Кравцов Г.О., директор ТОВ «НВФ «БКП-консалтинг», м.Київ

NIST 800-22 українською мовою. Набір статистичних тестів для генераторів випадкових та псевдовипадкових чисел для криптографічних додатків.

Метою цієї статті є подання матеріалу NIST 800-22 українською мовою, що сприятиме більш глибокому розумінню викладачами та студентами вищих навчальних закладів України принципів визначення ознак випадковості послідовностей чисел при використанні в криптографічних методах Національного інституту стандартів і технологій Сполучених Штатів Америки (НІСТ США). Оригінал тексту наведено на сайті НІСТ США та доступний за адресою http://csrc.nist.gov/publications/nistpubs/.

Щирі слова подяки від автора **Коростильову Олександру Сергійовичу** за допомогу у підготовці статті.

Статистичні тести НІСТ США використовуються для визначення якісних та кількісних ознак випадковості послідовності чисел. Розуміння математичної суті цих тестів є необхідною для криптографів та криптоаналітиків, особливо при розробці (аналізу) ключової послідовності (гами) шифру Вернама (G.S. Vernam).

В цій статті викладено 16 тестів у тому ж порядку, як в NIST 800-22. Кожен тест подається у розрізі мети, позначення, статистик, опису та правил інтерпретації результатів.

Компанією «БКП-консалтинг» розроблено програмний засіб - криптолабораторію, яка включає в себе всі описані нижче тести.

1. Частотний (монобітний) тест

1. Мета тесту.

   В центрі уваги цього тесту пропорція нулів та одиниць в усій послідовності. Мета тесту – визначити, чи буде кількість нулів та одиниць у послідовності приблизно така ж як у дійсно випадкової послідовності. Тест оцінює наскільки близька пропорція одиниць до ½. Тобто кількість нулів та одиниць в послідовності має бути приблизно однакова.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

3. Статистика тесту та граничний розподіл.

s_{obs} - абсолютна величина суми X_i по всій довжині послідовності, що поділена на корінь квадратний з довжини послідовності. (Тут $X_i = 2 \cdot \varepsilon_i - 1$, $X_i \in \{-1,1\}$).

Граничний розподіл тестової статистики при великих n - напівнормальний. Якщо послідовність випадкова, тоді $+1$ та -1 будуть компенсувати один одного та статистика тесту буде мати значення близькі до нуля. Якщо в послідовності дуже багато нулів або одиниць, тоді значення статистики тесту будуть значно відхилятися від нуля.

4. Опис тесту.

(1) Перетворення до ± 1: всі нулі вхідної послідовності ε замінюємо на (-1), тобто будуємо нову послідовність $X_1, X_2, ..., X_n$, де $X_i = 2 \cdot \varepsilon_i - 1$. Підраховуємо величину $S_n = X_1 + X_2 + ... + X_n$.

Наприклад, якщо $\varepsilon = 1011010101$, тоді $n = 10$ і
$S = 1 + (-1) + 1 + 1 + (-1) + 1 + (-1) + 1 + (-1) + 1 = 2$.

(2) Підраховуємо значення статистики тесту $s_{obs} = \frac{S_n}{\sqrt{n}}$.

Для прикладу з цього пункту, $s_{obs} = \frac{|2|}{\sqrt{10}} = 0.632455532$

(3) Підраховуємо $Pvalue = erfc\left(\frac{s_{obs}}{\sqrt{2}}\right)$, де $erfc$ - комплементарна функція похибки

що визначається таким чином: $erfc(z) = \frac{2}{\sqrt{\pi}} \cdot \int_z^{+\infty} e^{-u^2} du$.

Для прикладу з цього пункту, $Pvalue = erfc\left(\frac{0.632455532}{\sqrt{2}}\right) = 0.527089$

5. Вирішуюче правило (для рівня значущості 1%).

Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

Так як значення $Pvalue$, що отримане в пункті 1.4.(3) є ≥ 0.01 ($Pvalue = 0.527089$), робимо висновок, що послідовність випадкова.

Зауважимо, що якщо значення $Pvalue$ мале (< 0.01), тоді це означає, що значення $|S_n|$ та $|s_{obs}|$ занадто великі. Великі додатні значення S_n свідчать про занадто

велику кількість одиниць у вхідній послідовності, великі від'ємні значення S_n свідчать про занадто велику кількість нулів.

7. Рекомендації по вхідним розмірам.

   Рекомендується, щоб кожна послідовність, що тестується складалася як мінімум зі 100 бітів ($n \geq 100$).

2. Частотний тест по блокам.

1. Мета тесту.

   Увага цього тесту спрямована на пропорцію нулів та одиниць в M -бітових блоках. Мета тесту – визначити, чи буде кількість одиниць в середині кожного блоку приблизно дорівнювати $M/2$, як це очікується від випадкової послідовності. Якщо розмір блоку $M = n$, тоді цей тест перетворюється у частотний тест, розглянутий в параграфі 1.

2. Позначення.

   M - довжина кожного блоку в бітах.

   n - довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

3. Статистика тесту та граничний розподіл.

   $\chi^2(obs)$ - міра того, як добре пропорція одиниць в межах даних M -бітових блоків відповідає пропорції, що очікується за припущенням випадковості послідовності ($1/2$).

   Граничний розподіл такої статистики є χ^2 -розподіл.

4. Опис тесту.

   (1) Ділимо вхідну послідовність на $N = \left\lfloor \frac{n}{M} \right\rfloor$ блоків, що не перетинаються.

   Відкидаємо останні біти, які не утворюють повного M -бітового блоку (якщо такі є).

   Наприклад, якщо $n = 10$, $M = 3$ і $\varepsilon = 0110011010$, тоді будуть утворені такі 3 блоки ($N = 3$): $011, 001, 101$. Останній 0 буде відкинуто.

   (2) Визначаємо пропорцію π_i для кожного з блоків $i = 1, 2, ..., N$ за формулою: $\pi_i = \frac{\sum_{j=1}^{M} \varepsilon_{(i-1)\cdot M + j}}{M}$.

   Для прикладу з цього пункту: $\pi_1 = 2/3, \quad \pi_2 = 1/3, \quad \pi_3 = 2/3$.

(3) Підраховуємо значення статистики: $\chi^2(obs) = 4 \cdot M \cdot \sum_{i=1}^{N} \left(\pi_i - \frac{1}{2} \right)^2$.

Для прикладу з цього пункту: $\chi^2(obs) = 4 \cdot 3 \cdot \left(\left(\frac{2}{3} - \frac{1}{2} \right)^2 + \left(\frac{1}{3} - \frac{1}{2} \right)^2 + \left(\frac{2}{3} - \frac{1}{2} \right)^2 \right) = 1$

(4) Підраховуємо значення $Pvalue = igamc\left(\frac{N}{2}, \frac{\chi^2(obs)}{2} \right)$, де $igamc(\cdot)$ - неповна гамма функція що визначається наступним чином: $igamc(a, b) = \frac{1}{\Gamma(a)} \cdot \int_b^{+\infty} e^{-u} \cdot u^{a-1} du$.

Для прикладу з цього пункту: $Pvalue = igamc\left(\frac{3}{2}, \frac{1}{2} \right) = 0.801252$

5. Вирішуюче правило (для рівня значущості 1%).
   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.
6. Висновки та інтерпретація результатів тесту.
   Так як значення $Pvalue$, яке отримане в 2.4.(4) є ≥ 0.01 ($Pvalue = 0.801252$), робимо висновок, що послідовність ε випадкова.
7. Рекомендації по вхідним розмірам.
   Рекомендується, щоб кожна послідовність, що тестується складалася як мінімум зі 100 бітів ($n \geq 100$). Зауважимо, що $n \geq M \cdot N$. Розмір блоку M має бути таким, що $M \geq 20$, $M > 0.01 \cdot n$, і $N < 100$.

3. Тест серій.

1. Мета тесту.
   Увага в цьому тесті спрямована на загальну кількість серій в усій послідовності. Під серією розуміється неперервна послідовність однакових бітів. Серія довжини k складається рівно з k однакових бітів і обмежена на початку і в кінці бітами протилежного значення. Мета тесту – визначити, чи буде загальна кількість серій з одиниць та нулів різної довжини такою, яка очікується від випадкової послідовності бітів. В частинному випадку цей тест визначає чи є коливання між нулями та одиницями дуже швидке чи дуже повільне.
2. Позначення.
   n – довжина вхідної послідовності в бітах.
   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

3. Статистика тесту та граничний розподіл.

$V_n(obs)$ - загальна кількість серій (тобто загальна кількість серій з одиниць + загальна кількість серій з нулів) по всім n бітам послідовності.

Граничний розподіл статистики є χ^2-розподіл.

4. Опис тесту.

Зауваження: тест серій включає в себе попередній частотний тест.

(1) Для попереднього тесту визначаємо пропорцію π одиниць в послідовності:

$\pi = \frac{1}{n} \cdot \sum_{j=1}^{n} \varepsilon_j$.

Наприклад, якщо $\varepsilon = 1001101011$, тоді $n = 10$ і $\pi = 6/10 = 3/5$.

(2) Визначаємо, чи проходить послідовність попередній тест: Якщо $|\pi - 1/2| \geq \tau$, тоді тест серій не потрібно виконувати (оскільки ця послідовність не проходить частотний тест (параграф 1)). В цьому випадку покладаємо $Pvalue = 0.0000$. Для цього тесту покладено $\tau = \frac{2}{\sqrt{n}}$.

Для прикладу цього пункту: $\tau = \frac{2}{\sqrt{10}} = 0.63246$, тоді $|\pi - 1/2| = |3/5 - 1/2| = 0.1 < \tau$ і послідовність проходить попередній тест. Тому далі переходимо до тесту серій.

(3) Підраховуємо значення статистики тесту: $V_n(obs) = \sum_{k=1}^{n-1} r(k) + 1$, де $r(k) = 0$, якщо $\varepsilon_k = \varepsilon_{k+1}$, і $r(k) = 1$ інакше.

