametrów, zwanych *błądzeniem przypadkowym*, choć w pełni opisanych i uzasadnionych matematycznie (Bobrowski, 2002; Feller, 2006), a ponadto zweryfikowanych pomiarowo (Leśniewicz, 2009).

Wykr. 1. Zjawisko błądzenia przypadkowego – ciąg rzutów monetą



Źródło: Feller (2006).

Najwięcej trudności przysparza odpowiedź na najprostsze pytanie: czy w ciągu n-elementowym, stanowiącym realizację ciągu niezależnych zmiennych losowych o prawdopodobieństwie $P(X) = 1/2$, występuje równowaga elementów, umownie nazywanych {0} i {1}? W praktyce – a teoria w pełni to uzasadnia – taka równowaga występuje bardzo rzadko, czego przykładem mogą być wyniki rzutu monetą, zilustrowane na wykr. 1. Wielu matematyków rzucało nią nawet kilkadziesiąt tysięcy razy i przekonało się, że teoria mówi prawdę – co Feller (2006), już po 10 000 takich osobistych doświadczeń, skomentował następująco: „niewielu ludzi uwierzy, że uczciwa moneta może dać niedorzeczny ciąg, w którym zmiana prowadzenia nie nastąpi w ciągu milionów kolejnych doświadczeń, jest to jednak zjawisko raczej powszechne dla dobrych monet" (s. 87). „Dobrych monet", czyli zmiennych losowych

o prawdopodobieństwach „orła" i „reszki" równych $P(O) = P(R) = 1/2$. Ten przykład pokazuje, że ciągi losowe nie lubią stanu równowagi i zwykle w dłuższych przedziałach utrzymują prowadzenie po jednej ze stron. Jeśli jednak moneta jest „dobra", to ciąg wraca do stanu równowagi, by jednak szybko go opuścić i dalej błądzić przypadkowo, zwykle już po drugiej stronie realizacji zmiennej losowej, co zgodnie z prawami wielkich liczb sprowadza statystyki z próby $p(X)$ do wartości definicyjnego prawdopodobieństwa $P(X) = 1/2$. Zjawisko to jest ściśle opisane wieloma wzajemnie dopełniającymi się zależnościami probabilistycznymi, m.in. prawami arcusa sinusa czy prawami iterowanego logarytmu (Feller, 2006).

Powyższy przykład – który dotyczy każdego generatora i wytwarzanych przezeń ciągów losowych – wskazuje, że prawdopodobieństwa $P(X_1, ..., X_N) = 1/2^N$ odpowiadają średnim częstościom realizacji N-wymiarowych zmiennych losowych $X_1, ..., X_N$, jako $p(X_1, ..., X_N)$, ale dopiero dla ciągów o bardzo dużej liczebności, zgodnie z prawami wielkich liczb. W związku z tym matematycznie nie ma sensu pytanie, czym jest „liczba losowa" i czy np. liczba $0110\ 1000\ 0011\ 1000\ 1101\ 1001\ 0001_B = 109\ 284\ 753_D$ jest czy nie jest losowa. Zasadne i weryfikowalne jest natomiast pytanie w sensie hipotezy, czy konkretna realizacja ciągu losowego o liczebności co najmniej 1 000 000 elementów może być realizacją doskonale losowego ciągu N-wymiarowych zmiennych losowych o $P(X_1, ..., X_N) = 1/2^N$.

Wynika z tego, że nie ma również sensu badanie próby ciągu $n = 100$ elementów, zawierającego np. kolejnych $m = 20$ elementów o wartości {0} lub {1}, i arbitralne stwierdzenie, że z tego powodu jest on nielosowy. Jeśli ta próba jest podciągiem ciągu doskonale losowego o liczebności 1 GB, tzn. liczącego $n = 8 \times 2^{30} = 8\ 589\ 934\ 592$ elementy, to w takiej próbie z natury rzeczy zdarzają się realizacje o $m = 20$ kolejnych elementach o wartościach {0} lub {1}, ponieważ prawdopodobieństwo ich wystąpienia wynosi $P = 2^{-m} = 2^{-20} = 1/1\ 048\ 576$. Jak zatem łatwo obliczyć, w próbie ciągu o takiej liczebności występuje ich średnio po $n \times 2^{-m} = 8192$, typowo od 8000 do 8400, więc nie tylko mogą, ale muszą występować w takim właśnie przedziale; w przeciwnym wypadku na pewno nie jest to próba ciągu doskonale losowego. Mówiąc obrazowo – im próba ciągu będzie liczebniejsza, tym większe prawdopodobieństwo, że znajdą się w niej prawie wszystkie potencjalne realizacje zmiennych losowych o dowolnie dużych wymiarach N, a w próbie ciągu doskonale losowego o liczebności 1 GB można spodziewać się incydentalnie nawet 32-elementowych podciągów o wartościach {0} lub {1}.

Wniosek z tej analizy jest taki, że badanie losowości ciągu ma sens tylko dla prób o dużej, najlepiej możliwie największej technicznie liczebności i dla zmiennych losowych o jak największych wymiarach N. Założenie to stoi w sprzeczności z założeniami wykorzystania dotychczas stosowanych testów, takich jak testy National Institute of Standards and Technology (NIST; Rukhin i in., 2010), przewidujących badania ciągów o liczebności rzędu $n = 100$ czy nawet 10 000 elementów, a ponadto dla zmiennych

losowych o małej liczbie wymiarów N. Losowość tak mało liczebnych prób jest nieweryfikowalna, zakres badań nieadekwatny do potencjalnych nielosowości, a powtarzanie badań na podobnie mało liczebnych próbach może tylko wprowadzać w błąd z powodu niepowtarzalności wyników pomiarów.

3. Rzeczywisty ciąg losowy

Jak już wspomniano, przykład uczciwej monety mogącej dać niedorzeczny ciąg, w którym zmiana prowadzenia nie nastąpi w ciągu milionów kolejnych doświadczeń, został dokładnie opisany analitycznie wieloma zależnościami probabilistycznymi. W dalszej części artykułu wykorzystane są jednak tylko te zależności, które dotyczą ciągów doskonale losowych i znajdują zastosowanie w proponowanej tutaj metodzie.

Suma doskonale losowego ciągu N-wymiarowych zmiennych losowych $X_1, ..., X_N$ jest w pełni opisywana liczebnością elementów n, prawdopodobieństwami $P(X_1, ..., X_N) = 1/2^N$ oraz unormowaną wartością oczekiwaną $E(X_1, ..., X_N) = 1/2^N$ i unormowaną wariancją $V(X_1, ..., X_N) = [\, P(X_1, ..., X_N)\,(1 - P(X_1, ..., X_N))\,]\; N / n = 1/2^N\,(1 - 1/2^N)\; N / n$, gdzie n / N stanowi liczebność realizacji zmiennych losowych N-wymiarowych.

Wariancja $V(X_1, ..., X_N)$, mimo że malejąca w funkcji liczebności n, nigdy nie osiąga wartości zerowej, a jej zmniejszanie się jest relatywnie wolne – w relacji $1/n$. Z tak rozumianej wariancji można wyprowadzić zależność na odchylenie standardowe $\sigma(X_1, ..., X_N) = V(X_1, ..., X_N)^{1/2}$, ale nie ma ono intuicyjnej ilościowej interpretacji liczbowej. Taką ilościową – w sensie pomiarowym – interpretację ma natomiast moduł średniego odchylenia od wartości oczekiwanej $E(X_1, ..., X_N) = P(X_1, ..., X_N) = 1/2^N$, równy dla dowolnej N-wymiarowej zmiennej losowej (Leśniewicz, 2009):

$$\alpha(X_1, \ldots, X_N)_{śr} = \left|\frac{n_i N}{n} - E(X_1, \ldots, X_N)\right| = \left|\frac{n_i N}{n} - \frac{1}{2^N}\right| \cong \sqrt{\frac{2}{\pi} V(X_1, \ldots, X_N)} = \sqrt{\frac{(1 - 1/2^N)\, N}{2^{N-1}\pi n}}, \tag{2}$$

gdzie n_i stanowi liczbę realizacji każdej ze zmiennych losowych $X_1, ..., X_N$.

Znak ≅ wynika z przybliżenia Stirlinga dla pewnych zależności kombinatorycznych, więc zależność jest dokładna dla dużych n. Odchylenie standardowe i moduł średniego odchylenia wiąże relacja $\sigma(X_1, ..., X_N) = (\pi/2)^{1/2}\; \alpha(X_1, ..., X_N)$. Wszystkie te odchylenia mają źródła nie w potencjalnej nielosowości ciągu, lecz w naturalnym wpływie niezerowej wariancji ciągu o skończonej liczebności n i podobnie jak wariancji nie można ich zmniejszać w żaden inny sposób niż poprzez zwiększenie liczebności n.

W praktyce te odchylenia mają bardzo stabilny charakter – w zasadzie nigdy nie są zerowe, a ich wartości w każdej odpowiednio liczebnej próbie są bliskie zależności teoretycznej, co dość czytelnie ilustruje wykr. 1. Można też zauważyć, że zbieżność

tych wartości zależy od wymiaru zmiennej losowej N i przykładowo dla $N = 1$ zawierają się one zwykle w przedziale (–50%, +50%) wartości teoretycznej, ale dla $N = 16$ – w znacznie węższym przedziale (–2%, +2%). Ta pozornie dziwna właściwość ma wyjaśnienie analityczne – prawdopodobieństwo równowagi, rozumianej jako prawdopodobieństwo wystąpienia w n-elementowej próbie dokładnie $2^{-N}\ n/N$ realizacji dowolnej N-wymiarowej zmiennej losowej, dane jest zależnością (Bobrowski, 2002; Leśniewicz, 2009):

$$P_{\mathrm{R}}(n, N) = \sqrt{\frac{2^{N-1}\,N}{(1-1/2^{N})\pi\, n}}, \tag{3}$$

z której wynika, że prawdopodobieństwo takiej równowagi jest znacznie większe dla zmiennych losowych o większych wymiarach i np. dla próby o liczebności $n = 1\,048\,576$ elementów $P_{\mathrm{R}}(n,\ N = 16) = 0{,}4$, a $P_{\mathrm{R}}(n,\ N = 1) = 0{,}0024$. Z porównania tych prawdopodobieństw nie wynikają jednak wartości powyższych przedziałów i taka interpretacja ma charakter tylko jakościowy.

Wyniki analiz wartości tego rodzaju odchyleń, zwłaszcza w przypadku wielowymiarowych zmiennych losowych, są niestety mało czytelne. Po prostu jeśli N jest duże, to niewiele mówi stwierdzenie, że w danej próbie wszystkie z 2^N wartości $\sigma(X_1, ..., X_N)$ czy $\alpha(X_1, ..., X_N)_{\text{śr}}$ są dla każdej zmiennej losowej $X_1, ..., X_N$ nawet bardzo zbliżone.

Okazuje się, że zależności te przekładają się również na entropię informacyjną Shannona, która po uwzględnieniu wariancji $V(X_1, ..., X_N)$ w funkcji liczebności n przyjmuje dla ciągu doskonale losowego postać zilustrowaną na wykr. 2, daną jako entropia oczekiwana (Leśniewicz, 2009):

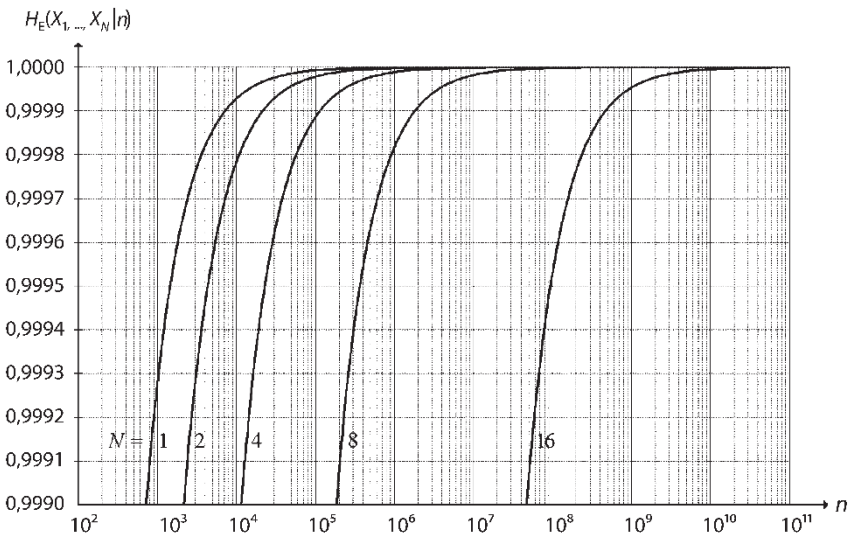
$$H_E(X_1, \ldots, X_N|n) = -\frac{1}{N}\sum_{i=1}^{k}(n_i N/n)\log_2(n_i N/n) = 1 - \frac{2^N - 1}{2\ln(2)\, n}, \tag{4}$$

gdzie n_i stanowi liczbę realizacji każdej ze zmiennych losowych $X_1, ..., X_N$.

