§3. Testowanie generatorów

Robert Janczewski

http://www.kaims.pl/~robert/MISS/

Celem testów jest ustalenie, czy i na ile dany generator odbiega od idealnego.

Testom podlegają rozmaite właściwości, które generator powinien mieć (np. wartość średnią taką samą jak generator idealny), lub których mieć nie powinien (np. okres).

Testy mogą mieć charakter teoretyczny lub empiryczny.

Testy teoretyczne badają globalne właściwości generatorów. Umożliwiają one ustalenie m.in. czy generator

- jest okresowy i jakie są możliwe wartości jego okresu;
- generuje wszystkie liczby z interesującego nas zakresu;
- generuje liczby równomiernie.

Częstym efektem ubocznym testów teoretycznych są własności, które parametry generatora muszą spełniać, by generowane przez niego liczby miały dobre własności statystyczne.

Testy teoretyczne są w ogólności trudne do przeprowadzenia, ponieważ wymagają umiejętności analizy wzorów definiujących generowane liczby.

Dla przykładu rozważmy generator liniowy kongruencyjny o parametrach $m = 2^{32}$, $a = 3$, $c = 1$ (jest on dany wzorem $X_{k+1} = (aX_k + c) \bmod m$).

Interesuje nas, jakie jest prawdopodobieństwo tego, że wygeneruje on kolejno dwie takie same liczby.

Z definicji tego generatora wynika, że wygenerowanie liczby x dwa razy pod rząd jest możliwe tylko wtedy, gdy $m|(a-1)x + c$.

To jednak nigdy nie zajdzie, bo m jest parzyste, a $(a-1)x + c$ nieparzyste.

Interesujące nas prawdopodobieństwo wynosi zatem 0. Ile wynosi jego wartość dla generatora idealnego?

Nietrudno jest zauważyć, że dla $x = 0, 1, \ldots, m-1$ mamy

$$Pr\{X_1 \neq X_2, \ldots, X_{k-1} \neq X_k, X_k = X_{k+1} = x\} = \frac{(m-1)^{k-1}}{m^{k+1}}.$$

Stąd

$$\begin{aligned} Pr\{\exists_k X_k = X_{k+1} = x\} &= \sum_{k=1}^{\infty} \frac{(m-1)^{k-1}}{m^{k+1}} \\ &= \frac{1}{m^2} \sum_{k=1}^{\infty} \left(1 - \frac{1}{m}\right)^{k-1} = \frac{1}{m}, \end{aligned}$$

więc $Pr\{\exists_k X_k = X_{k+1}\} = \sum_{x=0}^{m-1} Pr\{\exists_k X_k = X_{k+1} = x\} = 1 \neq 0.$

Test empiryczny jest 3-etapowy. W pierwszym etapie generujemy próbę losową, czyli pobieramy z generatora ciąg liczb lub bitów.

W drugim etapie wyznaczamy statystykę testową, czyli liczbę, która wyznaczana jest na podstawie otrzymanej wcześniej próby losowej.

W trzecim etapie porównujemy otrzymaną liczbę z rozkładem teoretycznym i decydujemy, czy jest ona akceptowalna (generator zaliczył test) czy też nie (generator nie zaliczył testu).

Testy empiryczne powtarza się wielokrotnie, obserwując ile testów generator zaliczył, a ile oblał.

Na tej podstawie wyciąga się wnioski odnośnie generatora – generator o dobrych właściwościach powinien zaliczyć większość testów.

(Należy być ostrożnym podczas wyciągania wniosków z testów empirycznych, ponieważ potrafią być one mylące.)

Istnieją programy komputerowe, zwane bateriami testów, które umożliwiają mniej lub bardziej automatyczne testowanie generatorów.

Jednym z nich jest Dieharder, dostępny pod adresem http://www.phy.duke.edu/~rgb/General/dieharder.php.

Dieharder zawiera implementacje wielu klasycznych generatorów liczb losowych oraz wielu testów empirycznych, w tym niektórych z omówionych na następnych stronach.

Test χ^2 jest jednym z tzw. testów zgodności. Testy zgodności sprawdzają, czy generator X generuje liczby, które są niezależne i mają zadany rozkład.

W pierwszym kroku ustalamy parametry testu:

- liczby a i b takie, że $Pr\{X < a\} = 0$ i $Pr\{X < b\} = 1$ (może być $a = -\infty$ i $b = +\infty$);
- liczbę naturalną k oraz punkty $a = a_0 < a_1 < \ldots < a_k = b$;
- prawdopodobieństwa $p_i = Pr\{a_{i-1} \leq X < a_i\}$ oraz długość próbki losowej n;
- poziom istotności α.