Для прикладу з цього пункту: $V_{10}(obs) = (1+0+1+0+1+1+1+1+0)+1 = 7$

(4) Підраховуємо значення $Pvalue = erfc\left(\frac{|V_n(obs) - 2 \cdot n \cdot \pi \cdot (1-\pi)|}{2 \cdot \sqrt{2 \cdot n} \cdot \pi \cdot (1-\pi)}\right)$.

Для прикладу з цього пункту: $Pvalue = erfc\left(\frac{|7 - 2 \cdot 10 \cdot 3/5 \cdot (1-3/5)|}{2 \cdot \sqrt{2 \cdot 10} \cdot 3/5 \cdot (1-3/5)}\right) = 0.147232$

5. Вирішуюче правило (для рівня значущості 1%).

Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

Так як значення $Pvalue$, що отримане в пункті 3.4.(4) є ≥ 0.01 ($Pvalue = 0.147232$), робимо висновок, що послідовність випадкова.

Зауважимо, що великі значення $V_n(obs)$ свідчать про дуже швидкі коливання між нулями та одиницями, малі значення $V_n(obs)$ свідчать про занадто повільні коливання. (Під коливаннями розуміється переходи від нуля до одиниці і навпаки). Швидкі коливання спостерігаються при великій кількості переходів (наприклад як в послідовності 010101010101). Послідовність з повільними коливаннями має менше серій, ніж очікується від випадкової послідовності. Наприклад послідовність, що містить 100 одиниць за якими слідує 73 нулі а потім 127 одиниць (довжина 300 біт) буде мати всього 3 серії, в той час коли для випадкової послідовності такої ж довжини очікується приблизно 150 серій.

7. Рекомендації по вхідним розмірам.

   Рекомендується, щоб кожна послідовність, що тестується складалася як мінімум зі 100 бітів ($n \geq 100$).

4. Тест найдовшої серії з одиниць.

1. Мета тесту.

   Увага спрямована на найдовшу серію одиниць в межах M -бітових блоків. Мета тесту – визначити, чи відповідає довжина найдовшої серії з одиниць (в послідовності, що тестується) довжині найдовшої серії з одиниць, що очікується в випадковій послідовності. Зауважимо, що відхилення очікуємої довжини найдовшої послідовності з одиниць означає також відхилення очікуємої довжини найдовшої серії з нулів.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

   M - довжина кожного блоку. В коді програми зафіксовані такі можливі значення для M: $M = 8, M = 128, M = 10^4$. Обране M має узгоджуватися з таблицею:

Мінімальне n	M
128	8
6272	128
750000	10^4

   N - кількість блоків; обирається згідно обраного значення M.

3. Статистика тесту та граничний розподіл.

   $\chi^2(obs)$ - міра того, як добре довжина найдовшої серії у межах M -бітових блоків,

що спостерігається співпадає з довжиною найдовшої серії у межах M -бітових блоків за припущенням випадковості.

Граничний розподіл статистики є χ^2 -розподіл.

4. Опис тесту.

(1) Ділимо ввідну послідовність на M -бітові блоки.

(2) Для кожного блоку знаходимо довжину найдовшої серії з одиниць $l_i^{\max}$, $i = 1, 2, ..., N$. Набір всіх цих довжин $\{l_i^{\max}\}_{i=1}^{N}$ розіб'ємо на класи V_k і будемо підраховувати кількість ν_k елементів, що потрапили до кожного класу. В залежності від M кількість класів, і умови, за якими довжини відносяться до того чи іншого класу приведені в наступній таблиці:

Клас V_k	$M = 8$	$M = 128$	$M = 10^4$
V_0	≤ 1	≤ 4	≤ 10
V_1	2	5	11
V_2	3	6	12
V_3	≥ 4	7	13
V_4		8	14
V_5		≥ 9	15
V_6			≥ 16

(3) Підраховуємо статистику $\chi^2(obs) = \sum_{i=0}^{K} \frac{(\nu_i - N \cdot \pi_i)^2}{N \cdot \pi_i}$, де величини π_i (наведені в NIST 800-22 в п.3.4. на сторінці 68). Величини K і N визначаються за M згідно наступної таблиці:

M	K	N
8	3	16
128	5	49
10^4	6	75

Для прикладу з пункту 4.8:

$$\chi^2(obs) = \frac{(4 - 16 \cdot 0.2148)^2}{16 \cdot 0.2148} + \frac{(9 - 16 \cdot 0.3672)^2}{16 \cdot 0.3672} + \frac{(3 - 16 \cdot 0.2305)^2}{16 \cdot 0.2305} + \frac{(0 - 16 \cdot 0.1875)^2}{16 \cdot 0.1875} = 4.882605$$

(4) Підраховуємо значення $Pvalue = igamc\left(\frac{K}{2}, \frac{\chi^2(obs)}{2}\right)$.

Для прикладу з пункту 4.8: $Pvalue = igamc\left(\frac{3}{2}, \frac{4.882605}{2}\right) = 0.180598$

5. Вирішуюче правило (для рівня значущості 1%).

   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 4.4.(4) є ≥ 0.01 ($Pvalue = 0.180598$), робимо висновок, що послідовність випадкова. Зауважимо, що великі значення $\chi^2(obs)$ свідчать про те, що послідовність має кластери одиниць.

7. Рекомендації по вхідним розмірам.

   Рекомендується, щоб кожна послідовність, що тестується складалася як мінімум з такої кількості бітів, що вказана в пункті 4.2.

8. Приклад.

   Для випадку $K = 3$ і $M = 8$.

   Нехай

   $\varepsilon = 11001100000101010110100010011001111000000000001001$
   $00110101010001000100111101011010000000110101111100$
   $11001110011011011000010110010$

   Тоді $n = 128$. Отримаємо такі блоки:

Блок	Довжина найдовшої серії з одиниць	Блок	Довжина найдовшої серії з одиниць
11001100	2	00010101	1
01101100	2	01001100	2
11100000	3	00000010	1
01001101	2	01010001	1
00010011	2	11010110	2
10000000	1	11010111	3
11001100	2	11100110	3
11011000	2	10110010	2

Підраховуємо кількості ν_k: $\nu_0 = 4, \nu_1 = 9, \nu_2 = 3, \nu_3 = 0$

Статистика $\chi^2(obs) = 4.882605$, $Pvalue = 0.180598$.

Так як $Pvalue = 0.180598 \geq 0.01$ робимо висновок, що послідовність випадкова.

5. Тест рангу бінарних матриць.

1. Мета тесту.

   Увага тесту зосереджена на ранзі матриць, що утворені з під послідовностей вхідної послідовності, які йдуть одна за другою. Мета тесту – зафіксувати лінійну залежність серед підрядків фіксованої довжини вхідної послідовності.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

   M - кількість рядків в кожній матриці. У програмі M фіксоване і дорівнює 32. Для інших M необхідно переобчислити наближення.

   Q - кількість колонок в кожній матриці. У програмі Q фіксоване і дорівнює 32. Для інших Q необхідно переобчислити наближення.

3. Статистика тесту та граничний розподіл.

   $\chi^2(obs)$ - міра того, як добре кількість рангів різного порядку у вхідній послідовності наближується до кількості рангів відповідного порядку за умови випадковості послідовності.

   Граничний розподіл статистики є χ^2-розподіл.

4. Опис тесту.

   (1) Ділимо вхідну послідовність на блоки довжиною $M \cdot Q$ біт, що йдуть один за іншим. Отримаємо $N = \left\lfloor \frac{n}{M \cdot Q} \right\rfloor$ блоків. Останні біти, що не утворюють повного блоку відкидаємо. Утворюємо з кожного блоку матрицю розмірності $M \times Q$. Кожний рядок матриці складається з Q-бітового блоку вхідної послідовності. Наприклад, якщо $n = 20$, $M = Q = 3$ і $\varepsilon = 01011001001010101101$, розділяємо ε на $N = \left\lfloor \frac{20}{3 \cdot 3} \right\rfloor = 2$ матриці розмірності 3×3. При цьому останні 2 біти (01) відкидаємо. Матриці будуть

такі: $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ та $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$. Перша матриця складається з перших трьох бітів, що записуються в перший рядок, других трьох бітів, що записуються у другий рядок і третіх трьох бітів, що записані в третьому рядку. Друга матриця утворюється подібним чином але для наступних дев'яти бітів.

(2) Визначаємо бінарний ранг кожної матриці R_l, $l = 1, 2, ..., N$. Метод визначення бінарного рангу матриці описаний в додатку А.

Для прикладу з цього пункту, ранг першої матриці дорівнює 2 ($R_1 = 1$) і ранг другої матриці дорівнює 3 ($R_2 = 3$).

(3) Позначимо через F_M кількість матриць, що має ранг $R_l = M$ (повний ранг), через F_{M-1} - кількість матриць з рангом $R_l = M - 1$, тоді $N - F_M - F_{M-1}$ - кількість матриць з рангом, меншим за $M - 1$.

Для прикладу з цього пункту, $F_M = F_3 = 1$, $F_{M-1} = F_2 = 1$, і не існує матриць з меншим рангом.

(4) Підраховуємо статистику

$$\chi^2(obs) = \frac{(F_M - 0.2888 \cdot N)^2}{0.2888 \cdot N} + \frac{(F_{M-1} - 0.5776 \cdot N)^2}{0.5776 \cdot N} + \frac{(N - F_M - F_{M-1} - 0.1336 \cdot N)^2}{0.1336 \cdot N}$$

Для прикладу з цього пункту,

$$\chi^2(obs) = \frac{(1 - 0.2888 \cdot 2)^2}{0.2888 \cdot 2} + \frac{(1 - 0.5776 \cdot 2)^2}{0.5776 \cdot 2} + \frac{(2 - 1 - 1 - 0.1336 \cdot 2)^2}{0.1336 \cdot 2} = 0.596953$$

(5) Підраховуємо $Pvalue = e^{-\chi^2(obs)/2}$. Так як в прикладі $M = 3$,

$$Pvalue = igamc\left(\frac{3}{2}, \frac{\chi^2(obs)}{2}\right).$$

Для прикладу з цього пункту, $Pvalue = e^{-0.596953/2} = 0.741948$.