Zależność i wykres można interpretować następująco. Dla doskonale losowego ciągu N-wymiarowych zmiennych losowych $X_1, ..., X_N$ o $P(X_1, ..., X_N) = 1/2^N$ definicyjna entropia ma wartość $H(X_1, ..., X_N) = 1$, ale oczekiwana entropia próby n-elementowej $H_{\mathrm{E}}(X_1, ..., X_N \mid n)$ ma charakter asymptotyczny i nie osiąga tej wartości nawet w przypadku prób o bardzo dużej liczebności. Źródło zjawiska jest to samo – nie jest to wynik potencjalnej nielosowości, tylko naturalny wpływ niezerowej wariancji ciągu o skończonej liczebności n. Patrząc na to z punktu widzenia teorii układów dynamicznych, ograniczonej liczebnie próbie ciągu można przypisać niepełne wymieszanie przestrzeni fazowej w sensie niezawierania się w niej wszystkich potencjalnych

realizacji zmiennych losowych o dowolnie dużych wymiarach N, co również w tej teorii skutkuje niepełną entropią.

Wykr. 2. Zjawisko asymptotycznego wzrostu entropii oczekiwanej w funkcji liczebności próby



Źródło: Leśniewicz (2009).

W praktyce – z powodu wpływu tego zjawiska – entropia oczekiwana i jej pomiary są mało użyteczne w przypadku analiz realizacji wielowymiarowych zmiennych losowych, nawet już dla $N = 8$. Wynika to z dwóch przyczyn: po pierwsze, dla n-elementowej próby ciągu liczba realizacji N-wymiarowych zmiennych losowych wynosi już tylko n/N, a po drugie (właściwie przede wszystkim) liczba potencjalnych realizacji (stopni swobody) zmiennych N-wymiarowych wynosi $k = 2^N$, co powoduje, że liczebność realizacji dla każdej zmiennej wynosi średnio $n_i = 2^{-N}\, n/N$. Przykładowo przy próbie o liczebności $n = 1\,048\,576$ elementów (1 Mbit) dla $N = 1$ liczebność zer i jedynek wynosi średnio po $n_i = 524\,288$ elementów, ale już dla $N = 8$ na jedną zmienną $X_1, ..., X_8$ przypada średnio tylko $n_i = 512$ realizacji. W rezultacie zbieżność entropii jest znacznie wolniejsza i wymagane są próby o bardzo dużej liczebności. Ze wzoru (4) i wykr. 2 widać, że dla tego samego poziomu zbieżności entropii dla $N = 8$ w porównaniu z $N = 1$ potrzebna byłaby próba o 256 razy większej liczebności, a więc w powyższym przykładzie $n = 268\,435\,456$ elementów (256 Mbit). Dla $N = 16$ byłaby to już próba o liczebności aż $n = 68\,719\,476\,736$ elementów (64 Gbit).

Trudno też stwierdzić, przy jakiej wartości zmierzonej entropii $H_P(X_1, ..., X_N \mid n)$ próby o danej liczebności n można taką próbę zakwalifikować jako realizację ciągu losowego, ponieważ w praktyce pomiarowej żadna próba ciągu nawet doskonale

losowego nie osiąga wartości $H_P(X_1, ..., X_N \mid n) = 1$, jako że teoretycznie nie osiąga jej przecież również entropia oczekiwana $H_E(X_1, ..., X_N \mid n)$.

Z powyższych względów entropia informacyjna Shannona jest doskonałym narzędziem analitycznym i interpretacyjnym, pozwalającym na określanie granicznych wartości liczebności elementów, przy których entropia oczekiwana osiąga wartości dowolnie bliskie $\lim_{n\to\infty} H_E(X_1, ..., X_N \mid n) = 1$. Nie ma to jednak przełożenia na praktykę pomiarową, ponieważ przy przeprowadzaniu dowodu bezpieczeństwa kryptograficznego, w którym zakłada się racjonalny czas analizy statystycznej, liczony w czasie istnienia Wszechświata, już w przypadku $N = 2$ otrzymuje się wartości bliskie $H_E(X_1, X_2 \mid n) = 1 - 10^{-32}$, ale dla zupełnie niefizycznych liczebności rzędu $n = 10^{27}$; dla wymiarów $N > 2$ są one odpowiednio większe (Leśniewicz, 2009).

4. Założenia konstrukcji obiektywnego testu ciągów doskonale losowych

Testy do weryfikacji losowości ciągów, intencjonalnie doskonale losowych, budzą liczne zastrzeżenia. Skoro ciąg doskonale losowy ma tak prosty model probabilistyczny i teorioinformacyjny, to dlaczego istniejące testy *de facto* z niego nie korzystają, a zamiast tego używają mechanizmów weryfikacji dziesiątek arbitralnie spreparowanych, specyficznych cech, których ciąg doskonale losowy nie ma, ponieważ mieć nie powinien, więc oczekiwanie takiej cechy i jej niewykrycie może stanowić o odrzuceniu hipotezy, że ciąg jest doskonale losowy, kiedy w rzeczywistości taki jest.

Część tych mechanizmów jest ewidentnie ukierunkowana na wykrywanie cech nielosowości właściwych określonej klasie ciągów pseudolosowych (L'Ecuyer i Simard, 2013); tymczasem dana cecha nielosowości może nie występować w innej klasie ciągów pseudolosowych i dany mechanizm jest wówczas nieużyteczny – na pewno w przypadku ciągu doskonale losowego. Dowodem powyższego są opisy generatorów ciągów (Jacak i in., 2021), które spełniają wymagania jednego zestawu testów, ale nie spełniają wymagań innych zestawów, co świadczy o specyficznych cechach takich ciągów, akceptowanych przez jeden zestaw i nieakceptowanych przez inny, a w konsekwencji źle świadczy zarówno o testach (dają się oszukać), jak i o samych ciągach (mają one cechy spreparowane pod specyficzne wymagania danego testu). Ciąg doskonale losowy w każdej próbie powinien spełniać każdy test, oczywiście poprawnie skonstruowany matematycznie.

Historia rozwoju metod statystycznego testowania ciągów losowych liczy sobie ponad 80 lat. Tymczasem w 2022 r. NIST, uznawany za czołową instytucję normalizującą, nagle po 12 latach wycofał swoje rekomendacje dla zestawu testów do badania losowości ciągów (NIST, 2022), pozostawiając badaczy i użytkowników w rozterce, że w tym czasie

stosowali niewiarygodne narzędzie matematyczne, a w konsekwencji uzyskane przez nich wyniki naukowo-techniczne weryfikowane tym narzędziem mogą być fałszywe.

Powyższe wątpliwości prowadzą do następujących wniosków:

1. Obiektywny test ciągu doskonale losowego musi opierać się na modelach probabilistycznym i teorioinformacyjnym (opisanych w częściach 3 i 4).
2. Wynik takiego testu powinien być binarny – próba ciągu bezwarunkowo spełnia wymagania testu i ciąg jest kwalifikowany jako losowy lub nie spełnia wymagań testu i ciąg jest kwalifikowany jako nielosowy.
3. Ciągi pseudolosowe nie są ciągami losowymi, tylko deterministycznymi, powinny być zatem źródłowo konstruowane zgodnie z prawami algebry i teorii liczb oraz mieć *a priori* dowiedzione referencyjne właściwości i parametry statystyczne wynikające z zastosowania odpowiednich funkcji algebraicznych, które *a posteriori* może i powinien sprawdzić na etapie konstrukcji ciągu jego twórca. Użytkownik takiego ciągu nie powinien już ich testować, ponieważ każda próba ciągu powinna mieć wzmiankowane referencyjne właściwości i parametry statystyczne. Kryptograficznymi przykładami generatorów ciągów pseudolosowych są funkcja *sponge*, jak również wcześniejsze konstrukcje szyfrów typu DES i AES czy funkcja skrótu SHA-3.
4. Ciągi prawdziwie i doskonale losowe są domeną generatorów sprzętowych, więc testy powinny być ukierunkowane na wykrywanie cech nielosowości właściwych ciągom wytwarzanym przez te generatory, z uwzględnieniem tego, że w praktyce:

- 90% błędów nielosowości wynika z niezrównoważenia zmiennych losowych o wymiarze $N = 1$, czyli statystyk {0} i {1}, zatem nierównowagi $P(0) \neq P(1) \neq 1/2$, i choć są to zazwyczaj minimalne różnice, to wpływają one na niezrównoważenia wszystkich wyższych zmiennych losowych o wymiarze $N \geq 2$;
- 5% błędów nielosowości wynika z niezrównoważenia zmiennych losowych o wymiarze $N = 2$, czyli statystyk par {0, 0}, {0, 1}, {1, 0} i {1, 1}, zatem nierównowagi $P(0,0) \neq P(0,1) \neq P(1,0) \neq P(1,1) \neq 1/4$, co wynika z inercyjnych właściwości źródeł losowości, np. szumów elektrycznych, i przekłada się na korelacje międzyelementowe oraz klasyczną, Markowowską zależność 1. rzędu, tzn. $P(X_N \mid X_1, ..., X_{N-1}) = P(X_N \mid X_{N-1}) \neq 1/2$;
- 5% błędów nielosowości wynika z przypadkowych właściwości i parametrów elektronicznych układów przetwarzających sygnały losowe. Wynikowo powoduje to pojawianie się niezrównoważonych zmiennych losowych o znacząco większych wymiarach $N \geq 8$, co powoduje, że np. statystyki {1} i {0}, a nawet pary {0, 0}, {0, 1}, {1, 0} i {1, 1} mogą być referencyjne, ale pojawiają się niezrównoważenia w sensie np. nieznacznego nadmiaru jednych i niedomiaru innych realizacji o wymiarze $N = 16$, jeśli dany układ przetwarzający pracuje w arytmetyce 16-bitowej;
- analizy wielowymiarowych zmiennych losowych są pracochłonne i żmudne, wyniki mało czytelne, a stosowalność entropii informacyjnej Shannona jest ograniczona

wpływem niezerowej wariancji ciągów losowych, mającej źródło w ograniczonej liczebności badanych prób ciągów.

5. Test zgodności chi-kwadrat dla ciągu doskonale losowego

Trudno byłoby przecenić użyteczność testu zgodności χ^2 w weryfikacji hipotez statystycznych, które dotyczą zjawisk polegających na porównywaniu częstości zdarzeń doświadczalnych ze spodziewanymi czy zależności między zmiennymi losowymi.

Zastosowanie statystyki testu χ^2 do porównywania częstości zdarzeń doświadczalnych ze spodziewanymi, zdefiniowanej przez Pearsona jako

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = n \sum_{i=1}^{k} \frac{(n_i/n - p_i)^2}{p_i}, \tag{5}$$

odpowiada weryfikacji hipotezy, że badana próba ciągu o liczebności n elementów, $k = 2^N$ stopniach swobody i nieznanym rozkładzie jest próbą doskonale losowego ciągu o dyskretnym rozkładzie równomiernym i prawdopodobieństwach zmiennych losowych $p_i = P(X_1, ..., X_N) = 1/2^N$.

Bezpośrednie zastosowanie takiej statystyki jest możliwe, nie wykorzystuje ono jednak pewnych ukrytych zależności właściwych ciągom doskonale losowym, które specyficznie i znacząco zmieniają właściwości i parametry uniwersalnej, definicyjnej statystyki testu χ^2.

Łatwo zauważyć, że wartość w liczniku drugiej wersji zależności, czyli statystyczna wariancja $(n_i / n - p_i)^2$, odpowiada probabilistycznej wariancji $V(X_i) = P(X_i)\,(1 - P(X_i)) / n$, która dla doskonale losowego ciągu N-wymiarowych zmiennych losowych dane jest jako $V(X_i) = 1/2^N\,(1 - 1/2^N)\,N / n = (n_i / n/N - 1/2^N)^2$. Wartość w mianowniku to oczywiście $1/2^N$, a stąd

$$\chi^2 = n / N \sum_{1=1}^{k} \frac{v(x_i)}{P_i} = n / N \sum_{i=1}^{k} \frac{1/2^N(1-1/2^N)N/n}{1/2^N}, \tag{6}$$

która po prostych przekształceniach i zsumowaniu wszystkich $k = 2^N$ składników sumy przyjmuje dla każdej N-wymiarowej zmiennej losowej właściwą jej jednowartościową postać, niezależną od liczebności próby n:

$$\chi^2(N)_{\mathrm{L}} = 2^N\,(1 - 1/2^N) = 2^N - 1 = df\,(N) - 1, \tag{7}$$

czyli wartość stopnia swobody $df\,(N)$ pomniejszoną o 1.