Następnie generujemy próbkę losową $X_1, X_2, \ldots, X_n$ i na jej podstawie wyznaczamy liczby Y_i zgodnie z wzorem

$$Y_i = |\{j : 1 \leq j \leq n \wedge a_{i-1} \leq X_j < a_i\}|$$

oraz statystykę testową

$$V = \sum_{i=1}^{k} \frac{(Y_i - np_i)^2}{np_i}.$$

Przy dużych wartościach n statystyka V będzie miała rozkład χ^2 o $k-1$ stopniach swobody.

Wynik testu odczytujemy z wykresu dystrybuanty rozkładu χ^2 – jeśli wartość dystrybuanty w punkcie V nie przekracza $1 - \alpha$, to test został zaliczony.

Istotną sprawą jest wybór wartości n. Zbyt mała wartość spowoduje, że rozkład V będzie bardzo odbiegał od χ^2 i wynik nie będzie wiarygodny. Zbyt duża zaś może ukryć rozmaite anomalie.

Zaleca się, żeby $np_i \geq 5$ dla każdego i. Test χ^2 powinno się powtórzyć kilkukrotnie, za każdym razem zmieniając wartość n.

Wybrane wartości krytyczne testu χ^2, w zależności od liczby stopni swobody i poziomu istotności:

	0.01	0.05	0.10	0.25	0.50
1	6.63490	3.84146	2.70554	1.32330	0.45494
2	9.21034	5.99146	4.60517	2.77259	1.38629
3	11.3449	7.81473	6.25139	4.10835	2.36597
4	13.2767	9.48773	7.77944	5.38527	3.35669
5	15.0863	11.0705	9.23636	6.62568	4.35146
6	16.8119	12.5916	10.6446	7.84080	5.34812
7	18.4753	14.0671	12.0170	9.03715	6.34581
8	20.0902	15.5073	13.3616	10.2189	7.34412
9	21.6660	16.9190	14.6837	11.3888	8.34283

Test Kołmogorowa to drugi z testów zgodności. Sprawdza on, czy generator tworzy liczby o zadanej dystrybuancie F.

Parametrami tego testu są długość próbki losowej n oraz poziom istotności α.

Generujemy próbkę losową $X_1, X_2, \ldots, X_n$ i na jej podstawie wyznaczamy tzw. dystrybuantę empiryczną

$$F_n(x) = \frac{\left|\{i : 1 \leq i \leq n \wedge X_i \leq x\}\right|}{n}.$$

Wraz ze wzrostem n dystrybuanta empiryczna powinna dążyć do F.

Aby sprawdzić, jak blisko znajdują się F i F_n, wprowadzamy dwie miary odległości

$$K_n^+ = \sqrt{n} \sup_{x \in \mathbb{R}} (F_n(x) - F(x)) \text{ i } K_n^- = \sqrt{n} \sup_{x \in \mathbb{R}} (F(x) - F_n(x)).$$

Miary te mają identyczny rozkład o dystrybuancie danej wzorem

$$Pr\{K_n^+ \leq x\} = Pr\{K_n^- \leq x\} \approx 1 - e^{-2x^2} + \frac{2}{3\sqrt{n}} e^{-2x^2}.$$

(We wzorze dokładnym po prawej stronie równości występuje $1 - \frac{x\sqrt{n}}{n^n} \sum_{x\sqrt{n} \leq k \leq n} \binom{n}{k} (k - x\sqrt{n})^k (x\sqrt{n} + n - k)^{n-k-1}$.)

Wynik testu odczytujemy z dystrybuant K_n^+, K_n^- – jeśli wartość wspólnej dystrybuanty w punktach K_n^+, K_n^- nie przekracza $1-\alpha$, to test został zaliczony.

Zamiast dwóch odległości można by wprowadzić jedną $K_n = \max\{K_n^+, K_n^-\}$, ale jej dystrybuanta wyraża się bardziej skomplikowanym wzorem niż dystrybuanty K_n^+, K_n^-.

Podobnie jak w przypadku testu χ^2, ten test trzeba powtórzyć kilkukrotnie, zmieniając za każdym razem wartość n.

Wzór opisujący K_n^+ i K_n^- nie nadaje się do zastosowania w obliczeniach komputerowych. Dlatego przekształca się go do postaci

$$K_n^+ = \sqrt{n} \max_{1 \le j \le n} \left(\frac{j}{n} - F(X_{j:n}) \right)$$

i

$$K_n^- = \sqrt{n} \max_{1 \le j \le n} \left(F(X_{j:n}) - \frac{j-1}{n} \right)$$

gdzie $X_{j:n}$ oznacza j-tą co do wielkości spośród liczb $X_1, X_2, \ldots, X_n$.