5. Вирішуюче правило (для рівня значущості 1%).

Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 5.4.(5) є ≥ 0.01 ($Pvalue = 0.741948$), робимо висновок, що послідовність випадкова.

   Зауважимо, що великі значення $\chi^2(obs)$ (і, як наслідок, мале значення $Pvalue$) свідчить про відхилення розподілу рангів від розподілу, що має випадкова послідовність.

7. Рекомендації по вхідним розмірам.

   Ймовірності для $M = Q = 32$ підраховані і зафіксовані у програмі. Можна обрати і інші значення для M та Q, але при цьому необхідно перерахувати ймовірності. Мінімальна довжина послідовності має бути така, що $n \geq 38 \cdot M \cdot Q$ (тобто, як мінімум повинно бути 38 матриць). Для $M = Q = 32$, послідовність має складатися як мінімум з 38912 біт.

6. Тест на основі дискретного перетворення Фур'є.

1. Мета тесту.

   Увага тесту зосереджена на висоті відліків дискретного перетворення Фур'є (ДПФ) вхідної послідовності. Мета тесту – виявити періодичний характер (тобто визначити наявність фрагментів, схожих між собою і періодично ???) вхідної послідовності, який буде свідчити про відхилення від припущення випадковості. Для цього з'ясовується, чи буде кількість відліків ДПФ з висотою більшою за 95% поріг значно відрізнятися від 5% від загальної кількості відліків. Під висотою відліку розуміється модуль компоненти ДПФ (яка в загальному випадку є комплексним числом).

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

3. Статистика тесту та граничний розподіл.

   d - нормалізована різниця між спостерігаємою та очікуємою кількістю частотних компонент (відліків ДПФ), що перевищують за модулем 95% поріг.

   Граничний розподіл статистики – нормальний розподіл.

4. Опис тесту.

   (1) Нулі та одиниці вхідної послідовності перетворюємо на (-1) та (+1) відповідно, утворюючи таким чином послідовність $X = x_1, x_2, ..., x_n$, де

$x_i = 2 \cdot \varepsilon_i - 1, \quad i = 1,2,...,n$.

Наприклад, якщо $n = 10$ і $\varepsilon = 1001010011$, тоді $X = 1,-1,-1,1,-1,1,-1,-1,1,1$

(2) Застосовуємо ДПФ до послідовності X і отримуємо послідовність S комплексних чисел: $S = ДПФ(\mathrm{X})$. Ця послідовність характеризує періодичні компоненти, що присутні у вхідній послідовності з різними частотами.

(3) Підраховуємо модулі комплексних чисел послідовності S утворюючи послідовність M: $M = |S'|$, де S' - послідовність з перших $n/2$ елементів послідовності S.

(4) Підраховуємо величину $T = \sqrt{3 \cdot n}$ - 95% поріг для висот відліків ДПФ. При справедливості припущення про випадковість вхідної послідовності 95% відліків її ДПФ не повинні перевищувати величини T.

(5) Підраховуємо $N_0 = 0.95 \cdot n/2$. N_0 - теоретична кількість відліків модуль яких менше за T (при умові випадковості послідовності).

Для прикладу з цього пункту, $N_0 = 4.75$.

(6) Підраховуємо N_1 - кількість елементів послідовності M, що менші за T.

Для прикладу з цього пункту, $N_1 = 4$.

(7) Підраховуємо статистику $d = \dfrac{N_1 - N_0}{\sqrt{n \cdot 0.95 \cdot 0.05/2}}$

Для прикладу з цього пункту, $d = \dfrac{4 - 4.75}{\sqrt{10 \cdot 0.95 \cdot 0.5/2}} = -1.538968$.

(8) Підраховуємо $Pvalue = erfc\left(\dfrac{|d|}{\sqrt{2}}\right)$.

Для прикладу з цього пункту, $Pvalue = erfc\left(\dfrac{1.538968}{\sqrt{2}}\right) = 0.123812$.

5. Вирішуюче правило (для рівня значущості 1%).

Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

Так як значення $Pvalue$, що отримане в пункті 6.4.(8) є ≥ 0.01 ($Pvalue = 0.123812$), робимо висновок, що послідовність випадкова.

Занадто малі значення d свідчать про дуже малу кількість (<95%) відліків з модулем меншим за T, і велику кількість (>5%) з модулем більшим за T.

7. Рекомендації по вхідним розмірам.

   Рекомендується, щоб кожна послідовність, що тестується складалася мінімум з 1000 бітів ($n \geq 1000$).

7. Тест на співпадіння з шаблоном без перекриття.

1. Мета тесту.

   Увага тесту зосереджена на тому, яку кількість разів зустрічається наперед заданий рядок у вхідній послідовності. Мета тесту – виявити генератор, який формує послідовність, що містить дуже велику кількість заданого неперіодичного (аперіодичного) шаблону. В цьому тесті, для пошуку заданого m -бітного шаблону використовується m -бітне вікно. Вікно зсувається вправо на один біт, якщо під ним не спостерігається заданий шаблон, і зсувається вправо на m біт, якщо послідовність, що в ньому знаходиться співпадає з шаблоном.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

   m - довжина шаблону в бітах.

   B - m -бітний шаблон, пошук якого проводиться. Фактично, це рядок з нулів та одиниць.

   N - кількість незалежних блоків. У програмі N фіксоване $N = 8$.

   M - довжина блоків в бітах, на які розбивається вхідна послідовність (визначається згідно N та n).

3. Статистика тесту та граничний розподіл.

   $\chi^2(obs)$ - міра того, як добре кількість «співпадінь» з шаблоном у вхідній послідовності відповідає очікуємій кількості «співпадінь» за умови випадковості послідовності.

   Граничний розподіл статистики – χ^2 -розподіл.

4. Опис тесту.

   (1) Розділяємо вхідну послідовність ε на N блоків довжини M .
   Наприклад, якщо $\varepsilon = 10100100101110010110$, тоді $n = 20$. Якщо $N = 2$ та $M = 10$, тоді буде сформовано два блоки: 1010010010 і 1110010110 .

   (2) Позначимо через W_j $(j = 1, 2, ..., N)$ кількість разів, яку шаблон B зустрічається в блоці номер j . Пошук проводиться за допомогою m -

бітового вікна, що пересувається вздовж послідовності. Якщо виділена вікном підпослідовність бітів співпадає з шаблоном B, тоді зсуваємо вікно на m біт праворуч; якщо ж не співпадає – на один біт праворуч. Наприклад, при $m = 3$, якщо вікно містить біти 3, 4, 5 та вони співпадають з шаблоном, то після зсуву вікна воно буде містити біти 6, 7, 8 вхідної послідовності; якщо ж спів падання немає, то після зсуву вікно буде містити біти 4, 5, 6. Для прикладу з цього пункту якщо $m = 3$ і $B = 001$ процес пошуку буде проходити наступним чином:

Розташування вікна, біти	Блок 1		Блок 2	
	Біти під вікном	W_1	Біти під вікном	W_2
1-3	101	0	111	0
2-4	010	0	110	0
3-5	100	0	100	0
4-6	001	1	001	1
5-7	Не розгляд.		Не розгляд.	
6-8	Не розгляд.		Не розгляд.	
7-9	001	2	011	1
8-10	Не розгляд.		110	1

Таким чином отримали, що $W_1 = 2$ і $W_2 = 1$.

(3) За умови випадковості, підраховуємо теоретичне середнє μ та дисперсію σ^2: $\mu = \frac{M - m + 1}{2^m}$, $\sigma^2 = M \cdot \left(\frac{1}{2^m} - \frac{2 \cdot m - 1}{2^{2 \cdot m}} \right)$.

Для прикладу з цього пункту, $\mu = \frac{10 - 3 + 1}{2^3} = 1$ і

$$\sigma^2 = 10 \cdot \left(\frac{1}{2^3} - \frac{2 \cdot 3 - 1}{2^{2 \cdot 3}} \right) = 0.46875$$

(4) Підраховуємо статистику $\chi^2(obs) = \sum_{j=1}^{N} \frac{(W_j - \mu)}{\sigma^2}$.

Для прикладу з цього пункту, $\chi^2(obs) = \frac{(2-1)^2 + (1-1)^2}{0.46875} = 2.133333$.

(5) Підраховуємо $Pvalue = igamc\left(\frac{N}{2}, \frac{\chi^2(obs)}{2} \right)$. Зауважимо, що необхідно обчислити стільки величин $Pvalue$, скільки різних шаблонів

використовується в тесті – на кожний шаблон відповідна $Pvalue$. Для $m = 9$ можливі до 148 значень $Pvalue$; для $m = 10$ - до 284.

Для прикладу з цього пункту, $Pvalue = igamc\left(\frac{2}{2}, \frac{2.133333}{2}\right) = 0.344154$

5. Вирішуюче правило (для рівня значущості 1%).

   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 7.4.(5) є ≥ 0.01 ($Pvalue = 0.344154$), робимо висновок, що послідовність випадкова.

   Якщо $Pvalue$ дуже мале (<0.01), тоді вхідна послідовність містить таку кількість шаблонних підпослідовностей, що не узгоджується з гіпотезою про випадковість.

7. Рекомендації по вхідним розмірам.

   Програма дозволяє шукати шаблони довжиною $m = 2, 3, ..., 32$. Рекомендовано використовувати $m = 9$ або $m = 10$ для отримання значущих результатів. Крім того $N = 8$ і зафіксовано у програмі, хоча можна використовувати і інші значення. Однак для забезпечення достовірності отриманих $Pvalue$, N має задовольняти умову: $N \leq 100$. M та N мають бути обраними так, щоб $M > 0.01 \cdot n$ та $N = \left\lfloor \frac{n}{M} \right\rfloor$.