Są to statystyki testu χ^2 właściwe ciągom doskonale losowym, a ich wartości nie są funkcją potencjalnej nielosowości, ale naturalnego wpływu niezerowych wariancji dla wszystkich potencjalnych realizacji zmiennej losowej $X_1, ..., X_N$. Ściśle mówiąc, właściwość ta charakteryzuje nie tylko ciągi doskonale losowe, lecz także wszystkie ciągi losowe, mające wariancje w postaciach dających się uprościć względem prawdopodobieństw zgodnie z zależnością (6).

Statystyki te można opisać następująco:

1. Gdyby każda próba badanego ciągu była próbą ciągu doskonale losowego i dla każdej N-wymiarowej zmiennej losowej występowałoby $n_i = n / N\ 1/2^N$ realizacji, to $\chi^2(N) = 0$. Jednak tak nigdy nie jest i próba ciągu o $\chi^2(N) = 0$ na pewno nie jest realizacją ciągu losowego.
2. Żadna realizacja badanego ciągu doskonale losowego nie może mieć zatem wartości statystyki $\chi^2(N) << 2^N - 1$.
3. Wariancje $V(X_1, ..., X_N)$ mają zawsze niezerowe, powtarzalne i stabilne wartości, dlatego wartość $\chi^2(N)_L = 2^N - 1$ stanowi najmniejszą wartość statystyki, właściwą jednak tylko ciągowi losowemu, w rozważanym przypadku opisanemu prawdopodobieństwami $P(X_1, ..., X_N) = 1/2^N$.
4. Żadna realizacja ciągu o wartości statystyki $\chi^2(N) >> 2^N - 1$ nie może być zatem realizacją ciągu doskonale losowego.
5. Model testu spełnia założenia testu zgodności χ^2, włącznie z tym, że dla odpowiednio liczebnych prób rozkłady wszystkich N-wymiarowych zmiennych losowych są normalne.
6. Porównywanie statystyk dla różnych wymiarów N zmiennej losowej może być mało czytelne, ponieważ dla większych wymiarów N przyjmują one znacznie większe wartości, więc wskazana jest ich normalizacja względem wartości $df(N) - 1 = 2^N - 1$ i zastosowanie zredukowanej statystyki testu χ^2, tzn.

$$\chi^2(N)_{\mathrm{LN}} = \frac{2^N - 1}{df(N) - 1} = 1, \tag{8}$$

gdzie wartość $\chi^2(N)_{LN} = 1$ stanowi ponadto jej wartość oczekiwaną $E(N)$.

W tabl. 1 zamieszczono oczekiwane wartości statystyk obliczone dla różnych wymiarów N. Dla porównania w ostatniej kolumnie zamieszczono wartości zredukowanych standardowych statystyk $\chi^2(N)_N$ dla typowo przyjmowanych poziomów istotności (od 0,05 do 0,001). Jak widać, statystyki $\chi^2(N)_{LN}$ dla ciągów doskonale losowych mają wyraźnie mniejsze wartości w porównaniu ze statystykami $\chi^2(N)_N$ w standardowej postaci. Łatwo jednak dostrzec, że – zwłaszcza dla zmiennych losowych o dużych wymiarach N

– statystyki $\chi^2(N)$ w wersji zredukowanej też wyraźnie dążą do wartości $\chi^2(N)_{\mathrm{N}} = 1$[3]. Korzystając od razu z tej zależności, można zauważyć, że standardowe statystyki $\chi^2(N)_{\mathrm{N}}$ dla różnych poziomów istotności operują przedziałami o wartościach tylko dodatnich i wyraźnie szerszych od σ_{N} – z natury rzeczy tym szerszych, im poziom istotności jest mniejszy. Statystyka $\chi^2(N)_{\mathrm{LN}}$ reprezentuje więc najwyższy poziom losowości w sensie najniższej granicznej wartości statystyki $\chi^2(N)_{\mathrm{N}}$. Nie zmienia to istoty testu χ^2, a tylko stawia wyższe wymagania jego statystyce, które może spełnić ciąg doskonale losowy.

Tabl. 1. Statystyki χ^2 dla *N*-wymiarowych zmiennych losowych

N	Liczba stopni swobody $k = 2^N$	Statystyka $\chi^2(N)_{\mathrm{L}} = 2^N - 1$	Standardowa statystyka $\chi^2(N)_{\mathrm{N}} = \chi^2 / (df(N) - 1)$ dla poziomów istotności od 0,05 do 0,001[a]
1	2	1	4÷10 / 1 = 4÷10 = 1 + 3,0÷9,00 $\sigma_{N=1}$
2	4	3	8÷16 / 3 = 2,67÷5,33 = 1 + 2,4÷6,10 $\sigma_{N=2}$
3	8	7	14÷24 / 7 = 2,00÷3,43 = 1 + 2,0÷4,86 $\sigma_{N=3}$
4	16	15	25÷38 / 15 = 1,67÷2,53 = 1 + 1,9÷4,32 $\sigma_{N=4}$
5	32	31	45÷61 / 31 = 1,45÷1,97 = 1 + 1,8÷3,88 $\sigma_{N=5}$
6	64	63	83÷103 / 63 = 1,32÷1,63 = 1 + 1,8÷3,56 $\sigma_{N=6}$
7	128	127	154÷182 / 127 = 1,21÷1,43 = 1 + 1,7÷3,44 $\sigma_{N=7}$
8	256	255	293÷330 / 255 = 1,15÷1,29 = 1 + 1,7÷3,28 $\sigma_{N=8}$
9	512	511	565÷613 / 511 = 1,11÷1,20 = 1 + 1,7÷3,21 $\sigma_{N=9}$
10	1 024	1 023	1 100÷1 170 / 1 023 = 1,075÷1,14 = 1 + 1,7÷3,18 $\sigma_{N=10}$

a Ostatni wyraz to 1 + przedział × σ_N = 1 + przedział × $2^{-(N-1)/2}$.

Uwaga. Statystyka $\chi^2(N)_{\mathrm{LN}} = 1$.

Źródło: opracowanie własne.

Pomiarowa statystyka $\chi^2(n, N)_{\mathrm{P}}$ próby ciągu doskonale losowego o liczebności *n* dana jest jako

$$\chi^2(n,N)_p = \frac{2^N}{2^N-1}\frac{n}{N}\sum_{i=1}^{k}(n_i N / n - 1 / 2^N)^2, \tag{9}$$

gdzie n_i stanowi liczbę realizacji każdej z $k = 2^N$ zmiennych losowych $X_1, ..., X_N$.

Taka postać statystyki ma trzy oczywiste zalety:

- statystyka $\chi^2(n, N)_{\mathrm{P}}$ zachowuje wszystkie właściwości i parametry definicyjnej statystyki $\chi^2(N)$;
- wyniki dla każdej z *N*-wymiarowych zmiennych losowych są weryfikowane pod względem zgodności z jedną znormalizowaną wartością $\chi^2(N)_{\mathrm{LN}} = 1$, co bardzo ułatwia analizy porównawcze;

[3] W części 6 wyznaczane są wariancja i odchylenie standardowe tego rozkładu dla zmiennych losowych o dowolnym wymiarze *N* w postaci $\sigma_N = 2^{-(N-1)/2}$.

- wszystkie składniki sumy są dodatnie, więc statystyka jest rosnąca i nie wykazuje właściwości wzajemnego kompensowania się realizacji częściej i rzadziej występujących. Oznacza to, że dla ciągu, którego statystyki byłyby zmienne w funkcji liczebności n lub wymiaru N, statystyka będzie zbierała i sumowała wszystkie odchylenia i na końcu okaże je jako zbiorczy parametr nielosowości.

Warto jeszcze rozważyć, jakie liczebności prób ciągów powinny być używane, aby statystyka była wiarygodna, tzn. zbieżna. W tabl. 2 zamieszczono obliczone liczebności dla większych wymiarów N, ponieważ to właśnie one wymagają znacznie liczebniejszych prób. Przyjęto następującą konwencję wartości: 1000 MB = 1 GB = $=1000 \times 8 \times 1\,048\,576$ elementów.

Tabl. 2. Liczebność prób niezbędna do spełnienia wymagań statystyki $\chi^2(n, N)_P$

N	2^N = liczba przedziałów	$n_i > 30$	$n_i > 100$	$\Delta n_i / n_i < 1/8$	$\Delta n_i / n_i < 1/16$
		w MB			
16	65 536	3,93	13,1	5,34	21,4
18	262 144	17,7	59	24	96
20	1 048 576	78,6	262	107	428
22	4 194 304	345	1 150	470	1 880
24	16 777 216	1 510	5 030	2 050	8 200
26	67 108 864	6 540	21 800	8 890	35 600
28	268 435 456	28 200	94 000	38 300	153 000
30	1 073 741 824	121 000	403 000	164 000	656 000

Źródło: opracowanie własne.

Obliczono niezbędne liczebności dla czterech warunków – średnich liczebności realizacji $n_i > 30$ i $n_i > 100$ oraz liczebności zapewniających dokładność każdej i-tej realizacji lepszą od $\Delta n_i / n_i < 1/8$ i $\Delta n_i / n_i < 1/16$. Jak się okazuje, dla danego N mają one zbliżone wartości. Weryfikacja ich rzeczywistych, tzn. niezbędnych, liczebności została przeprowadzona doświadczalnie. Wynika z niej, że górną granicą badanych wymiarów może być $N = 20$, ponieważ generowanie i pomiary prób o liczebności wielu gigabajtów wymagałyby specjalnych generatorów o wyjątkowo dużych wydajnościach, a czas obliczeń na typowych komputerach byłby dość długi[4].

6. Test zgodności chi-kwadrat dla rzeczywistego ciągu doskonale losowego

Przedstawiona w części 5 statystyka testu $\chi^2(N)_{LN}$ dla ciągów doskonale losowych może wydawać się zbyt prosta i – z pozornie niejasnych powodów – niezależna od

[4] Te problemy opisano szerzej w części 10.

liczebności próby. Te wątpliwości najlepiej wyjaśnić dzięki doświadczalnemu wykonaniu pomiarowej weryfikacji wyprowadzonych analitycznie teoretycznych wartości $\chi^2(N)_{LN}$. Weryfikacja powinna polegać na pomiarach statystyk $\chi^2(n, N)_P$ dla prób ciągów, które przy zastosowaniu innych metod zostały uznane za doskonale losowe, i porównaniu ich z wartościami $\chi^2(N)_{LN}$.

Wygenerowano 10 prób ciągów o liczebności n = 1000 MB. Obserwowano bieżące wartości $\chi^2(n, N)_P$ kolejnych podciągów o liczebności 1 MB i wartości dla sumy tych podciągów. Wyniki uśrednionych pomiarów w sensie wyznaczenia dla każdej liczebności statystyki testu $\chi^2(n, N)_P$ zawiera tabl. 3. W kolumnie „rozmiar próby" uwzględniono liczebności sumy podciągów 1 MB, przy których zauważono wyraźne zmiany wartości statystyki. Statystyki wyznaczono dla realizacji zmiennych losowych o wymiarach N = 1, 2, 4, 8, 16 i 20.

Tabl. 3. Statystyki $\chi^2(n, N)_P$ rzeczywistego ciągu doskonale losowego

Rozmiar próby w MB	$\chi^2(N=1)_P$	$\chi^2(N=2)_P$	$\chi^2(N=4)_P$	$\chi^2(N=8)_P$	$\chi^2(N=16)_P$	$\chi^2(N=20)_P$
1	0÷4	0÷4	0,6÷1,4	0,8÷1,3	0,99÷1,01	0,998÷1,002
50	0÷4	0÷4	0,7÷1,3	0,9÷1,1	0,99÷1,01	0,999÷1,001
500	<1	<1	0,7÷1,3	0,9÷1,1	0,995÷1,005	0,999÷1,001
1 000	<1	<1	0,8÷1,2	0,95÷1,05	0,995÷1,005	0,999÷1,001
$\pm\sigma = \pm 2^{-(N-1)/2}$	±1,00	±0,71	±0,35	±0,088	±0,0055	±0,00135
$\pm 3\sigma = \pm 3\times 2^{-(N-1)/2}$	±3,00	±2,12	±1,06	±0,265	±0,0166	±0,00405

Źródło: opracowanie własne.