Ten wzór pozwala na wyznaczenie obu miar w czasie $O(n \log n)$.

Zamiast wykonywać test raz i wyznaczać jego statystykę, możemy wykonać test wielokrotnie, uzyskując w efekcie ciąg statystyk. Na otrzymanym ciągu można wykonać test χ^2 lub Kołmogorowa.

Powyższa procedura nosi nazwę testu dwupoziomowego.

W tym teście badamy, czy pary kolejnych liczb generowanych przez generator są rozmieszczone równomiernie i czy są generowane niezależnie.

Jedynymi parametrami tego testu są poziom istotności α i długość próbki losowej n (w tym teście n musi być parzyste). Zakładamy przy tym, że generator generuje liczby z zakresu $0, 1, \ldots, m-1$.

Generujemy próbkę losową $X_1, X_2, \ldots, X_n$ a z otrzymanych liczb tworzymy pary $(X_1, X_2), (X_3, X_4), \ldots, (X_{n-1}, X_n)$.

Dla wszystkich liczb p, q z zakresu $0, 1, \ldots, m-1$ sprawdzamy, ile razy w otrzymanych parach występuje wartość (p, q).

Na otrzymanych wynikach wykonujemy test χ^2.

W analogiczny sposób można przeprowadzić test częstości trójek, czwórek itd.

Ten test staje się niepraktyczny, gdy m jest duże, wówczas stosujemy inne testy, ewentualnie badamy fragmenty otrzymanych liczb, np. ich reszty z dzielenia przez ustaloną liczbę.

Ten test sprawdza, czy t-elementowe krotki liczb generowanych przez generator mają odpowiedni rozkład.

Parametrami tego testu są liczba t, długość próbki losowej n (n musi być podzielne przez t) oraz poziom istotności α.

Z próbki losowej $X_1, X_2, \ldots, X_n$ tworzymy ciąg $V_i = \max\{X_{ti-t+1}, X_{ti-t+2}, \ldots, X_{ti}\}$, $i = 1, 2, \ldots, n/t$.

Na otrzymanym ciągu wykonujemy test Kołmogorowa, pamiętając o tym, że dystrybuanta V_i ma postać x^t, jeśli zmienne X_i są niezależne i mają rozkład równomierny w $[0, 1)$.

Test sum wykonujemy tak samo jak test największy z t, ale zmieniamy wzór opisujący V_i na następujący

$$V_i = \sum_{j=ti-t+1}^{ti} X_j.$$

Zmienia się rzecz jasna dystrybuanta V_i. Dla $t = 2$ ma ona postać:

$$F(x) = \begin{cases} \frac{1}{2}x^2, & \text{gdy } 0 \leq x \leq 1, \\ 1 - \frac{1}{2}(2-x)^2, & \text{gdy } 1 < x \leq 2. \end{cases}$$

Test permutacji także bada zachowanie t-elementowych krotek zbudowanych z liczb wygenerowanych przez generator.

Parametrami tego testu są liczba t, długość próbki losowej n (n musi być podzielne przez t) oraz poziom istotności α.

Próbkę losową $X_1, X_2, \ldots, X_n$ dzielimy na t-elementowe krotki i przekształcamy zgodnie z wzorem

$$(T_1, T_2, T_3, \ldots, T_t) \mapsto (f(T_1), f(T_2), \ldots, f(T_t)),$$

gdzie $f(j) = |\{i \leq t\colon\ T_i \leq T_j\}|$.

Następnie zliczamy, ile razy w otrzymanych krotkach pojawiła się każda z możliwości i stosujemy test χ^2.

Test ten stosujemy niemal wyłącznie wtedy, gdy elementy w krotkach się nie powtarzają. W tym przypadku prawdopodobieństwo każdej z otrzymanych permutacji jest równe $1/t!$.

Ten test bada odstępy, jakie pojawiają się pomiędzy kolejnymi liczbami z interesującego nas przedziału $[a, b]$.

Poprzez odstęp rozumiemy tutaj taki podciąg próbki losowej, że:

- element poprzedzający podciąg, o ile istnieje, należy do $[a, b]$;
- element następujący po podciągu należy do $[a, b]$;
- elementy podciągu nie należą do $[a, b]$.

Parametrami tego testu są liczby n, t, przedział $[a, b]$ i poziom istotności α.

Generujemy próbkę losową do momentu, aż uzyskamy n odstępów pomiędzy kolejnymi liczbami z przedziału $[a, b]$. Zliczamy liczbę odstępów długości $0, 1, \ldots, t-1$ i dłuższych niż $t-1$.