8. Тест шаблонов з перекриттям.

1. Мета тесту.

   Тест оснований на аналізі кількості, яку наперед заданий шаблон зустрічається у вхідній послідовності. Так само як і попередній, для пошуку m-бітного шаблону цей тест використовує m-бітне вікно. Якщо виділений вікном підрядок загальної послідовності не співпадає з заданим шаблоном, вікно зсувається на один біт праворуч. На відміну від попереднього тесту, якщо навіть цей підрядок співпадає з шаблонним, вікно все одно зсувається на *один* біт праворуч. Після зсуву вікна пошук продовжується аналогічно.

2. Позначення

   m - довжина шаблонного рядка в бітах.

   n - довжина вхідної послідовності.

   ε - вхідна послідовність бітів, $\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$

   B - шаблонний рядок (шаблон).

K - кількість степенів свободи. У програмі зафіксовано $K = 5$.

N - кількість незалежних блоків, на які поділяється послідовність ε.

M - довжина кожного такого блоку.

3. Статистика тесту та граничний розподіл

$\chi^2(obs)$ - міра того, як добре кількість, яку заданий шаблон зустрічається у послідовності ε відповідає цій кількості але за умови випадковості послідовності. Граничний розподіл статистики - χ^2-розподіл.

4. Опис тесту.

(1) Поділяємо послідовність ε на N блоків довжиною M біт кожний. При цьому останні біти послідовності, що не утворюють цілий блок відкидаються.

Наприклад, якщо

$\varepsilon = 10111011110010110100011100101110111110000101101001$, $n = 50$

$K = 2$, $M = 10$ і $N = 5$,

тоді буде створено 5 блоків: 1011101111, 0010110100, 0111001011, 1011111000, і 0101101001.

(2) Підраховуємо кількість, яку зустрічається шаблон у кожному з N блоків. Для кожного блоку кількість шукається окремо. Перед аналізом нового блоку лічильник встановлюємо в нуль. При пошуку шаблона створюється m-бітне вікно, що пересувається вздовж вхідної послідовності. Якщо виділена вікном послідовність біт співпадає з шаблоном, тоді лічильник збільшуємо на одиницю. Не залежно від результату порівняння, вікно зсувається на один біт праворуч. В залежності від значення лічильника після проходження блоку збільшуємо на одиницю одну з величин ν_i, $i = 0,1,...K$. Якщо значення лічильника дорівнює нулю, тоді збільшуємо ν_0, якщо одиниці - ν_1, і так далі, якщо значення більше рівне п'яти, тоді збільшуємо на одиницю ν_5.

Для прикладу з цього пункту, якщо $m = 2$ і $B = 11$ процес пошуку по першому блоку 1011101111 можна зобразити таблицею:

Позиція бітів	Біти	Кількість «зустрічей» шаблону $B = 11$
1-2	10	0
2-3	01	0

3-4	11 (співпадіння)	1
4-5	11 (співпадіння)	2
5-6	10	2
6-7	01	2
7-8	11 (співпадіння)	3
8-9	11 (співпадіння)	4
9-10	11 (співпадіння)	5

Після проходження першого блоку маємо 5 разів зустрітий шаблон B.

Збільшуємо на один ν_5 і отримуємо: $\nu_0 = \nu_1 = ... = \nu_4 = 0$, $\nu_5 = 1$

Подібним чином проходимо по всім іншим блокам. В блоці 2 маємо 2 рази зустрітий шаблон B, в блоці 3 – 3 рази, в блоці 4 – 4 рази, в блоці 5 – 1 раз. Після проходження всіх блоків маємо:

$\nu_0 = 0,\ \nu_1 = 1,\ \nu_2 = 1,\ \nu_3 = 1,\ \nu_4 = 1,\ \nu_5 = 1$

(3) Підраховуємо значення λ і η, які будуть використовуватися при визначення теоретичних ймовірностей π_i, що відповідають класам ν_i:

$$\lambda = \frac{M - m + 1}{2^m},\ \eta = \frac{\lambda}{2}.$$

Для прикладу з цього пункту, $\lambda = \frac{10 - 2 + 1}{2^2} = 2.25$, $\eta = \frac{2.25}{2} = 1.125$.

(4) Підраховуємо статистику $\chi^2(obs) = \sum_{i=0}^{K} \frac{(\nu_i - N \cdot \pi_i)^2}{N \cdot \pi_i}$, де $\pi_0 = 0.367879$,

$\pi_1 = 0.183940$, $\pi_2 = 0.137955$, $\pi_3 = 0.099634$, $\pi_4 = 0.077147$,

$\pi_5 = 0.140657$. (Як визначати π_i викладено у 3.8.)

Для прикладу цього пункту величини π_i були пере обчислені так як приклад не задовольняє умовам параграфу 3.8. Тому використовуються такі значення: $\pi_0 = 0.324652$, $\pi_1 = 0.182617$, $\pi_2 = 0.142670$, $\pi_3 = 0.106645$, $\pi_4 = 0.077147$, $\pi_5 = 0.166269$. Тоді значення статистики

$$\chi^2(obs) = \frac{(0 - 5 \cdot 0.324652)^2}{5 \cdot 0.324652} + \frac{(1 - 5 \cdot 0.182617)^2}{5 \cdot 0.182617} + \frac{(1 - 5 \cdot 0.142670)^2}{5 \cdot 0.142670} +$$
$$+ \frac{(1 - 5 \cdot 0.106645)^2}{5 \cdot 0.106645} + \frac{(1 - 5 \cdot 0.077147)^2}{5 \cdot 0.077147} + \frac{(1 - 5 \cdot 0.166269)^2}{5 \cdot 0.166269} = 3.167729$$

(5) Підраховуємо величину $Pvalue = igamc\left(\frac{K}{2}, \frac{\chi^2(obs)}{2}\right)$

Для прикладу з цього пункту, $Pvalue = igamc\left(\frac{5}{2}, \frac{3.167729}{2}\right) = 0.274932$.

5. Вирішуючи правило (для рівня значущості 1%).

   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 8.4.(5) є ≥ 0.01 ($Pvalue = 0.274932$), робимо висновок, що послідовність випадкова.

   Зауважимо, що для 2-бітного шаблону ($B = 11$), якщо послідовність має дуже багато двобітних серій з одиниць, тоді: 1) v_5 буде дуже велике 2) тестова статистика буде великою 3) $Pvalue$ буде малим (< 0.01) 4) буде зроблен висновок про не випадковість послідовності.

7. Рекомендації по вхідним розмірам.

   Значення K, N, M мають бути обраними так, щоб вхідна послідовність складалася як мінімум з 10^6 біт ($n \geq 10^6$). Можуть бути обрані різні значення для параметру m, але NIST рекомендує обирати $m = 9,\ m = 10$. При бажанні використовувати інші значення параметрів, необхідно додержуватися вимог:

   - $n \geq M \cdot N$
   - N обирається так, щоб $N \cdot \min(\pi_0, \pi_1, ..., \pi_K) > 5$
   - $\lambda = (M - m + 1)/2^2 \approx 2$
   - m обирається так, щоб $m \approx \log_2 M$
   - K обирається так, щоб $K \approx 2 \cdot \lambda$.

   Для $K \neq 5$ необхідно пере обчислити значення π_i, $i = 1, 2, ..., K$ (див. 3.8)

9. Універсальний статистичний тест Мауера.

1. Мета тесту.

   Тест оснований на підрахунку кількості біт, що розташовані між співпадаючими шаблонами (міра, яка відноситься до довжини стиснутої послідовності). Мета тесту – визначити, чи може вхідна послідовність бути значно стиснута без втрати інформації. Послідовність, що значно стискується важається не випадковою.

2. Позначення.

n – довжина вхідної послідовності в бітах.

ε - вхідна послідовність біт, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

L - довжина кожного блоку, на які поділяється вхідна послідовність.

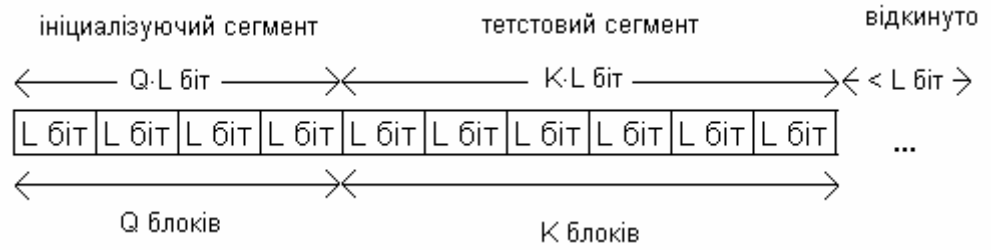
Q - кількість блоків, що утворюють ініціалізуючий сегмент.

3. Статистика тесту та граничний розподіл.

f_n - сума двійкових логарифмів відстаней між співпадаючими L-бітними шаблонами. Відстанню вважається кількість біт між L-бітними шаблонами. Граничний розподіл – напівнормальний.

4. Опис тесту.

(1) Вхідна n-бітна послідовність ε розбивається на два сегменти: ініціалізуючий, який складається з Q L-бітний блоків, що не перекриваються, та тестовий сегмент, який складається з K L-бітний блоків, що не перекриваються. Останні біти послідовності, що не утворюють цілий L-бітний блок відкидаються. (Див. мал.)



Перші Q блоків використовуються для ініціалізації тесту. Наступні K блоків тестуються. ($K = \left\lfloor \frac{n}{L} \right\rfloor - Q$).

Наприклад, якщо $\varepsilon = 01011010011101010111$, $n = 20$, $L = 2$, $Q = 4$, тоді $K = \left\lfloor \frac{20}{2} \right\rfloor - 4 = 6$. Ініціалізуючий сегмент 01011010, тестовий сегмент 011101010111. L-бітні блоки зображені в наступній таблиці:

Блок	Тип	Вміст
1	Ініціалізуючий сегмент	01
2		01
3		10
4		10

5	Тестовий сегмент	01
6		11
7		01
8		01
9		01
10		11

(2) Використовуючи ініціалізуючий сегмент, створюємо таблицю для всіх можливих L-бітних слів (використовуємо десяткове значення L-бітного слова як індекс у таблиці). В таблицю заносимо номер блока в якому востаннє зустрічається відповідне номеру L-бітне слово (тобто для всіх $i = 1, 2, ..., Q$ $T_j = i$, де j - десяткове значення i-ого L-бітного слова).