Z pomiarów wynika, że:

- dla dowolnej N-wymiarowej zmiennej losowej i dowolnej liczebności próby n wyniki obliczeń wartości statystyki $\chi^2(n, N)_P$ są ściśle zbieżne z wyznaczonymi teoretycznie wartościami statystyki $\chi^2(N)_{LN} = 1$;
- przedziały zbieżności statystyk wokół wartości $\chi^2(n, N)_{LN}$ są tym mniejsze, im liczebniejsza była badana próba. W praktyce dostatecznie zbieżne wyniki otrzymuje się już dla prób o liczebności 1 MB, a dwukrotne zawężenie przedziału zbieżności do stabilnej wartości dla dowolnego wymiaru N wymaga zwiększenia liczebności próby do 1 GB, zbieżność ta jest zatem bardzo wolna. Można szacować, że dla bardzo dużych liczebności rzędu 1 GB przedziały zbieżności dla różnych wymiarów N dane są w przybliżeniu jako $\Delta\chi^2(n, N)_P / \chi^2(n, N)_P \approx \pm 2^{-N/2}$;
- szerokość przedziału zbieżności można jednak łatwo określić teoretycznie. $\chi^2(N)_{LN}$ pozostaje statystyką $\chi^2(N)$, której wartość oczekiwana dla ciągu doskonale losowego $E(N) = k = 2^N$, a wariancja $V(N) = \sigma^2 = 2k = 2^{N+1}$, dlatego stosunek odchylenia standardowego do wartości oczekiwanej $\sigma_{LN} / E(N)_{LN} = 2^{-(N-1)/2}$. Dla $E(N)_{LN} = 1$ wartość $\sigma^2_{LN} = 2^{-N+1}$ można zatem nazwać *unormowaną wariancją statystyki* $\chi^2(N)_{LN}$ dla ciągów doskonale losowych. Przy uwzględnieniu, że dla dużych wymiarów N i dużych

liczebności n statystyka χ^2 dąży do rozkładu normalnego, wskazane byłoby jednak przyjęcie szerokości przedziałów zbieżności odpowiadających kryterium $\pm 3\sigma$, zawierających $p \geq 0{,}997$ wyników, i wtedy przedziały zbieżności będą dane jako $\pm 3\sigma_{\mathrm{LN}} / E(N)_{\mathrm{LN}} = \pm 3\sqrt{2} \times 2^{-N/2} \approx \pm 4{,}24 \times 2^{-N/2}$. To przybliżenie jest jednak dokładne tylko dla wymiarów $N \geq 5$, ponieważ dla mniejszych N lewe ogony statystyk $\chi^2(N)$ nie przypominają ogonów rozkładu normalnego. Biorąc to pod uwagę, należy założyć zerową graniczną wariancję $\chi^2(N)$ dla lewej strony statystyki;

- w przypadku prób o zbyt małej liczebności wyniki dla małych wymiarów zmiennych losowych N są niezbyt wiarygodne, bo rozproszone, jednak w każdym przypadku wokół oczekiwanych wartości $\chi^2(N)_{\mathrm{LN}} = 1$;
- powyższe spostrzeżenia ściśle korespondują również z wynikami przedstawionymi w tabl. 1.

Wyniki pomiarów statystycznych prób ciągów doskonale losowych o dowolnej liczebności n potwierdzają zatem w pełni poprawność analiz probabilistycznych, jednak zbieżność do wartości $\chi^2(N)_{\mathrm{LN}} = 1$, zwłaszcza dla zmiennych losowych o małych wymiarach $N = 1$ i 2, następuje dopiero w przypadku prób o bardzo dużej liczebności, sięgającej 1 GB. Powyższe spostrzeżenia są zatem zgodne z kolejnymi, opisanymi w części 3, właściwościami ciągów doskonale losowych – modułem średniego odchylenia od wartości oczekiwanej $\alpha(X_1, ..., X_N)_{\mathrm{śr}}$ i prawdopodobieństwem równowagi dowolnej N-wymiarowej zmiennej losowej $P_{\mathrm{R}}(n, N)$.

7. Test zgodności chi-kwadrat dla rzeczywistego ciągu niedoskonale losowego

Wyniki badań prób ciągów doskonale losowych w pełni potwierdziły matematyczną poprawność i aplikacyjną użyteczność w sensie ścisłej zbieżności statystyki $\chi^2(n, N)_{\mathrm{P}}$ do $\chi^2(N)_{\mathrm{LN}}$. Interesujące byłoby teraz sprawdzenie, choćby dla pokazania zdolności identyfikacji potencjalnych nielosowości, jaka jest wrażliwość tej statystyki na potencjalne nielosowości badanego ciągu.

Najprostszym sposobem, i zarazem najlepiej odzwierciedlającym sprzętowe uwarunkowania generowania ciągów losowych, jest użycie ciągu z elementarnego sprzętowego generatora, który źródłowo wytwarza ciągi prawdziwie losowe, ale bardzo dalekie od doskonałości (Leśniewicz, 2009). Wyraża się to nierównowagą (ang. *bias*), opisaną jako $P(0) = 1/2 - s$ oraz $P(1) = 1 - P(0) = 1/2 + s$, gdzie typowa wartość $s = 0{,}005$, oraz korelacjami międzyelementowymi, modelowanymi właściwością Markowa 1. rzędu, opisaną jako $P(X_N \mid X_{N-1}) \neq 1/2$, co sprowadza prawdopodobieństwa par elementów {0} i {1} do wartości $P(0, 0) = 1/4 - s + 1/4\ K$, $P(0, 1) = 1/4 - 1/4\ K$, $P(1, 0) = 1/4 - 1/4\ K$ i $P(1, 1) = 1/4 + s + 1/4\ K$, gdzie $K = P(0, 0) + P(1, 1) - P(0, 1) - P(1, 0) - (2s)^2$ o typowej wartości $K = 0{,}005$. Wartości parametrów s i K wynikają

z elektrycznych właściwości elementów, stanowiących źródła losowości, i charakteryzują się dobrą stabilnością w czasie i zmiennych warunkach klimatyczno-mechanicznych, ale są różne dla każdego egzemplarza generatora, a rozrzut ich wartości sięga zwykle ±50%.

Wyznaczenie oczekiwanych wartości statystyk $\chi^2(n, N)$ dla różnych n i N przy powyższych założeniach jest możliwe, ale bardzo pracochłonne i – zwłaszcza dla większych wymiarów N – zależne od relacji wartości s i K, więc prostsze jest doświadczalne sprawdzenie wrażliwości statystyki $\chi^2(n, N)_P$ na nielosowości badanego ciągu. Badanie przeprowadzono na jednej próbie ciągu o liczebności 1 GB. Zarejestrowano statystyki dla sum kolejnych podciągów o liczebności 1 MB. Wybrane wyniki ilustruje tabl. 4.

Tabl. 4. Statystyki $\chi^2(n, N)_P$ rzeczywistego ciągu niedoskonale losowego z $s = 0{,}005$ i $K = 0{,}005$

Rozmiar próby w MB	$\chi^2(N=1)_P$	$\chi^2(N=2)_P$	$\chi^2(N=4)_P$	$\chi^2(N=8)_P$	$\chi^2(N=16)_P$	$\chi^2(N=20)_P$
1	1 200	750	53	10	1,039	1,0001
10	11 300	7 100	460	85	1,325	1,0070
100	111 000	70 000	4 710	830	4,220	1,0680
1 000	1 160 000	720 000	46 000	8 500	34,20	1,6600

Źródło: opracowanie własne.

Z danych zawartych w tabl. 4 wynika oczywisty wniosek, że statystyka $\chi^2(n, N)_P$ jest niezależna od liczebności próby n tylko dla ciągów doskonale losowych, natomiast dla niedoskonale losowych wykazuje tym większą wrażliwość, im badana próba ciągu jest liczebniejsza. Stanowi to bardzo cenną właściwość pomiarową, ponieważ gdyby statystyka $\chi^2(n, N)_P$ przekraczała nieznacznie, ale zauważalnie wartość $\chi^2(N)_{LN} = 1$, można byłoby dokonać badania znacznie liczebniejszej próby. Jeśli wartość $\chi^2(n, N)_P$ odpowiednio wzrosłaby, to stanowiłoby to dowód na występowanie ewidentnej nielosowości.

Pewne wątpliwości może budzić progresja wartości statystyki $\chi^2(n, N)_P$ w zależności od wymiaru N; dla $N = 1, 2, 4$ i 8 jest ona w przybliżeniu liniowa w funkcji liczebności próby n, ale dla większych wymiarów $N = 16$ i 20 wydaje się, że wrażliwość rozpatrywanej statystyki jest znacznie mniejsza. Wynika to po pierwsze z tego, że w badanym typie nielosowości, tzn. nierównowagi i korelacji międzyelementowych, są one – z natury rzeczy – szybko identyfikowane przez statystykę $\chi^2(n, N)_P$ dla $N = 1$ (nierównowaga) i $N = 2$ (korelacje), a dla zmiennych losowych o większych wymiarach N rozkładają się one równomiernie w dłuższych realizacjach. Po drugie, należy pamiętać, że w przypadku ciągu doskonale losowego rozproszenie wyników np. dla $\chi^2(n, N = 16)_P$ i liczebności do 100 MB wynosi od 0,99 do 1,01, zatem w przypadku ciągu nielosowego już dla 1 MB widać wyraźną, dodatkową nielosowość, która dla 10 MB rośnie ponad 30-krotnie, a dla 100 MB czy 1 GB przyjmuje nawet wartości znacznie większe od 1.

W takiej sytuacji należy powtórzyć badanie dla co najmniej 10-krotnie większej próby i stwierdzić, że wartości $\chi^2(n, N)_P$ dla $N = 1, 2, 4$ i 8 wzrosły w przybliżeniu kolejne 10 razy, a dla $N = 16$ i 20 – wyraźnie przekroczyły wartość 1. W całości taki raport stanowi o ciągu bardzo dalekim od losowości doskonałej już przy próbie o liczebności 1 MB.

Aby sprawdzić wrażliwość testu na znacznie mniejsze nielosowości, wygenerowano próbę ciągu o nierównowadze $s = 0{,}0002$ i korelacjach $K = 0{,}00002$. Wybrane wyniki zawiera tabl. 5.

Tabl. 5. Statystyki $\chi^2(n, N)_P$ rzeczywistego ciągu niedoskonale losowego z $s = 0{,}0002$ i $K = 0{,}00002$

Rozmiar próby w MB	$\chi^2(N=1)_P$	$\chi^2(N=2)_P$	$\chi^2(N=4)_P$	$\chi^2(N=8)_P$	$\chi^2(N=16)_P$	$\chi^2(N=20)_P$
1	1,11	0,91	0,96	1,05	1,0025	1,0005
10	7,46	2,52	1,55	1,13	1,0040	1,0004
100	158,00	53,00	11,10	1,30	1,0090	1,0009
1 000	1 630,00	540,00	108,00	4,76	1,0260	1,0015

Źródło: opracowanie własne.

Wnioski z wyników pomiarów są takie same jak w poprzednim przypadku – ciągi zostaną bezwarunkowo zakwalifikowane jako niedoskonale losowe, ale dopiero dla prób o liczebności powyżej 10 MB. Można wspomnieć, że większość używanych obecnie testów zakwalifikowałaby taki ciąg jako doskonale losowy. Wypływa z tego wniosek, że badanie zbyt małych prób ciągów uniemożliwia wykrycie potencjalnych nielosowości. Można jednak zauważyć, że aplikacyjne użycie próby ciągu o liczebności uniemożliwiającej wykrycie takich nielosowości można byłoby uznać za dopuszczalne, jeśli byłoby jednorazowe. Taki kompromis jest jednak nie do przyjęcia w przypadku kolejnych prób, co wyraźnie pokazują kolejne wiersze tabl. 5.

Aby ocenić graniczne możliwości testu, warto sprawdzić jego wrażliwość przy jeszcze mniejszych nielosowościach – nierównowadze $s = 0{,}00001$ i korelacjach $K = 0{,}000007$. Wyniki przedstawiono w tabl. 6.