Na otrzymanych wynikach wykonujemy test χ^2, pamiętając, że prawdopodobieństwo uzyskania odstępu długości (co najmniej) d jest równe $(1-p)^d p$ ($(1-p)^d$), gdzie $p = Pr\{a \leq X \leq b\}$.

Tego testu nie da się wykonać, jeżeli generator nie generuje liczb należących do przedziału $[a, b]$.

W tym teście badamy właściwości 5-elementowych krotek generowanych przez generator. Zakładamy, że generator może wygenerować $m \geq 5$ różnych liczb.

Parametrami testu są długość próbki losowej n (musi być podzielna przez 5) oraz poziom istotności α.

Generujemy próbkę losową, dzielimy ją na 5-elementowe krotki i zliczamy liczbę krotek następujących typów:

- *abcde* – każda liczba w krotce jest inna;
- *aabcd* – dwie liczby w krotce są takie same, pozostałe mają różne wartości;

- *aabbc* – w krotce są dwie pary identycznych liczb, ostatnia liczba jest od nich różna;
- *aaabc* – trzy liczby w krotce są takie same, pozostałe mają różne wartości;
- *aaabb* – w krotce jest para i trójka identycznych liczb;
- *aaaab* – cztery liczby w krotce są takie same, ostatnia liczba jest od nich różna;
- *aaaaa* – wszystkie liczby w krotce są takie same.

Na otrzymanych wynikach wykonujemy test χ^2.

Prawdopodobieństwa uzyskania poszczególnych typów krotek są następujące:

Typ	Prawdopodobieństwo
$abcde$	$m(m-1)(m-2)(m-3)(m-4)/m^5$
$aabcd$	$10m(m-1)(m-2)(m-3)/m^5$
$aabbc$	$15m(m-1)(m-2)/m^5$
$aaabc$	$10m(m-1)(m-2)/m^5$
$aaabb$	$10m(m-1)/m^5$
$aaaab$	$5m(m-1)/m^5$
$aaaaa$	$1/m^4$

Ten test sprawdza, ile liczb trzeba wygenerować, aby uzyskać wszystkie liczby z pewnego zakresu.

Parametrami tego testu są liczby d i t oraz poziom istotności α.

Generujemy próbkę losową tak długo, aż wystąpią w niej wszystkie liczby z zakresu $0, 1, \ldots, d-1$.

Statystyką testową, do której stosujemy test χ^2 jest długość otrzymanej próbki (wszystkie próbki długości większej lub równej t liczymy razem).

Ten test sprawdza, jak długie monotoniczne serie liczb generuje nasz generator.

Parametrami tego testu są długość próbki losowej n, liczba t i poziom istotności α.

Generujemy próbkę losową i zliczamy, jakiej długości serie monotoniczne w niej występują (serie długości większej lub równej t liczymy razem).

Zliczając serie należy pamiętać o tym, aby liczby następującej bezpośrednio za monotoniczną serią nie uwzględniać w obliczeniach, bo nie jest niezależna od poprzednich.

Na otrzymanych wynikach wykonujemy test χ^2.

Ten test sprawdza, jak często pojawiają się w generowanych liczbach wszystkie krotki o ustalonej długości.

Parametrami testu są długość próbki losowej n, liczba k i poziom istotności α.

Generujemy próbkę losową, dzielimy na krotki długości k i zliczamy, ile razy dana krotka wystąpiła.

Na otrzymanych wynikach wykonujemy test χ^2.

Ten test bada, czy pary liczb generowanych przez generator są równomiernie rozmieszczone w kwadracie o boku 1.

Parametrami tego testu są długość próbki losowej n $(2|n)$ oraz dopuszczalny błąd ϵ.

Próbkę losową dzielimy na pary, które interpretujemy jako współrzędne punktów na płaszczyźnie. Wyznaczamy iloraz

$$P = \frac{2N}{n},$$

gdzie N jest liczbą punktów, które zmieściły się w kole o środku w $(1/2, 1/2)$ i promieniu $1/2$.

Test zaliczamy, jeśli $|P - \pi/4| \leq \epsilon$.

1. Jakie jest prawdopodobieństwo tego, że generator idealny wygeneruje wyłącznie liczby parzyste? Czy wynik będzie taki sam dla generatora liniowego kongruencyjnego?
2. Przeprowadź test χ^2 dla generatora liniowego kongruencyjnego.
3. Przeprowadź test Kołmogorowa dla generatora liniowego kongruencyjnego.
4. Wyznacz dystrybuantę statystyki testu sum dla parametru $t = 3$.
5. Wyznacz dystrybuantę statystyki testu kolekcjonera.