Для прикладу з цього пункту по першим 4 блокам створена наступна таблиця:

	Можливі L-бітні значення			
	00 (зберіг. в T_0)	01 (зберіг. в T_1)	10 (зберіг. в T_2)	11 (зберіг. в T_3)
Ініціалізоване значення	0	2	4	0

(3) Аналізуємо кожний з K блоків тестового сегменту і визначаємо, скільки блоків пройдено з останнього блоку, що містив таке ж саме слово як і в поточному блоці (тобто обчислюємо відстань $i - T_j$). Додаємо двійковий логарифм від обчисленої відстані до накопичувача суми $sum = sum + \log_2(i - T_j)$. Замінюємо значення в таблиці номером поточного блоку ($T_j = i$).

Для прикладу з цього пункту таблиця і сума, що накопичується змінюється таким чином:

Для блоку 5 (перший тестовий блок): вміст - 01, в таблиці це «01-й» рядок, тобто T_1, $sum = \log_2(5 - 2) = 1.584962501$

Для блоку 6 (другий тестовий блок): вміст - 11, в таблиці це «11-й» рядок, тобто T_3, $sum = 1.584962501 + \log_2(6 - 0) = 4.169925002$

Для блоку 7: вміст - 01, в таблиці це «01-й» рядок, тобто T_1, $sum = 4.169925002 + \log_2(7 - 5) = 5.169925002$

Для блоку 8: вміст - 01, в таблиці це «01-й» рядок, тобто T_1,

$sum = 5.169925002 + \log_2(8-7) = 5.169925002$

Для блоку 9: вміст - 01, в таблиці це «01-й» рядок, тобто T_1,

$sum = 5.169925002 + \log_2(9-8) = 5.169925002$

Для блоку 10: вміст - 11, в таблиці це «11-й» рядок, тобто T_3,

$sum = 5.169925002 + \log_2(10-6) = 7.169925002$

Стан таблиці:

Номер поточного блоку	**Можливі L-бітні значення**			
	00	**01**	**10**	**11**
4	0	2	4	0
5	0	5	4	0
6	0	5	4	6
7	0	7	4	6
8	0	8	4	6
9	0	9	4	6
10	0	9	4	10

(4) Підраховуємо статистику тесту: $f_n = \frac{1}{K} \cdot \sum_{i=Q+1}^{Q+K} \log_2(i - T_j)$.

Для прикладу з цього пункту, $f_n = \frac{7.169925002}{6} = 1.1949875$

(5) Підраховуємо значення $Pvalue = erfc\left(\left|\frac{f_n - ОчікуємеЗначення(\mathrm{L})}{\sqrt{2}\cdot\sigma}\right|\right)$, де

$erfc(z) = \frac{2}{\sqrt{\pi}} \cdot \int_z^{+\infty} e^{-u^2}\,du$, і $ОчікуємеЗначення(\mathrm{L})$ та σ беруться з наведених нижче таблиць. За припущенням випадковості послідовності, вибіркове середнє $ОчікуємеЗначення(\mathrm{L})$, це теоретично очікуєме значення обчисленої статистики для заданого значення L. Теоретичне стандартне відхилення визначається за формулою:

$\sigma = c \cdot \sqrt{\frac{\mathrm{var}(L)}{K}}$, де $c = 0.7 - \frac{0.8}{L} + \left(4 + \frac{32}{L}\right) \cdot \frac{K^{-3/L}}{15}$.

L	*ОчікуємеЗначення*	var
6	5,2177052	2,954
7	6,1962507	3,125
8	7,1836656	3,238
9	8,1764248	3,311
10	9,1723243	3,356
11	10,170032	3,384
12	11,168765	3,401
13	12,168070	3,410
14	13,167693	3,416
15	14,167488	3,419
16	15,167379	3421

Для прикладу з цього пункту,

$$Pvalue = erfc\left(\left|\frac{1.1949875 - 1.5374383}{\sqrt{2}\cdot\sqrt{1.338}}\right|\right) = 0.767189$$

Зауважимо, що *ОчікуємеЗначення* для $L = 2$ не вказано у наведених таблицях, оскільки $L = 2$ не рекомендовано використовувати при тестуванні.

5. Вирішуюче правило (для рівня значущості 1%).

   Якщо підраховане значення *Pvalue* менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення *Pvalue*, що отримане в пункті 9.4.(5) є ≥ 0.01 ($Pvalue = 0.767189$), робимо висновок, що послідовність випадкова.

   Якщо f_n дуже відрізняється від *ОчікуємеЗначення*(L), тоді послідовність може бути значно стиснута.

7. Рекомендації по вхідним розмірам.

   Тест вимагає довгої послідовності бітів ($n \geq (Q + K)\cdot L$, що поділяється на два сегменти L-бітних блоків. L має бути обраним так, що $6 \leq L \leq 16$. Перший сегмент складається з Q блоків, Q має бути таким, що $Q = 10\cdot 2^L$. Другий сегмент складається з K блоків, де $K = \left\lfloor \frac{n}{L} \right\rfloor - Q \approx 1000\cdot 2^L$. Значення L, Q, n мають узгоджуватися з таблицею:

n	L	$Q = 10 \cdot 2^L$
387840	6	640
904960	7	1280
2068480	8	2560
4654080	9	5120
10342400	10	10240
22753280	11	20480
49643520	12	40960
107560960	13	81920
231669760	14	163840
496435200	15	327680
1059061760	16	655360

10. Тест на основі стискання Лемпеля-Зіва.

1. Мета тесту.

   Тест оснований на загальній кількості різних шаблонів (слів), що містяться у вхідній послідовності. Мета тесту – визначити наскільки можна стиснути послідовність. Послідовність вважається не випадковою, якщо її можна значно стиснути.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - вхідна послідовність біт, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

3. Статистика тесту та граничний розподіл.

   W_{obs} - кількість послідовно приєднаних один до іншого різних слів у послідовності.

   Граничний розподіл – нормальний.

4. Опис тесту.

   (1) Розбиваємо послідовність слідуючі один за одним слова, які утворюють «словник» послідовності. Це досягається шляхом створення підрядка з послідовно стоячих бітів, який збільшується (за рахунок приєднання наступних бітів послідовності) до тих пір, поки не утвориться слово, якого ще немає у «словнику». Отриманий підрядок – нове слово, що додається у «словник». Кількість таких слів у «словнику», що отриманий після аналізу всієї послідовності і позначено через W_{obs} .

   Наприклад, якщо $\varepsilon = 010110010$ процес створення словника має вигляд:

Номер біту	Біт	Нове слово?	Нове слово
1	0	Так	0 (біт 1)
2	1	Так	1 (біт 2)
3	0	Ні	
4	1	Так	01 (біти 3-4)
5	1	Ні	
6	0	Так	10 (біти 5-6)
7	0	Ні	
8	1	Ні	
9	0	Так	010 (біти 7-9)

І отримуємо 5 слів у «словнику»: 0, 1, 01, 10, 010. Звідси $W_{obs} = 5$.

(2) Підраховуємо значення $Pvalue = \frac{1}{2} \cdot erfc\left(\frac{\mu - W_{obs}}{\sqrt{2 \cdot \sigma^2}} \right)$, де

$erfc(z) = \frac{2}{\sqrt{\pi}} \cdot \int_z^{+\infty} e^{-u^2} du$, а при $n = 10^6$ маємо $\mu = 69586.25$ та

$\sigma = \sqrt{70.448718}$.

Для інших значень n необхідно знов обчислити параметри μ та σ.

Зауважимо, що на сьогодні не існує теорії, яка дозволяє точно обчислити ці параметри. Для отримання наведених вище значень використовувався (з відповідним припущенням про випадковість) SHA-1. При використанні генератора Blum-Blum-Shub отримані аналогічні значення.

Так як послідовність, що використовується у прикладі цього пункту занадто мала (відносно тої довжини, що рекомендована), наведені значення для μ та σ не можна використовувати. Тому, вважаючи, що тест був застосований до послідовності довжиною $n = 10^6$ біт і що була отримана величина $W_{obs} = 69600$, підрахуємо значення

$$Pvalue = \frac{1}{2} \cdot erfc\left(\frac{69586.25 - 69600}{\sqrt{2 \cdot 70.448718}} \right) = 0.949310$$

5. Вирішуюче правило (для рівня значущості 1%).

   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 10.4.(2) є ≥ 0.01 ($Pvalue = 0.949310$), робимо висновок, що послідовність випадкова.

   Зауважимо, що при $n = 10^6$, якщо W_{obs} виявиться менше за 69.561, тоді буде зроблено висновок, що послідовність є такою, що значно стискається і тому не є випадковою.

7. Рекомендації по вхідним розмірам.

   Рекомендується, щоб кожна послідовність, що тестується складалася як мінімум з 10^6 бітів ($n \geq 10^6$).

11. Тест лінійної складності.

1. Мета тесту.

   Тест оснований на підрахунку довжини лінійного зсувного регістру (ЛЗР). Мета тесту – визначити, чи є вхідна послідовність достатньо складною, щоб вважатися випадковою. Випадкові послідовності характеризуються довгими ЛЗР. Короткі ЛЗР свідчать про не випадковість.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

   M - довжина блоків, на які розбивається вхідна послідовність (в бітах).

   K - кількість степенів свободи. У програмі зафіксовано $K = 6$.

3. Статистика тесту та граничний розподіл.

   $\chi^2(obs)$ - міра того, як добре та кількість ЛЗР фіксованої довжини, що спостерігається в послідовності ε, відповідає кількості ЛЗР фіксованої довжини у випадковій послідовності.