Tabl. 6. Statystyki $\chi^2(n, N)_P$ rzeczywistego ciągu niedoskonale losowego z $s = 0{,}00001$ i $K = 0{,}000007$

Rozmiar próby w MB	$\chi^2(N=1)_P$	$\chi^2(N=2)_P$	$\chi^2(N=4)_P$	$\chi^2(N=8)_P$	$\chi^2(N=16)_P$	$\chi^2(N=20)_P$
1	0,57	0,31	0,94	1,070	1,0092	1,0016
10	0,15	0,73	1,03	0,960	1,0071	1,0021
100	0,83	0,35	1,08	0,940	0,9989	0,9989
1 000	1,35	0,88	1,09	0,960	1,0008	1,0008
2 000	7,91	2,78	1,13	1,006	0,9981	0,9991
3 000	11,40	4,08	1,29	1,037	0,9944	0,9995
4 000	15,90	5,50	1,33	1,070	0,9953	0,9994

Źródło: opracowanie własne.

Dla prób o liczebności do 1 GB wszystkie znane testy, nawet $\chi^2(n, N)_P$, uznałyby taki ciąg za doskonale losowy. Można to łatwo wyjaśnić, porównując wartość nierównowagi $s = 0{,}00001$ z wartością modułu średniego odchylenia od wartości oczekiwanej dla zmiennej losowej o wymiarze $N = 1$, danej jako $\alpha(X_1) = (2\pi\, n)^{-1/2}$, a dla liczebności $n = 8 \times 1000 \times 1\,048\,576$ elementów równej $\alpha(X_1) = 0{,}0000043$, zatem około dwukrotnie mniejszej od $s = 0{,}00001$. Jak widać, $\chi^2(n, N)_P$ bardzo rygorystycznie reaguje na objawy nierównowagi i wszystkie inne odchylenia od naturalnej losowości ciągu. Ponownie można też zauważyć, że w badanym typie nielosowości statystyka $\chi^2(n, N)_P$ dla $N = 1$ i 2 wykazuje największą wrażliwość, natomiast dla $N = 8$, 16 i 20 – z opisanych już powodów – znacznie mniejszą. Wynika z tego ponownie, że potencjalne mniejsze nielosowości ciągu mogą się ujawnić dopiero w przypadku prób o odpowiednio większej liczebności. Problem z ich wykrywaniem jest jednak taki, że generator ciągu musi mieć niezbędną wydajność, a samo testowanie wymaga czasu[5].

8. Test zgodności chi-kwadrat dla ciągów pseudolosowych

Kolejnym sprawdzeniem użyteczności statystyki $\chi^2(n, N)_P$ były badania ciągów pseudolosowych. Pominięto ciągi uznawane za słabe, natomiast dokładnie zbadano powszechnie stosowany ciąg Mersenne Twister (Sulewski, 2019). Jest to ciąg generowany źródłowo z dyskretnym rozkładem równomiernym i uznawany za jeden z najlepszych ciągów pod względem szybkości generowania i dobrych statystyk, zwłaszcza imponująco długiego okresu, o dowiedzionej wartości $L = 2^{19937} - 1 \approx 4{,}3 \times 10^{6001}$.

Wygenerowano trzy próby ciągów o liczebności $n = 1$ GB, wykorzystując 32-bitową wersję algorytmu, oznaczoną jako MT19937. Każda próba została wygenerowana z użyciem innych warunków początkowych w postaci tablicy inicjującej (Sulewski, 2019).

W ten sam sposób zbadano statystyki $\chi^2(n, N)_P$ funkcji kryptograficznych *sponge*, AES, DES i 3DES oraz SHA-3, konfigurując ich wejścia do pracy w najprostszych trybach licznikowych, przy stałych wektorach inicjujących.

Metodyka i wyniki badań wszystkich ciągów pseudolosowych były identyczne jak dla ciągów prawdziwie i doskonale losowych i nie odbiegały od wartości przedstawionych w tabl. 3. Również w tym przypadku badano tylko statystyki, nie wnikając w algorytmy generowania w sensie badań kryptoanalitycznych. Można jednak wspomnieć, że użycie programowych, tzn. komputerowych generatorów ciągów pseudolosowych, nawet tych poprawnie skonstruowanych, nie musi prowadzić do uzyskania ciągów o domyślnie referencyjnych właściwościach i parametrach statystycznych.

[5] Temu zagadnieniu jest poświęcona część 10.

W czasie badań stwierdzono bowiem, że:

- nie wszystkie kody źródłowe generatorów tych ciągów mają poprawne implementacje;
- każdy z generatorów musi być poprawnie zainicjowany w sensie trybu pracy i odpowiednio wybranych warunków początkowych. O ile w przypadku funkcji kryptograficznych sprawa jest dość prosta, o tyle właściwe przeprowadzenie procedury inicjującej tablicę do poprawnego wygenerowania ciągu Mersenne Twister jest krytyczne, ponieważ w przypadku błędnej inicjacji otrzymuje się ciągi wyglądające jak losowe, ale statystycznie bardzo dalekie od losowości doskonałej;
- testowanie wszelkich pseudolosowych ciągów z generatorów programowych, w tym również funkcji kryptograficznych, ma głęboki sens kontrolny – obowiązkowo po każdej kompilacji programu na etapie prototypowania – i profilaktyczny, nawet na niezbyt liczebnej próbie, np. 1 MB, przed każdym operacyjnym użyciem komputera czy innego układu generowania takiego ciągu. Należy bowiem pamiętać, że nawet poprawna wersja programowa nie będzie prawidłowo działała w układzie sprzętowym, który ulegnie uszkodzeniu, zakłóceniu itp.

W przypadku wystąpienia opisanych wyżej błędów implementacyjnych i aplikacyjnych ciągi pseudolosowe zwykle będą miały skrajnie złe statystyki, które można jednak szybko identyfikować poprzez profilaktyczne testowanie ich przed użyciem.

9. Powiązanie testu zgodności chi-kwadrat z entropią informacyjną Shannona

Wracając do prostoty i niezależności od liczebności próby, charakteryzujących wyznaczoną statystykę $\chi^2(N)_{LN}$, a w konsekwencji $\chi^2(n, N)_P$, można teraz skonfrontować je z entropią informacyjną Shannona, która dla ciągów doskonale losowych również korzysta z analizy i obliczeń wariancji ciągu. Jeśli porównamy statystyki $\chi^2(n, N)_P$, entropii oczekiwanej $H_E(X_1, ..., X_N \mid n)$ i zmierzonej entropii próby $H_P(X_1, ..., X_N \mid n)$ dla ciągów doskonale losowych, tzn.

$$x^2(n,N)_p = \frac{2^N}{2^N-1}\frac{n}{N}\sum_{i=1}^{k}(n_iN / n - 1 / 2^N)^2, \tag{10}$$

$$H_E(X_1, \ldots, X_N | n) = 1 - \frac{2^N - 1}{2\ln(2)\ n} \tag{11}$$

i

$$H_E(X_1, \ldots, X_N|n) = 1 - \frac{2^{N-1}}{N\ ln(2)}\sum_{i=1}^{k}(n_iN / n - 1 / 2^N)^2, \tag{12}$$

to łatwo wykażemy, że wiąże je zależność

$$H_P(X_1, \ldots, X_N \mid n) = 1 - \frac{2^N - 1}{2\ln(2)\ n}\chi^2(n, N)_P = \\ = 1 - [1 - H_E(X_1, \ldots, X_N \mid n)]\ \chi^2(n, N)_P, \tag{13}$$

która świadczy o tym, że dla entropii próby $H_P(X_1, \ldots, X_N \mid n)$ zmienna wartość $1 - H_E(X_1, \ldots, X_N \mid n)$ zawiera się w przedziale zbieżności statystyki $\chi^2(n, N)_P$, wynikającym ze statystyki $\chi^2(N)_{LN}$ i jej unormowanej wariancji $\sigma^2_{LN} = 2^{-N+1}$.

Zależność między statystyką $\chi^2(n, N)_P$ a entropiami $H_P(X_1, \ldots, X_N \mid n)$ i $H_E(X_1, \ldots, X_N \mid n)$ można przedstawić jeszcze prościej w sensie praktycznej interpretacji. W części 3 wzmiankowano, że wykorzystanie zmierzonej wartości entropii próby $H_P(X_1, \ldots, X_N \mid n)$ nie jest wsparte konkretnym praktycznym kryterium, wskazującym dla danej próby, jaka wartość mniejsza od oczekiwanej entropii $H_E(X_1, \ldots, X_N \mid n)$ może być akceptowana w sensie kwalifikacji danej próby jako realizacji ciągu doskonale losowego. Po prostych przekształceniach zależności $H_P(X_1, \ldots, X_N \mid n) = 1 - \{1 - H_E(X_1, \ldots, X_N \mid n)\}\ \chi^2(n, N)_P$ można napisać, że

$$\chi^2(n, N)_P = \frac{1 - H_P(X_1, \ldots, X_N \mid n)}{1 - H_E(X_1, \ldots, X_N \mid n)}, \tag{14}$$

i ocenić, jaka zależność wiąże stosunek wartości $1 - H_P(X_1, \ldots, X_N \mid n)$ i $1 - H_E(X_1, \ldots, X_N \mid n)$, czyli różnic między wartością informacyjnej entropii Shannona $H(X_1, \ldots, X_N) = 1$ dla ciągu doskonale losowego a mierzoną entropią próby ciągu $H_P(X_1, \ldots, X_N \mid n)$ i jej entropią oczekiwaną $H_E(X_1, \ldots, X_N \mid n)$. Statystyka $\chi^2(n, N)_P$ stanowi zatem niezależną od liczebności próby n miarę zbieżności mierzonej entropii próby z entropią oczekiwaną.

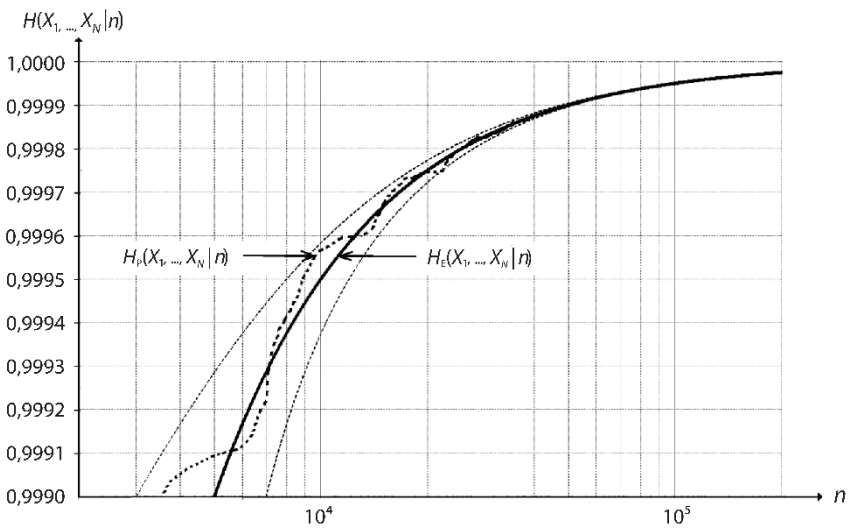
Można dodać, że pozornie prostsza miara stosunku wartości obu entropii, tzn. $H_P(X_1, \ldots, X_N \mid n) / H_E(X_1, \ldots, X_N \mid n)$, daje w rezultacie czytelną, ale niezbyt użyteczną postać, zależną ponadto od liczebności próby n:

$$\frac{H_P(X_1, \ldots, X_N \mid n)}{H_E(X_1, \ldots, X_N \mid n)} = \frac{2\ln(2)\ n - \chi^2(n, N)_P(2^N - 1)}{2\ln(2)\ n - (2^N - 1)}. \tag{15}$$

Jak widać, dla dużych liczebności prób n i nawet dużych wymiarów N miara ta przyjmuje wartości bardzo bliskie 1, co uniemożliwia identyfikację potencjalnej nielosowości ciągu. Aby ją wykryć, $\chi^2(n, N)_P$ musiałaby mieć wartości, jakie były dane w pierwszym przykładzie z części 7 (tabl. 4), ale nawet wtedy miara dla $N = 1$ przyjmowałaby wartości bliskie 0,9999, a dla większych N identyfikacja byłaby jeszcze mniej czytelna.

Ilustracją powyższych rozważań są przedstawione na wykr. 3 przykładowe realizacje entropii próby ciągu $H_P(X_1, ..., X_N \mid n)$, oznaczone linią przerywaną, na tle jej entropii oczekiwanej $H_E(X_1, ..., X_N \mid n)$, oznaczonej linią ciągłą. Widać na nim, że wartości $1 - H_P(X_1, ..., X_N \mid n)$ i $1 - H_E(X_1, ..., X_N \mid n)$ ciągle maleją, ale ich stosunek pozostaje stały i równy $\chi^2(N)_{LN}$ w zakresie unormowanej wariancji $\sigma^2_{LN} = 2^{-N+1}$.

Wykr. 3. Entropia próby $H_P(X_1, ..., X_N \mid n)$ na tle entropii oczekiwanej $H_E(X_1, ..., X_N \mid n)$



Źródło: opracowanie własne.

Interpretacja powyższych wyników jest następująca:

1. Statystyki $\chi^2(N)_{LN}$ i $\chi^2(n, N)_P$ zostały wyprowadzone z ogólnych przesłanek statystyki matematycznej i źródłowo nie opierały się na pojęciu entropii informacyjnej Shannona.
2. Entropia próby ciągu losowego $H_P(X_1, ..., X_N \mid n)$ była analizowana, mierzona i estymowana (Leśniewicz, 2009), ale nie znano dotąd analitycznej postaci rozproszenia wyników pomiarów $H_P(X_1, ..., X_N \mid n)$ względem oczekiwanej wartości $H_E(X_1, ... , X_N \mid n)$.
3. Okazuje się, że obie entropie wiąże statystyka $\chi^2(n, N)_P$, która dla ciągu doskonale losowego nie zależy od liczebności próby n.
4. Interpretacja statystyki $\chi^2(n, N)_P$ jest w tym przypadku prosta – im wartości $H_P(X_1, ..., X_N \mid n)$ i $H_E(X_1, ..., X_N \mid n)$ są sobie bliższe, tym zbieżność statystyk wokół wartości $\chi^2(n, N)_{LN} = 1$ jest większa. Nie oznacza to jednak, że $H_P(X_1, ..., X_N \mid n)$ i $H_E(X_1, ..., X_N \mid n)$ powinny być sobie równe – w przypadku ciągu doskonale losowego powinny różnić się w zakresie określonym wariancją σ^2_{LN} statystyki $\chi^2(N)_{LN}$.

5. Statystyki $\chi^2(N)_{\mathrm{LN}}$ i $\chi^2(n, N)_{\mathrm{P}}$ mogą zatem być uznane za samodzielne statystyczne miary doskonałości ciągów losowych i w taki sposób wykorzystywane, ale z punktu widzenia teorii informacji mają dużo głębsze znaczenie. Dowodzą bowiem analitycznie, że istnieje niezależna od liczebności n zbieżność entropii próby ciągu doskonale losowego do entropii oczekiwanej i pozwalają określać ją praktycznie. Nie ma w tym jednak niczego dziwnego, skoro wszystkie opierają się na tym samym modelu probabilistycznym w postaci dyskretnego rozkładu równomiernego oraz na analizach i obliczeniach wariancji ciągów zmiennych losowych o dowolnych wymiarach N. Warto jednak wspomnieć, że zdaniem Kołmogorowa „teoria informacji powinna wychodzić przed probabilistykę, a nie opierać się na niej" (Johnson, 2004, s. 8).
6. Niezależnie od interpretacji pochodzenia statystyki $\chi^2(N)_{\mathrm{LN}}$ i $\chi^2(n, N)_{\mathrm{P}}$ pozwalają według tych samych kryteriów i wartości jednoznacznie kwalifikować próby ciągów doskonale losowych i dyskwalifikować próby ciągów niedoskonale losowych, ponieważ powyższe wnioski i zależności dotyczą tylko ciągów doskonale losowych, a w przypadku ciągów niedoskonale losowych nie są zgodne.

10. Kryterium zgodności testu chi-kwadrat dla rzeczywistego ciągu doskonale losowego

Z matematycznego punktu widzenia wyniki pomiarów statystyk $\chi^2(n, N)_{\mathrm{P}}$ dla ciągów doskonale losowych świadczą o tym, że jeśli losowość ciągu jest bliska doskonałości, to ich wartości dla wszystkich N-wymiarowych zmiennych losowych są bliskie oczekiwanej wartości $\chi^2(N)_{\mathrm{LN}} = 1$, a skupienie wokół wartości oczekiwanej jest tym lepsze, im większe są liczebność próby n i wymiar testowanej zmiennej losowej N. Wyniki te mają ścisły związek z teorioinformacyjną analizą entropii Shannona, która korzysta z tego samego modelu probabilistycznego, ale pozwala prościej zinterpretować zjawisko losowości i wyrazić je globalną miarą w postaci entropii informacyjnej. To przenikanie się różnych nauk, prowadzące do wzajemnie zgodnych i konkretnych wyników, dających się ponadto zweryfikować pomiarowo, można uznać za największą zaletę proponowanej metody. W związku z tym powstaje pytanie, na podstawie jakich konkretnych kryteriów liczbowych powinna następować kwalifikacja doskonałości losowości prób generowanych ciągów losowych.

Na wstępie można powiedzieć, że użytkowników prób ciągów losowych w kryptografii, naukach technicznych i zapewne wielu innych zwykle nie interesuje, jak zostały one wygenerowane i wedle jakiego kryterium zostały zweryfikowane. Chcą tylko wiedzieć, czy na podstawie opinii ekspertów z dziedziny matematyki mogą przyjąć *a priori*, że są to faktycznie próby ciągów doskonale losowych, ponieważ konsekwencje użycia ciągów o potencjalnej nielosowości *a posteriori* mogą być fatalne – od fałszywych wyników badań naukowych po zagrożenie bezpieczeństwa kryptograficznego.

Dlatego należy zadać sobie pytanie, czy taka weryfikacja powinna polegać na przeprowadzeniu przez zaufanego eksperta klasycznej procedury testowania hipotezy statystycznej wobec każdej wygenerowanej próby ciągu, np. na podstawie hipotezy zerowej i alternatywnej, poziomu istotności i obszaru krytycznego statystyki testowej oraz innych pojęć takiej weryfikacji, czy może powinna opierać się na innych kryteriach, nieniosących ze sobą choćby werbalnie poczucia probabilistycznej niepewności wyniku takiej weryfikacji. Rozstrzygnięcia tego problemu nie ułatwiają trwające od wielu lat wśród statystyków dyskusje na temat ułomności procedury testowania istotności statystycznej hipotezy zerowej czy kryteriów wyboru poziomu istotności dla różnych rodzajów badań, zmierzające do całkowicie nowego zdefiniowania istotności statystycznej (Szreder, 2019, 2022). Problematyka ta, dotychczas tylko wzmiankowana w literaturze przedmiotu, znalazła ostatnio rzeczową i spójną postać monograficzną (Szreder i Kozłowski, 2024).

Wycofana rekomendacja NIST dla zestawu testów *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Application* opierała się właśnie na takich procedurach, a jednym z zasadniczych zarzutów wobec ich kryteriów była nietransparentność przyjętych wartości *p*, które w ogólności binarnie kwalifikowały badaną próbę jako spełniającą arbitralnie przyjęte kryteria losowości, ale w szczególności niczego nie mówiły o poziomie losowości danej próby, poza tym że przy kryterium $p > 0{,}01$ została ona w danym teście zakwalifikowana jako losowa. Krytykę takiego podejścia można zresztą znaleźć w wielu innych współczesnych publikacjach.

W praktyce decyzyjnej często wymagane jest binarne rozstrzygnięcie „tak – nie", ale nie oznacza to, że sama wartość *p* może przesądzić o tym, czy decyzja jest poprawna. Szeroko stosowana kategoria „istotności statystycznej" (zwykle interpretowana jako „$p < 0{,}05$") w sensie licencji na uznanie wniosków naukowych (lub na sugerowanie prawdy) prowadzi do poważnego wypaczenia procesu badawczego (Szreder, 2019, s. 55).

Co więcej, testy można było bardzo łatwo oszukać. Przykładowo dla ciągów opisanych w części 7:

- pomnożenie modulo 2 niedoskonale losowego ciągu z $s = 0{,}005$ i $K = 0{,}005$ z ciągłą sekwencją 101 ... 010 powodowało, że przed tą operacją spełniał on ok. 1/3 z 15 testów, a po niej – już 2/3 tych testów, z czego wynika, że były one wrażliwe na nierównowagę {0} i {1}, ale już nie zmiennych losowych o wymiarze $N \geq 2$, bo przecież statystyki takich zmiennych nie ulegają zmianie po tego rodzaju operacji;
- niedoskonale losowe ciągi z $s < 0{,}0005$ i $K < 0{,}0005$ spełniały wszystkie 15 testów, i to niezależnie od liczebności prób.

Nietrudno się domyślić, jakie zawodowe i etyczne rozterki może przeżywać przy takich wynikach konstruktor generatora ciągów losowych, który zwykle zna jego niedoskonałości, ale otrzymuje w każdym teście ocenę *passed*. Oczywiście w znacznie gorszej sytuacji byłby użytkownik ciągów z takiego generatora, ponieważ również

otrzymałby wyniki z oceną pozytywną, a używałby ciągów *de facto* nielosowych. Należy jednak zakładać, że konstruktor nie dopuści się takiego nadużycia zaufania wobec użytkownika, a w przypadku poddania generatora procedurze certyfikacji przez służby ochrony państwa nie byłoby szans na uzyskanie dla niego certyfikatu, ponieważ konstruktor jest prawnie zobowiązany do wskazania znanych sobie słabych miejsc w konstrukcji, a ich ujawnienie wyklucza skuteczne przeprowadzenie dowodu bezpieczeństwa kryptograficznego.

Podobne dylematy można by mnożyć. Warto powiedzieć jeszcze o dwóch, związanych z użytkowaniem. Potrzeby ilościowe w przypadku profesjonalnych zastosowań ciągów losowych są dzisiaj ogromne, liczone w gigabajtach, a użytkownik zwykle nie chce zbyt długo czekać na wygenerowanie ciągów, ich testowanie i wnioskowanie statystyczne, zwłaszcza przez osoby trzecie. Odpowiedzialność za niepoprawne wygenerowanie ciągów, ich testowanie i wnioskowanie statystyczne może mieć wymiar moralny w zastosowaniach naukowych, ale np. w kryptografii pociąga odpowiedzialność prawną, jeśli w wyniku błędów zostaną poniesione straty materialne czy szkody społeczne. Z powyższych powodów należy przyjąć następujące założenia do oceny:

- matematycznym kryterium oceny powinna być statystyka $\chi^2(N)_{\mathrm{LN}}$, za którą stoją interdyscyplinarność, jasność i prostota koncepcyjna oraz łatwość weryfikacji pomiarowej;
- zastosowania ciągów losowych są dużej wagi i należy wykluczyć wzięcie do danego zastosowania próby ciągu, co do której istnieje jakakolwiek wątpliwość dotycząca jej doskonałej losowości. W opozycji do zasady domniemania niewinności (czyli „lepiej uwolnić winnego niż skazać niewinnego") dla próby ciągu należy przyjąć zasadę domniemania nielosowości (innymi słowy, „lepiej odrzucić próbę ciągu doskonale losowego niż zaakceptować próbę ciągu nielosowego"). Powyższe założenie skutkuje tylko wydłużonym czasem generowania ciągów i pomiarów, gdyby któreś próby były dyskwalifikowane i odrzucane.

Najprościej i najbezpieczniej jest ustawić kryterium spełnienia testu na bezwarunkowej wartości $\chi^2(n, N)_{\mathrm{P}} = 1 \pm \varepsilon_N$, gdzie ε_N będzie arbitralnie wybranym, bezpiecznym odchyleniem, wynikającym z doświadczeń weryfikacji wielu prób ciągów bezdyskusyjnie bliskich losowej doskonałości. Muszą temu podlegać testy każdej N-wymiarowej zmiennej losowej, tzn. kiedy dla danej próby wszystkie wartości spełnią kryterium $\chi^2(n, N)_{\mathrm{P}} = 1 \pm \varepsilon_N$, ponieważ jego niespełnienie choćby dla jednej dowolnej N-wymiarowej zmiennej może oznaczać jakiś specyficzny typ nielosowości, który wyklucza uznanie danej próby za realizację ciągu doskonale losowego. Takie podejście można łatwo zalgorytmizować i będzie ono satysfakcjonowało zwłaszcza tych użytkowników, którzy – samodzielnie weryfikując ciągi – chcą mieć jednoznaczną odpowiedź na pytanie, czy dana próba ciągu spełnia kryterium losowej doskonałości.

Wartość ε_N będzie oczywiście inna dla każdego wymiaru zmiennej losowej N. Ponieważ wyniki podane w tabl. 3 są przedziałami, które w trakcie obserwacji rzadko

były przekraczane, to przyjmując kryterium $3\sigma = 3 \times 2^{-(N-1)/2}$, można założyć prawdopodobieństwo $P < 0{,}003$ i dopuścić 3 błędne decyzje o uznaniu ciągu losowego za nielosowy na 1000 wykonanych testów. Nie stanowiłoby to oczywiście o uznaniu generatora takiego ciągu za niespełniający wymagań, ponieważ powtórzenie badania z tymi samymi kryteriami wyklucza powtórzenie oceny negatywnej.