   Граничний розподіл статистики тесту χ^2-розподіл.

4. Опис тесту.

   (1) Ділимо вхідну послідовність на N блоків довжини M, $N = \left\lfloor \frac{n}{M} \right\rfloor$.

   Останні біти, що не утворюють цілого блоку відкидаються.

(2) Використовуючи алгоритм Берлекампа-Месі (Berlekamp-Massey) визначаємо лінійну складність L_i кожного з N блоків ($i = 1,2,...,N$). L_i - це довжина найкоротшої послідовності бітів блоку i яка дозволяє отримати всі біти блоку i. Тобто для будь-якої L_i-бітної підпослідовності блоку i можна обрати комбінацію її біт, що при додаванні їх за модулем 2 отримаємо наступний $(L_i + 1)$-й біт послідовності. (Набір позицій бітів, що входять в цю комбінацію має бути однаковим для всіх підпослідовностей).

Наприклад, якщо $M = 13$ і блок, що буде тестуватися 1101011110001, тоді $L_i = 4$ і послідовність блоку можна отримати шляхом додавання першого та другого бітів в межах 4-бітової підпослідовності і отримування таким чином наступного (5-го) біту. Процес аналізу блоку зображено в наступній таблиці:

	Біт 1	**Біт 2**	**Біт 3**	**Біт 4**	**Біт 5**
Біти 1-4 і результ. 5-й:	1	1	0	1	0
Біти 2-5 і результ. 6-й:	1	0	1	0	1
Біти 3-6 і результ. 7-й:	0	1	0	1	1
.	1	0	1	1	1
.	0	1	1	1	1
.	1	1	1	1	0
.	1	1	1	0	0
.	1	1	0	0	0
Біти 9-12 і результ. 13-й	1	0	0	0	1

Для цього блоку працює тривіальний алгоритм з зворотнім зв‘язком. Якщо такий варіант не підійде, будуть перевірятися інші алгоритми (наприклад, додавання 1-го та 3-го біту для отримання 5-го, або додавання 1–го, 2-го, 3-го для отримання 6-го, …)

(3) При умові випадковості, підраховуємо теоретичне середнє μ:

$$\mu = \frac{M}{2} + \frac{\left(9 + (-1)^{M+1}\right)}{36} - \frac{\left(M/3 - 2/9\right)}{2^M}.$$

Для прикладу з цього пункту:

$$\mu = \frac{13}{2} + \frac{\left(9 + (-1)^{13+1}\right)}{36} - \frac{\left(13/3 + 2/9\right)}{2^{13}} = 6.777222$$

(4) Для кожного блоку підраховуємо величину $T_i = (-1)^M \cdot (L_i - \mu) + \frac{2}{9}$.

Для прикладу з цього пункту: $T_i = (-1)^{13} \cdot (4 - 6.777222) + \frac{2}{9} = 2.999444$

(5) Записуємо T_i в масив $\nu_0, \nu_1, ..., \nu_K$ за таким правилом:

Якщо: $T_i \le -2.5$ збільшуємо на один ν_0;

$-2.5 \le T_i \le -1.5$ збільшуємо на один ν_1;

$-1.5 \le T_i \le -0.5$ збільшуємо на один ν_2;

$-0.5 \le T_i \le 0.5$ збільшуємо на один ν_3;

$0.5 \le T_i \le 1.5$ збільшуємо на один ν_4;

$1.5 \le T_i \le 2.5$ збільшуємо на один ν_5;

$T_i > 2.5$ збільшуємо на один ν_6;

(6) Підраховуємо $\chi^2(obs) = \sum_{i=0}^{K} \frac{(\nu_i - N \cdot \pi_i)^2}{N \cdot \pi_i}$, де $\pi_0 = 0.01047$, $\pi_1 = 0.03125$, $\pi_2 = 0.125$, $\pi_3 = 0.5$, $\pi_4 = 0.25$, $\pi_5 = 0.0625$, $\pi_6 = 0.02078$ обраховані згідно 3.11.

(7) Підраховуємо $Pvalue = igamc\left(\frac{K}{2}, \frac{\chi^2(obs)}{2}\right)$.

Для прикладу з цього пункту: $Pvalue = 0.949310$

5. Вирішуюче правило (для рівня значущості 1%).

Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

Так як значення $Pvalue$, що отримане в пункті 11.4.(7) є ≥ 0.01 ($Pvalue = 0.949310$), робимо висновок, що послідовність випадкова.

Зауважимо, що якщо значення $Pvalue$ мале (< 0.01), тоді це означає, що розподіл T_i, що отриманий при аналізі послідовності ε відрізняється від очікуємого (для випадкової послідовності). Очікується, що розподіл T_i має бути пропорційним до ймовірностей π_i, що наведені в 11.4.(6).

7. Рекомендації по вхідним розмірам.

   Обирайте $n \geq 10^6$. M має задовольняти умову $500 \leq M \leq 5000$ і $N \geq 200$ (для того, щоб отримати адекватне значення χ^2).

12. Тест серій.

1. Мета тесту.

   Тест оснований на підрахунку частоти усіх можливих m-бітних шаблонів (з перекриттям) в усій послідовності. Мета тесту – визначити, чи буде кількість, яку зустрічається в послідовності кожний з можливих m-бітних шаблонів (їх буде 2^m) приблизно дорівнювати такій кількості у випадковій послідовності. Випадкова послідовність характеризується рівномірністю. Тому кожний з m-бітових шаблонів має таку саму ймовірність появи як і будь-який інший m-бітовий шаблон. Зауважимо, що при $m = 1$ цей тест перетворюється в частотний тест (параграф 1).

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

   m - довжина блоків, на які розбивається вхідна послідовність (в бітах).

3. Статистика тесту та граничний розподіл.

   $\nabla\psi_m^2(obs), \nabla^2\psi_m^2(obs)$ - міри того, як добре частоти m-бітних шаблонів, що спострігаються, узгоджуються з частотами для випадкової послідовності.

   Граничний розподіл статистик тесту χ^2-розподіл.

4. Опис тесту.

   (1) Формуємо нову послідовність ε': додаємо перші $m-1$ біт послідовності ε в її кінець і отримуємо ε'.

   Наприклад, дано $n = 10$, $\varepsilon = 0011011101$. Якщо $m = 3$, тоді $\varepsilon' = 001101110100$, Якщо $m = 2$, тоді $\varepsilon' = 00110111010$. Якщо $m = 1$, тоді $\varepsilon' = 0011011101$.

   (2) Визначаємо частоти всіх можливих m-бітних шаблонів, $(m-1)$-бітних шаблонів, що перекриваються, $(m-2)$-бітних шаблонів. Позначимо через $\nu_{i_1, i_2, ..., i_m}$ частоту шаблону $i_1, i_2, ..., i_m$, через $\nu_{i_1, i_2, ..., i_{m-1}}$ частоту шаблону $i_1, i_2, ..., i_{m-1}$, через $\nu_{i_1, i_2, ..., i_{m-2}}$ частоту шаблону $i_1, i_2, ..., i_{m-2}$.

Для прикладу з цього пункту, якщо $m=3$, тоді $(m-1)=2$, $(m-2)=1$. І отримаємо такі частоти: для 3-бітних шаблонів - $\nu_{000}=0,\ \nu_{001}=1,\ \nu_{010}=1,\ \nu_{011}=2,\ \nu_{100}=1,\ \nu_{101}=2,\ \nu_{110}=2,\ \nu_{111}=0$, для 2-бітних - $\nu_{00}=1,\ \nu_{01}=3,\ \nu_{10}=3,\ \nu_{1}=3$, для 1-бітних - $\nu_{0}=4,\ \nu_{1}=6$.

(3) Підраховуємо:

$$\psi_m^2=\frac{2^m}{n}\cdot\sum_{i_1,\ldots,i_m}\left(\nu_{i_1,\ldots,i_m}-\frac{n}{2^m}\right)^2=\frac{2^m}{n}\sum_{i_1,\ldots,i_m}\left(\nu_{i_1,\ldots,i_m}\right)^2-n,$$

$$\psi_{m-1}^2=\frac{2^{m-1}}{n}\cdot\sum_{i_1,\ldots,i_{m-1}}\left(\nu_{i_1,\ldots,i_{m-1}}-\frac{n}{2^{m-1}}\right)^2=\frac{2^{m-1}}{n}\sum_{i_1,\ldots,i_{m-1}}\left(\nu_{i_1,\ldots,i_{m-1}}\right)^2-n,$$

$$\psi_{m-2}^2=\frac{2^{m-2}}{n}\cdot\sum_{i_1,\ldots,i_{m-2}}\left(\nu_{i_1,\ldots,i_{m-2}}-\frac{n}{2^{m-2}}\right)^2=\frac{2^{m-2}}{n}\sum_{i_1,\ldots,i_{m-2}}\left(\nu_{i_1,\ldots,i_{m-2}}\right)^2-n.$$

Для прикладу з цього пункту:

$$\psi_3^2=\frac{2^3}{10}\cdot(0+1+1+4+1+4+4+1)-10=2.8,$$

$$\psi_2^2=\frac{2^2}{10}\cdot(1+9+9+9)-10=1.2,$$

$$\psi_1^2=\frac{2}{10}\cdot(16+36)-10=0.4$$

(4) Підраховуємо:

$$\nabla\psi_m^2=\psi_m^2-\psi_{m-1}^2,$$

$$\nabla^2\psi_m^2=\psi_m^2-2\cdot\psi_{m-1}^2+\psi_{m-2}^2$$

Для прикладу з цього пункту:

$$\nabla\psi_m^2=\psi_m^2-\psi_{m-1}^2=\psi_3^2-\psi_2^2=2.8-1.2=1.6,$$

$$\nabla^2\psi_m^2=\psi_m^2-2\cdot\psi_{m-1}^2+\psi_{m-2}^2=\psi_3^2-2\cdot\psi_2^2+\psi_1^2=2.8-2\cdot1.2+0.4=0.8$$

(5) Підраховуємо: $Pvalue1=igamc\left(2^{m-2},\nabla\psi_m^2\right)$, $Pvalue2=igamc\left(2^{m-3},\nabla^2\psi_m^2\right)$.