Propozycje dopuszczalnych wartości statystyki $\chi^2(n, N)_P$ dla rzeczywistych ciągów doskonale losowych przedstawiono w tabl. 7. Formalnymi kryteriami weryfikacji i klasyfikacji danej próby ciągu jako ciągu doskonale losowego są wartości zawarte w ostatnim wierszu tabl. 7. W przypadku gdy liczebność próby nie może wynosić 1 GB, można testować mniejsze próby z tymi samymi kryteriami, ale próby, które nie będą się mieściły w proponowanych przedziałach, należy odrzucać, zgodnie z zasadą domniemania nielosowości.

Tabl. 7. Dopuszczalne wartości statystyki $\chi^2(n, N)_P$ dla rzeczywistych ciągów doskonale losowych

Rozmiar próby w MB	$\chi^2(N=1)_P = 1 \pm \varepsilon_1$	$\chi^2(N=2)_P = 1 \pm \varepsilon_2$	$\chi^2(N=4)_P = 1 \pm \varepsilon_4$	$\chi^2(N=8)_P = 1 \pm \varepsilon_8$	$\chi^2(N=16)_P = 1 \pm \varepsilon_{16}$	$\chi^2(N=20)_P = 1 \pm \varepsilon_{20}$
1	0÷4	0÷4	0,6÷1,4	0,7÷1,5	0,98÷1,02	0,995÷1,005
10	0÷4	0÷4	0,7÷1,3	0,8÷1,3	0,98÷1,02	0,998÷1,002
100	<1	<1	0,7÷1,3	0,85÷1,2	0,99÷1,01	0,998÷1,002
1 000	<1	<1	0,8÷1,2	0,9÷1,1	0,99÷1,01	0,998÷1,002
σ_N	1,00	0,71	0,35	0,088	0,0055	0,00135
$3\sigma_N = \varepsilon_N$	3,00	2,12	1,06	0,265	0,0166	0,00405
$1 \pm \varepsilon_N$ dla 1 GB	0÷4,0	0÷3,12	0÷2,06	0,735÷1,265	0,983÷1,017	0,996÷1,004

Źródło: opracowanie własne.

Wybór liczebności próby do takich badań jest uwarunkowany trzema czynnikami:

- wydajnością generatora ciągów – jeśli jest ona ograniczona do wartości BR = b (elementów na sekundę), to czas generowania próby o liczebności n będzie trwał $t = n$ / BR i przykładowo dla typowego generatora sprzętowego o BR = 8 Mbit/s na próbę $n = 8 \times 1\,048\,576$ elementów (1 MB) trzeba poczekać ok. 1 s, ale na próbę $n = 8 \times 1\,048\,576\,000$ elementów (1 GB) – już 1000 s, czyli ponad kwadrans. Te czasy znacznie się skrócą w przypadku nowoczesnych generatorów kwantowych, których wydajność będzie sięgała BR = 1 Gbit/s, co dla powyższych liczebności skróci je do odpowiednio $t = 0{,}008$ s i $t = 8$ s;
- czasem testowania ciągów – sprzętowe możliwości użytego komputera i oprogramowanie do obliczeń ograniczają czas weryfikacji losowości próby ciągu. Można szacować, że na współczesnym komputerze z optymalizowanym oprogramowaniem wynosi on do 10 min na 1 GB ciągu. Nie jest to dużo w przypadku jednej próby o liczebności 100 MB, ale przy kilku próbach o liczebności 1 GB czas wydłuży się nawet do godziny;
- potrzebami operacyjnymi użytkownika ciągów – jeśli badacz wykonujący np. symulacje stochastyczne potrzebuje próby o liczebności $n = 8 \times 1\,048\,576\,000$

elementów (1 GB), to dysponując generatorem o wydajności BR = 8 Mbit/s, musi na nią poczekać kwadrans, zweryfikować ją testem $\chi^2(n, N)_P$ i po pozytywnym wyniku testu może w całości użyć do swoich celów. Nieco inna jest sytuacja kryptologa – on też powinien wygenerować próbę o jak największej liczebności i przetestować ją, aby się upewnić, że w całości reprezentuje ona doskonałą losowość, a następnie wybrać z niej podciągi o takiej liczebności, jakich potrzebuje. Są to zwykle klucze szyfrujące o liczebności n = 256 lub 512 elementów czy ciągi do poszukiwań liczb pierwszych o liczebności n od 1000 do 2000 elementów. Kryptolog nie może jednak poprzestać na wygenerowaniu próby o mniejszej liczebności, ponieważ musi mieć absolutną gwarancję, że używane przez niego ciągi pochodzą z próby ciągu doskonale losowego.

Wstępna weryfikacja autorska dała wyniki przedstawione w tabl. 8. Zbadano 40 prób ciągów o liczebności 1 GB każda, uznanych za doskonale losowe. Dla czytelnego porównania w dwóch głównych kolumnach zamieszczono wyniki tylko dla wymiarów N = 1, 2, 8 i 16. Pogrubioną czcionką został zaznaczony kierunek odchylenia w przypadku przekroczenia kryterium przez daną próbę, odpowiednio:

- w lewej kolumnie dla kryterium $\chi^2(n, N)_P = 1 \pm \varepsilon_N = 1 \pm \sigma_N$;
- w prawej kolumnie dla kryterium $\chi^2(n, N)_P = 1 \pm \varepsilon_N = 1 \pm 3\sigma_N$.

Tabl. 8. Wartości statystyk $\chi^2(n, N)_P$ dla 40 prób rzeczywistych ciągów doskonale losowych

Próby	$\chi^2(n = 1\text{ GB}, N)_P = 1 \pm \varepsilon_N = 1 \pm \sigma_N$				$\chi^2(n = 1\text{ GB}, N)_P = 1 \pm \varepsilon_N = 1 \pm 3\sigma_N$			
	$\chi^2(N=1)_P = 1 \pm \sigma_1$	$\chi^2(N=2)_P = 1 \pm \sigma_2$	$\chi^2(N=8)_P = 1 \pm \sigma_8$	$\chi^2(N=16)_P = 1 \pm \sigma_{16}$	$\chi^2(N=1)_P = 1 \pm 3\sigma_1$	$\chi^2(N=2)_P = 1 \pm 3\sigma_2$	$\chi^2(N=8)_P = 1 \pm 3\sigma_8$	$\chi^2(N=16)_P = 1 \pm 3\sigma_{16}$
σ_N	1,00	0,71	0,088	0,0055	.	.	.	.
$3\sigma_N$	.	.	.	.	3,00	2,12	0,265	0,0166
$1 \pm \varepsilon_N$	0,0÷2,0	0,29÷1,71	0,912÷ ÷1,088	0,9945÷ ÷1,0055	0,0÷4,0	0,0÷3,12	0,735÷ ÷1,265	0,983÷ ÷1,017
1	↑ **4,398**	↑ **2,755**	1,069	0,996	↑ **4,398**	2,755	1,069	0,996
2	↑ **4,010**	↑ **2,105**	1,071	↑ **1,007**	4,010	2,105	1,071	1,007
3	0,212	0,629	1,038	0,998	0,212	0,629	1,038	0,998
4	0,542	1,357	↑ **1,111**	0,997	0,542	1,357	1,111	0,997
5	1,341	0,830	1,032	0,996	1,341	0,830	1,032	0,996
6	1,848	1,127	0,987	↑ **1,007**	1,848	1,127	0,987	1,007
7	0,736	0,606	1,004	0,995	0,736	0,606	1,004	0,995
8	0,388	0,358	↓ **0,893**	↑ **1,010**	0,388	0,358	0,893	1,010
9	↑ **3,192**	1,113	1,003	0,992	3,192	1,113	1,003	0,992
10	1,942	↑ **2,599**	↓ **0,767**	1,003	1,942	2,599	0,767	1,003
11	0,006	0,621	0,990	1,000	0,006	0,621	0,990	1,000
12	1,117	0,504	0,935	1,000	1,117	0,504	0,935	1,000
13	0,118	↓ **0,149**	↓ **0,863**	1,006	0,118	0,149	0,863	1,006
14	0,075	↓ **0,100**	0,932	1,000	0,075	0,100	0,932	1,000
15	0,027	0,636	1,036	0,995	0,027	0,636	1,036	0,995
16	1,812	1,477	↓ **0,868**	0,996	1,812	1,477	0,868	0,996

Tabl. 8. Wartości statystyk $\chi^2(n, N)_P$ dla 40 prób rzeczywistych ciągów doskonale losowych (dok.)

Próby	$\chi^2(n = 1\text{ GB}, N)_P = 1 \pm \varepsilon_N = 1 \pm \sigma_N$				$\chi^2(n = 1\text{ GB}, N)_P = 1 \pm \varepsilon_N = 1 \pm 3\sigma_N$			
	$\chi^2(N=1)_P = 1 \pm \sigma_1$	$\chi^2(N=2)_P = 1 \pm \sigma_2$	$\chi^2(N=8)_P = 1 \pm \sigma_8$	$\chi^2(N=16)_P = 1 \pm \sigma_{16}$	$\chi^2(N=1)_P = 1 \pm 3\sigma_1$	$\chi^2(N=2)_P = 1 \pm 3\sigma_2$	$\chi^2(N=8)_P = 1 \pm 3\sigma_8$	$\chi^2(N=16)_P = 1 \pm 3\sigma_{16}$
17	0,596	0,481	↑ **1,144**	0,996	0,596	0,481	1,144	0,996
18	↑ **2,707**	1,149	↑ **1,094**	0,998	2,707	1,149	1,094	0,998
19	0,526	0,793	↓ **0,867**	1,001	0,526	0,793	0,867	1,001
20	0,570	0,454	0,936	↑ **1,009**	0,570	0,454	0,936	1,009
21	0,958	0,921	1,020	1,001	0,958	0,921	1,020	1,001
22	1,996	1,733	0,974	1,001	1,996	1,733	0,974	1,001
23	0,055	↓ **0,203**	1,033	↑ **1,007**	0,055	0,203	1,033	1,007
24	↑ **3,083**	1,480	1,047	0,997	3,083	1,480	1,047	0,997
25	0,121	0,453	0,986	0,998	0,121	0,453	0,986	0,998
26	0,096	1,494	↓ **0,824**	1,004	0,096	1,494	0,824	1,004
27	1,223	0,512	↑ **1,196**	1,000	1,223	0,512	1,196	1,000
28	0,000	0,491	↑ **1,260**	1,001	0,000	0,491	1,260	1,001
29	↑ **3,343**	↑ **1,914**	0,949	0,998	3,343	1,914	0,949	0,998
30	0,643	0,510	↑ **1,106**	1,001	0,643	0,510	1,106	1,001
31	0,710	0,568	↑ **1,131**	0,995	0,710	0,568	1,131	0,995
32	0,033	↓ **0,270**	1,042	0,995	0,033	0,270	1,042	0,995
33	0,315	↓ **0,272**	↑ **1,123**	↓ **0,988**	0,315	0,272	1,123	0,988
34	0,005	↓ **0,074**	0,968	0,996	0,005	0,074	0,968	0,996
35	0,542	0,973	0,940	1,004	0,542	0,973	0,940	1,004
36	0,309	↓ **0,109**	↓ **0,802**	1,001	0,309	0,109	0,802	1,001
37	0,221	↓ **0,074**	1,019	0,998	0,221	0,074	1,019	0,998
38	0,820	0,462	1,043	1,000	0,820	0,462	1,043	1,000
39	1,028	1,068	↑ **1,102**	1,000	1,028	1,068	1,102	1,000
40	1,062	0,800	1,087	0,997	1,062	0,800	1,087	0,997
Udział w całości w %	3,75	7,50	10,60	3,75	0,63	0	0	0
Podsumowanie	41 wyników nieco ponad kryterium σ_N – 26%				1 wynik na granicy kryterium $3\sigma_N$ – 0,6%			

Źródło: opracowanie własne.

Kryterium $\chi^2(n, N)_P = 1 \pm \varepsilon_N = 1 \pm \sigma_N$ spełnia ponad 74% wyników testów. Wyniki, które go nie spełniają, mają typowo incydentalny charakter, tzn. nie dotyczą wszystkich zmiennych losowych, a zwykle jednej, osiem razy dwóch, a tylko w trzech przypadkach trzech zmiennych. Należy jednak zauważyć, że otrzymane wartości były na granicy kryterium i gdyby przyjąć je choćby jako $1{,}5\sigma_N$, ich liczba zmniejszyłaby się do zaledwie kilku.

Kryterium $\chi^2(n, N)_P = 1 \pm \varepsilon_N = 1 \pm 3\sigma_N$ spełniają w zasadzie wszystkie wyniki testów – zaledwie jeden przekroczył kryterium o niecałe 10%.

Można zatem przyjąć, że wyniki te odpowiadają z dobrym przybliżeniem kryteriom σ_N i $3\sigma_N$, właściwym rozkładowi normalnemu, dla którego w takich przedziałach oczekuje się odpowiednio 68,3% i 99,7% wyników. Niewielka różnica między 74% wyników testów i 68,3% właściwych rozkładowi normalnemu dla kryterium σ_N wynika prawdopodobnie z tego, że wariancja lewego ogona statystyki $\chi^2(N)_{LN}$ jest ograniczona wartością zerową i w tym zawężonym przedziale mieszczą się wszystkie wyniki

pomiarów, które dodają się do nominalnych 68,3%. Uzyskane wyniki kolejny raz potwierdzają poprawność przeprowadzonych analiz teoretycznych i dobrą zgodność założonego kryterium pomiarowego.

Poczynione ustalenia prowadzą do wniosku, że w badaniu losowości ciągów należy:

- wygenerować próbę ciągu o maksymalnej liczebności, na jaką pozwalają wydajność generatora ciągu i akceptowalny czas oczekiwania, ze wskazaniem na liczebność co najmniej $n = 1$ GB;
- sprawdzić dla wszystkich badanych wymiarów N, czy wartości $\chi^2(n, N)_P = 1 \pm \varepsilon_N$. Tylko w przypadku gdy wszystkie spełnią to kryterium, można uznać próbę za realizację ciągu doskonale losowego.

Na koniec należy rozważyć jeszcze problem stochastycznej stabilności generowanych ciągów, przez którą rozumie się ich stacjonarność i ergodyczność (Gray, 2013). Mówiąc w największym skrócie, właściwość stacjonarności w szerszym sensie można przypisać ciągom, których statystyki są takie same w każdej próbie, a ergodyczności – ciągom, których statystyki są takie same dla kolejnych podciągów danego ciągu. Ergodyczność jest zatem cechą mocniejszą i ciąg ergodyczny jest również stacjonarny, ale ciąg stacjonarny nie musi być ergodyczny.

Badanie ergodyczności próby ciągu najłatwiej przeprowadzić poprzez równomierne podzielenie próby na podciągi i sprawdzenie statystyk wszystkich kolejnych podciągów, które następnie są do siebie sukcesywnie dodawane i pokazują statystykę całej próby. Łatwo zauważyć, że test $\chi^2(n, N)_P$ umożliwia takie badanie ze względu na naturę swojej statystyki pomiarowej. Ilustrację badania ergodyczności ciągu przedstawiono na schemacie.

Schemat. Badanie ergodyczności ciągu według statystyk $x^2(n/p, N)$

Źródło: opracowanie własne.

Opisane w częściach 6–8 badania prób ciągów w sensie obserwacji bieżących wartości $\chi^2(n/p = 1 \text{ MB}, N)_P$ kolejnych podciągów o liczebności 1 MB, podsumowane wartością $\chi^2(n, N)_P$ sumy wszystkich podciągów, były właśnie badaniami ergodyczności.

Można od razu przyjąć, że bezwarunkowe uznanie próby za ergodyczną i doskonale losową następuje w przypadku jednoczesnego spełnienia wszystkich poniższych warunków:

- każdy podciąg ma wymagane statystyki $\chi^2(n/p, N)_P$, uwzględniające jego liczebność n/p;
- cała próba ma wymagane statystyki $\chi^2(n, N)_P$, uwzględniające jej liczebność n;
- statystyki $\chi^2(n/p, N)_P$ wszystkich podciągów mogą być rozproszone wokół statystyki całej próby $\chi^2(n, N)_P$ tylko z powodu ich ograniczonej liczebności.

Jednak błądzenie przypadkowe – zwłaszcza w próbach o bardzo dużej liczebności – może incydentalnie znacząco odchylić statystyki zmiennych losowych o dowolnych rozmiarach N od wartości oczekiwanych, co jest widoczne w tabl. 8. Próby ciągów, które są w niej opisane, były generowane jedna po drugiej, z kilkusekundowymi przerwami na zapisanie ich na dysku komputera, więc nie wystąpiły tu zakłócenia w postaci czynników starzeniowych, klimatyczno-mechanicznych czy elektromagnetycznych, które mogłyby zmieniać warunki pracy generatora wpływające na statystyki kolejnych prób ciągu i prowadzić do błędów nielosowych (ang. *non-random errors*; Szreder, 2021). Z późniejszej analizy podciągów o liczebności 1 MB wynikało, że lokalne nielosowości były okazywane przez podciągi o liczebności od kilku do kilkudziesięciu megabajtów, ale następne podciągi miały już oczekiwane statystyki, co wskazuje właśnie na incydentalne i niewielkie przekraczanie odchyleń standardowych. A zatem raczej intuicyjna hipoteza Fellera, że „uczciwa moneta może dać niedorzeczny ciąg, w którym zmiana prowadzenia nie nastąpi w ciągu milionów kolejnych doświadczeń" (Feller, 2006, s. 87; obszerniejszy cytat został zamieszczony w części 2), ma podstawy doświadczalne, nie mając jednak źródeł analitycznych, ponieważ probabilistyka nie zna zależności, która określałaby, czy są to tysiące, czy miliony doświadczeń.

W przypadku wystąpienia lokalnych nielosowości można przyjąć dwie strategie weryfikacji:

- bezwarunkową, kiedy cała próba zostaje zdyskwalifikowana i odrzucona z powodu przekroczenia dowolnego z powyżej opisanych kryteriów;
- warunkową, kiedy z próby zostają usunięte te podciągi, w których stwierdzono przekroczenia dowolnego z kryteriów w danym podciągu.

Weryfikacja warunkowa odpowiada niezależnemu generowaniu mniej liczebnych prób i odrzucaniu prób niespełniających kryteriów. W każdym przypadku można to dopuścić, ponieważ jest poprawne matematycznie: usuwanie dowolnych podciągów z ciągu losowego nie zmienia jego losowości (a w konsekwencji – statystyk), jednak prowadzi do wygenerowania prób o mniejszej liczebności, niedających gwarancji, że okażą potencjalne nielosowości, które opisano w części 7 i zilustrowano wynikami zamieszczonymi w tabl. 5 i 6.

Przy generowaniu ciągów doskonale losowych najlepiej zatem stosować weryfikację bezwarunkową, ponieważ kosztem nielicznych zdyskwalifikowanych i odrzuconych prób, a w konsekwencji – relatywnie niewielkiego wydłużenia czasu generowania ciągów i pomiarów daje ona znacznie większą pewność braku potencjalnych nielosowości w generowanych ciągach.

11. Podsumowanie

Przedstawiona w artykule propozycja autorskiej metody testowania ciągów losowych została opracowana z intencją obiektywnej oceny tych ciągów, wolnej od specyficznych właściwości i ograniczeń dotychczas stosowanych zestawów testów. Opiera się na rzeczywistych właściwościach ciągów losowych i ich elementarnych parametrach opisywanych przez probabilistykę, statystykę i teorię informacji. Wykorzystując je, wykazano, że statystyka χ^2_{LN} dla ciągów doskonale losowych teoretycznie nie zależy od liczebności badanej próby, jest jednoparametryczna, a po normalizacji względem wymiaru zmiennej losowej N jest równa 1. Nie zmienia to istoty testu χ^2, a tylko stawia jego statystyce wyższe wymagania, które w praktyce może spełnić jedynie ciąg doskonale losowy. Jednocześnie wykazano, że taka statystyka opisuje zgodność zbieżności entropii próby ciągu doskonale losowego względem jej wartości oczekiwanej, niezależnie od liczebności próby.

Metoda została pierwotnie opracowana jako narzędziowe wsparcie projektu kwantowego mikrofalowego generatora ciągów losowych o wydajności 1 Gbit/s, skonstruowanego w Wojskowym Instytucie Łączności w ramach własnych prac statutowych. Przewiduje się zastosowanie go jako elementu ogólnodostępnego serwera ciągów losowych do celów naukowo-technicznych. Serwer będzie umożliwiał bezpłatne pobieranie plików z ciągami losowymi, które będą miały referencyjne właściwości i parametry statystyczne potwierdzone pomiarami wykonywanymi dla każdej próby. Przewiduje się ciągłą dostępność prób ciągów o liczebności 1 GB każda. Pobrane próby będą automatycznie kasowane, co będzie stanowiło namiastkę niepowtarzalności. Skasowane próby będą zastępowane nowymi.

Pierwszym praktycznym i oficjalnym dowodem skuteczności metody było jej zastosowanie do oceny bezpieczeństwa kryptograficznego w procesie ustawowej certyfikacji prostszego generatora o wydajności 10 Mbit/s (Wojskowy Instytut Łączności, 2025), który uzyskał certyfikat nr 8/2025 wydany przez Jednostkę Certyfikującą Służby Kontrwywiadu Wojskowego i dopuszczający go do zastosowań w zakresie kryptograficznej ochrony informacji o klauzulach „ściśle tajne", „NATO Confidential" i „Confidentiel UE" / „EU Confidential".

Bibliografia

Bobrowski, D. (2002). *Ciągi losowe*. Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Feller, W. (2006). *Wstęp do rachunku prawdopodobieństwa*. Wydawnictwo Naukowe PWN.

Gray, R. M. (2013). *Entropy and Information Theory*. Springer.

International Telecommunication Union. (2019). *Recommendation ITU-T X.1702: Quantum noise random number generator architecture*. https://www.itu.int/rec/T-REC-X.1702-201911-I/en.

Jacak, M. M., Jóźwiak, P., Niemczuk, J., Jacak, J. E. (2021). Quantum generators of random numbers. *Scientific Reports*, *11*, 1–21. https://doi.org/10.1038/s41598-021-95388-7.

Johnson, O. (2004). *Information Theory and the Central Limit Theorem*. Imperial College Press. https://doi.org/10.1142/p341.

L'Ecuyer, P. (2012). Random Number Generation. W: J. Gentle, W. Härdle, Y. Mori (red.), *Handbook of Computational Statistics. Concepts and Methods* (s. 35–71). Springer. https://doi.org/10.1007/978-3-642-21551-3_3.

L'Ecuyer, P., Simard, R. (2013). *TestU01. A Software Library in ANSI C for Empirical Testing of Random Number Generators. User's Guide, compact version*. https://simul.iro.umontreal.ca/testu01/guideshorttestu01.pdf.

Leśniewicz, M. (2009). *Sprzętowa generacja losowych ciągów binarnych.* Wojskowa Akademia Techniczna.

National Instituite of Standards and Technology. (2022). *Decision to Revise NIST SP 800-22 Rev. 1a*. https://csrc.nist.gov/News/2022/decision-to-revise-nist-sp-800-22-rev-1a.

Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., Vo, S. (2010). *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. National Institute of Standards and Technology. https://csrc.nist.gov/pubs/sp/800/22/r1/upd1/final.

Shannon, C. E. (1949). Communication Theory of Secrecy Systems. *Bell System Technical Journal*, *28*(4), 656–715. https://doi.org/10.1002/j.1538-7305.1949.tb00928.x.

Sulewski, P. (2019). Porównanie generatorów liczb pseudolosowych. *Wiadomości Statystyczne. The Polish Statistician*, *64*(11), 5–31. https://doi.org/10.5604/01.3001.0013.7605.

Szreder, M. (2019). Istotność statystyczna w czasach big data. *Wiadomości Statystyczne. The Polish Statistician*, *64*(11), 42–57. https://doi.org/10.5604/01.3001.0013.7583.

Szreder, M. (2021). Błędy nielosowe i ich znaczenie w testowaniu hipotez. *Wiadomości Statystyczne. The Polish Statistician*, *66*(3), 7–21. https://doi.org/10.5604/01.3001.0014.7984.

Szreder, M. (2022). Szanse i iluzje dotyczące korzystania z dużych prób we wnioskowaniu statystycznym. *Wiadomości Statystyczne. The Polish Statistician*, *67*(8), 1–16. https://doi.org/10.5604/01.3001.0015.9704.

Szreder, M., Kozłowski, A. (2024). *Wnioskowanie na podstawie prób losowych i nielosowych*. Wydawnictwo Uniwersytetu Gdańskiego.

Wojskowy Instytut Łączności. (2025). *Sprzętowy generator ciągów losowych SGCL-1MB-2*. https://www.wil.waw.pl/wil/oferta/produkty/ochrona-kryptograficzna/23219,Sprzetowy-Generator-Ciagow-Losowych-SGCL-1MB-2.html.