Для прикладу з цього пункту: $Pvalue1=igamc(2,1.6)=0.808792$,

$Pvalue2=igamc(1,0.8)=0.670320$

5. Вирішуюче правило (для рівня значущості 1%).

Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

Так як значення $Pvalue$, що отримане в пункті 12.4.(5) є ≥ 0.01

($Pvalue1 = 0.808792$, $Pvalue2 = 0.670320$), робимо висновок, що послідовність випадкова.

Зауважимо, що якщо значення $\nabla^2\psi_m^2$ або $\nabla\psi_m^2$ велике, то це свідчить про нерівномірність розподілу m -бітних шаблонів.

7. Рекомендації по вхідним розмірам.

   Обирайте m та n так, щоб $m < \lfloor \log_2 n \rfloor - 2$

13. Тест на основі апроксимації ентропії.

1. Мета тесту.

   Як і попередній тест, цей оснований на підрахунку частот усіх можливих шаблонів довжиною m біт по всій вхідній послідовності. Мета тесту – порівняти частоти всіх можливих шаблонів довжини m та $(m+1)$, що перекриваються, з результатами, щодо випадкової послідовності.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

   m - перша довжина шаблонів.

   $m+1$ - друга довжина шаблонів. (Тест проводиться для шаблонів двох довжин)

3. Статистика тесту та граничний розподіл.

   $\chi^2(obs)$ - міри того, як добре спостерігаєме значення $ApEn(m)$ (див.13.4.(6)) узгоджуються з очікуємим для випадкової послідовності.

   Граничний розподіл статистики тесту χ^2 -розподіл.

4. Опис тесту.

   (1) Додаємо перші $m-1$ біт послідовності ε (для того, щоб створити n m -бітних підрядків, що перекриваються) в її кінець і отримуємо нову послідовність.

   Наприклад, якщо $n = 10$, $\varepsilon = 0100110101$, $m = 3$, тоді додамо перші два біти 01 в кінець послідовності і отримаємо послідовність 010011010101 (Зауважимо, що таку операцію додавання в кінець робимо окремо для кожного значення m).

   (2) Підраховуємо кількості всіх можливих шаблонів. Позначимо через C_i^m - частоту m -бітового шаблону, що має десяткове значення i .

Для прикладу з цього пункту можливі 8 шаблонів для $m = 3$. Позначимо через #B – кількість шаблону B. Тоді маємо:

#000 = 0, #001 = 1, #010 = 3, #011 = 1,

#100 = 1, #101 = 3, #110 = 1, #111 = 0.

(3) Підраховуємо $C_i^m = \frac{\#i}{n}$ для кожного i.

Для прикладу з цього пункту:

$C_{000}^3 = 0,\ C_{001}^3 = 1,\ C_{010}^3 = 0.3,\ C_{011}^3 = 0.1,$

$C_{100}^3 = 0.1,\ C_{101}^3 = 0.3,\ C_{110}^3 = 0.1,\ C_{111}^3 = 0$

(4) Підраховуємо $\varphi^{(m)} = \sum_{i=0}^{2^m-1} \pi_i \cdot \log \pi_i$, де $\pi_i = C_j^i,\ \ j = \log_2 i$.

Для прикладу з цього пункту:

$$\varphi^{(3)} = 0 \cdot \log 0 + 0.1 \cdot \log 0.1 + 0.3 \cdot \log 0.3 + 0.1 \cdot \log 0.1 + \\ + 0.1 \cdot \log 0.1 + 0.3 \cdot \log 0.3 + 0.1 \cdot \log 0.1 + 0 \cdot \log 0 = -1.64341772$$

(5) Повторюємо кроки (1)-(4) замінюючи m на $(m+1)$

Крок 1: Для прикладу з цього пункту, тепер $m = 4$ і послідовність, що тестується 0100110101010.

Крок 2: Шаблони, що зустрічаються в послідовності - 0100, 1001, 0011, 0110, 1101, 1010, 0101, 1010, 0101, 1010. Тоді #0011 = 1, #0100 = 1, #0101 = 2, #0110 = 1, #1001 = 1, #1010 = 3, #1101 = 1, і для всіх інших можливих шаблонів кількість дорівнює нулю.

Крок 3: $C_{0011}^4 = C_{0100}^4 = C_{0110}^4 = C_{1001}^4 = C_{1101}^4 = 0.1$, $C_{0101}^4 = 0.2$, $C_{1010}^4 = 0.3$, і всі інші дорівнюють нулю.

Крок 4:

$$\varphi^{(4)} = 0 + 0 + 0 + 0.1 \cdot \log 0.1 + 0.1 \cdot \log 0.1 + 0.2 \cdot \log 0.2 + 0.1 \cdot \log 0.1 + 0 + \\ + 0 + 0.1 \cdot \log 0.1 + 0.3 \cdot \log 0.3 + 0 + 0 + 0.1 \cdot \log 0.1 + 0 + 0 = -1.83437197$$

(6) Підраховуємо статистику $\chi^2(obs) = 2 \cdot n \cdot [\log 2 - ApEn(m)]$, де $ApEn(m) = \varphi^{(m)} - \varphi^{(m+1)}$.

Для прикладу з цього пункту:

$ApEn(3) = -1.643618 - (-1.834372) = 0.190954$,

$\chi^2(obs) = 2 \cdot 10 \cdot (0.693147 - 0.190954) = 0.502193$.

(7) Підраховуємо $Pvalue = igamc\left(2^{m-1}, \frac{\chi^2(obs)}{2}\right)$.

Для прикладу з цього пункту: $Pvalue = igamc\left(2^2, \frac{0.502193}{2}\right) = 0.261961$

5. Вирішуюче правило (для рівня значущості 1%).

   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 13.4.(7) є ≥ 0.01 ($Pvalue = 0.261961$), робимо висновок, що послідовність випадкова.

   Зауважимо, що якщо

7. Рекомендації по вхідним розмірам.

   Обирайте m та n так, щоб $m < \lfloor \log_2 n \rfloor - 2$

14. Тест накопичених сум.

1. Мета тесту.

   Тест оснований на визначені максимального відхилення (від нуля) випадкового блукання, що визначене як інтегральна сума конвертованої (до ± 1) послідовності. Мета тесту – визначити, чи будуть інтегральні суми частинних підпослідовностей вхідної послідовності занадто великими чи малими відносно очікуємої поведінки інтегральних сум для випадкової послідовності. Інтегральна сума розглядається як випадкове блукання. Для випадкової послідовності відхилення інтегральних сум має бути невеликим. Для невипадковостей певного типу відхилення інтегральних сум від нуля може бути досить значним.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

   mode - якщо дорівнює 0, тоді тест проводиться від початку до кінця послідовності, якщо дорівнює 1, тоді тест проводиться від кінця до початку послідовності .

3. Статистика тесту та граничний розподіл.

   z - максимальне відхилення від нуля часткової суми конвертованої послідовності.

   Граничний розподіл статистики тесту нормальний.

4. Опис тесту.

(1) Формуємо нову послідовність: нулі послідовності ε замінюємо на (-1), $X_i = 2 \cdot \varepsilon_i - 1$.

Наприклад, якщо $\varepsilon = 1011010111$, тоді $X = 1,-1,1,1,-1,1,-1,1,1,1$.

(2) Підраховуємо частинні суми S_i починаючи з X_1, якщо mode $= 0$ і з X_n, якщо mode $= 1$.

mode $= 0$ (вперед)	mode $= 1$ (назад)
$S_1 = X_1$	$S_1 = X_n$
$S_2 = X_1 + X_2$	$S_2 = X_n + X_{n-1}$
$S_3 = X_1 + X_2 + X_3$	$S_3 = X_n + X_{n-1} + X_{n-2}$
…	…
$S_n = X_1 + X_2 + ... + X_n$	$S_n = X_n + X_{n-1} + ... + X_1$

Для прикладу з цього пункту, при mode $= 0$ маємо:

$S_1 = 1$

$S_2 = 1 + (-1) = 0$

$S_3 = 1 + (-1) + 1 = 1$

$S_4 = 1 + (-1) + 1 + 1 = 2$

$S_5 = 1 + (-1) + 1 + 1 + (-1) = 1$

$S_6 = 1 + (-1) + 1 + 1 + (-1) + 1 = 2$

$S_7 = 1 + (-1) + 1 + 1 + (-1) + 1 + (-1) = 1$

$S_8 = 1 + (-1) + 1 + 1 + (-1) + 1 + (-1) + 1 = 2$

$S_9 = 1 + (-1) + 1 + 1 + (-1) + 1 + (-1) + 1 + 1 = 3$

$S_{10} = 1 + (-1) + 1 + 1 + (-1) + 1 + (-1) + 1 + 1 + 1 = 4$

(3) Підраховуємо статистику $z = \max\limits_{1 \le k \le n} |S_k|$ - максимальне відхилення часткових сум.

Для прикладу з цього пункту $z = 4$.

(4) Підраховуємо

$$Pvalue = 1 - \sum_{k=\left(\frac{-n}{z}+1\right)/4}^{\left(\frac{n}{z}-1\right)/4} \left(\Phi\left(\frac{(4\cdot k+1)\cdot z}{\sqrt{n}} \right) - \Phi\left(\frac{(4\cdot k-1)\cdot z}{\sqrt{n}} \right) \right) + \sum_{k=\left(\frac{-n}{z}-3\right)/4}^{\left(\frac{n}{z}-1\right)/4} \left(\Phi\left(\frac{(4\cdot k+3)\cdot z}{\sqrt{n}} \right) - \Phi\left(\frac{(4\cdot k+1)\cdot z}{\sqrt{n}} \right) \right),$$

де $\Phi(x) = \frac{1}{\sqrt{2\cdot\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$

Для прикладу з цього пункту $Pvalue = 0.4116588$

5. Вирішуюче правило (для рівня значущості 1%).

   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 13.4.(7) є ≥ 0.01 ($Pvalue = 0.4116588$), робимо висновок, що послідовність випадкова.

   Зауважимо, що при $\text{mode} = 0$, великі значення статистики говорять про те, що дуже багато нулів або одиниць на початку послідовності; при $\text{mode} = 1$, великі значення статистики говорять про те, що дуже багато нулів або одиниць в кінці послідовності. Малі значення статистики свідчать про рівномірне перемішування нулів та одиниць.

7. Рекомендації по вхідним розмірам.

   Рекомендується, щоб кожна послідовність, що тестується складалася як мінімум зі 100 бітів ($n \geq 100$).

15. Тест випадкових блукань.

1. Мета тесту.

   Тест оснований на підрахунку кількості циклів, що точно k разів переходять в певний стан інтегральної суми, яка розглядається як випадкове блукання. Інтегральні суми рахуються для послідовності, що отримана з вхідної шляхом заміни всіх нулів на мінус одиниці. Цикл складається з послідовності кроків випадково блукання, що починається та закінчується нульовим станом. Мета тесту – визначити, наскільки кількість перебувань в конкретно заданому стані в межах

одного циклу відхиляється від очікуємої від випадкової послідовності. Насправді цей тест представляє собою 8 тестів для кожного зі станів: -4, -3, -2, -1, 1, 2, 3, 4.

2. Позначення.

   n – довжина вхідної послідовності в бітах.

   ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

3. Статистика тесту та граничний розподіл.

   $\chi^2(obs)$ - для заданого стану x, міра того, наскільки кількість потраплянь в стан x в межах циклу відповідає очікуємій від випадкової послідовності.

   Граничний розподіл статистики тесту χ^2-розподіл.

4. Опис тесту.

   (5) Формуємо нову послідовність: нулі послідовності ε замінюємо на (-1), $X_i = 2 \cdot \varepsilon_i - 1$.

   Наприклад, якщо $\varepsilon = 0110110101$, тоді $X = -1,1,1,-1,1,1,-1,1,-1,1$.

   (6) Підраховуємо частинні суми S_i починаючи з X_1, формуємо множину $S = \{S_i\}$:

   $$S_1 = X_1$$
   $$S_2 = X_1 + X_2$$
   $$S_3 = X_1 + X_2 + X_3$$
   $$\dots$$
   $$S_n = X_1 + X_2 + ... + X_n$$

   Для прикладу з цього пункту маємо:

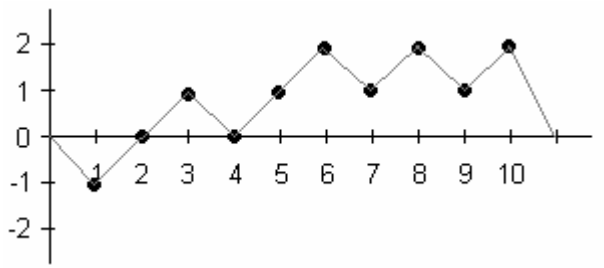
   $S_1 = -1$, $S_2 = 0$, $S_3 = 1$, $S_4 = 0$, $S_5 = 1$,

   $S_6 = 2$, $S_7 = 1$, $S_8 = 2$, $S_9 = 1$, $S_{10} = 2$.

   Отримали множину $S = \{-1,0,1,0,1,2,1,2,1,2\}$

   (7) Формуємо нову послідовність S' приєднуючи нулі перед і після послідовності S. $S' = 0, s_1, s_2, ..., s_n, 0$.

   Для прикладу з цього пункту $S' = 0,-1,0,1,0,1,2,1,2,1,2,0$. Результуюче випадкове блукання зображено нижче.



(8) Нехай J - кількість нулів в послідовності S' не враховуючи першого. Також J буде і кількістю циклів в S'. Під циклом будемо розуміти підпослідовність S', що починається і закінчується нулями та не містить нулів в середині. Якщо $J < 500$, зупиняємо тест.

Для прикладу з цього пункту: $J = 3$ (нулю дорівнює 3-й, 5-й, 12-й елемент послідовності S') і маємо 3 цикли $\{0,-1,0\}$, $\{0,1,0\}$, $\{0,1,2,1,2,1,2,0\}$.

(9) Для кожного циклу, для кожного стану x, що має значення $-4 \le x \le -1$, і $1 \le x \le 4$, підраховуємо кількість, яку він досягається в циклі.

Для прикладу з цього пункту: перший цикл має одну (-1), другий цикл має одну 1, третій цикл має три 1 і три 2. Зобразимо це у вигляді таблиці:

Стан x	Цикли		
	Цикл 1 $\{0,-1,0\}$	Цикл 2 $\{0,1,0\}$	Цикл 3 $\{0,1,2,1,2,1,2,0\}$
-4	0	0	0
-3	0	0	0
-2	0	0	0
-1	1	0	0
1	0	1	3
2	0	0	3
3	0	0	0
4	0	0	0

(10) Для кожного з 8 станів x, підраховуємо величину $\nu_k(x)$ - загальна кількість циклів, в яких стан x досягається рівно k разів, для кожного $k = 0,1,...,5$ (для $k = 5$ величина $\nu_5(x)$ - загальна кількість циклів, в яких стан x досягається ≥ 5 разів). Зауважимо, що $\sum_{k=0}^{5} \nu_k(x) = J$.

Для прикладу з цього пункту:

a. $\nu_0(-1)=2$ (стан (-1) досягається рівно 0 разів в двох циклах)

$\nu_1(-1)=1$ (стан (-1) досягається рівно 1 раз в одному циклі)

$\nu_2(-1)=\nu_3(-1)=\nu_4(-1)=\nu_5(-1)=0$ (стан (-1) досягається рівно 2, 3, 4, ≥ 5 разів в жодному з циклів).

b. $\nu_0(1)=1$ (стан 1 досягається рівно 0 разів в одному циклі)

$\nu_1(1)=1$ (стан 1 досягається рівно 1 раз в одному циклі)

$\nu_3(1)=1$ (стан 1 досягається рівно 3 рази в одному циклі)

$\nu_2(1)=\nu_4(1)=\nu_5(1)=0$ (стан 1 досягається рівно 2, 4, ≥ 5 разів в жодному з циклів).

c. $\nu_0(2)=2$ (стан 2 досягається рівно 0 разів в двох циклах)

$\nu_3(2)=1$ (стан 2 досягається рівно 3 рази в одному циклі)

$\nu_1(2)=\nu_2(2)=\nu_4(2)=\nu_5(2)=0$ (стан 2 досягається рівно 1, 2, 4, ≥ 5 разів в жодному з циклів).

d. $\nu_0(-4)=3$ (стан (-4) досягається рівно 0 разів в трьох циклах)

$\nu_1(-4)=\nu_2(-4)=\nu_3(-4)=\nu_4(-4)=\nu_5(-4)=0$ (стан (-4) досягається рівно 1, 2, 3, 4, ≥ 5 разів в жодному з циклів).

e. і так далі.

Це можна зобразити у таблиці:

Стан	**Кількість циклів**					
x	**0**	**1**	**2**	**3**	**4**	**5**
-4	3	0	0	0	0	0
-3	3	0	0	0	0	0
-2	3	0	0	0	0	0
-1	2	1	0	0	0	0
1	1	1	0	1	0	0
2	2	1	0	0	0	0
3	3	0	0	0	0	0
4	3	0	0	0	0	0

(11) Для кожного з 8 станів x підраховуємо статистику:

$$\chi^2(obs)=\sum_{k=0}^{5}\frac{(\nu_k(x)-J\cdot\pi_k(x))^2}{J\cdot\pi_k(x)},$$

Де $\pi_k(x)$ - ймовірність того, що стан x зустрінеться k разів у випадковій послідовності (див. 3.15).

Для прикладу з цього пункту, для стану $x = 1$:

$$\chi^2(obs) = \frac{(1 - 3 \cdot 0.5)^2}{3 \cdot 0.5} + \frac{(1 - 3 \cdot 0.25)^2}{3 \cdot 0.25} + \frac{(0 - 3 \cdot 0.125)^2}{3 \cdot 0.125} +$$
$$+ \frac{(1 - 3 \cdot 0.0625)^2}{3 \cdot 0.0625} + \frac{(0 - 3 \cdot 0.0312)^2}{3 \cdot 0.0312} + \frac{(0 - 3 \cdot 0.0312)^2}{3 \cdot 0.0312} = 4.333.033$$

(12) Для кожного стану x, підраховуємо $Pvalue = igamc\left(\frac{5}{2}, \frac{\chi^2(obs)}{2}\right)$.

Для прикладу з цього пункту, для стану $x = 1$:

$$Pvalue = igamc\left(\frac{5}{2}, \frac{4.333033}{2}\right) = 0.502529$$

5. Вирішуюче правило (для рівня значущості 1%).

   Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

6. Висновки та інтерпретація результатів тесту.

   Так як значення $Pvalue$, що отримане в пункті 15.4.(8) є ≥ 0.01 ($Pvalue = 0.502529$), робимо висновок, що послідовність випадкова.

   Зауважимо, що великі значення $\chi^2(obs)$ свідчать про відхилення від теоретичного розподілу для заданого стану.

7. Рекомендації по вхідним розмірам.

   Рекомендується, щоб кожна послідовність, що тестується складалася як мінімум з 10^6 бітів ($n \geq 10^6$).

16. Тест випадкових блукань.

8. Мета тесту.

   Тест оснований на підрахунку загальної кількості, яку досягається певний стан інтегральної суми, яка розглядається як випадкове блукання. Інтегральні суми рахуються для послідовності, що отримана з вхідної шляхом заміни всіх нулів на мінус одиниці. Мета тесту – визначити, наскільки кількість досягнень конкретно заданого стану відхиляється від очікуємої кількості у випадковій послідовності. Насправді цей тест представляє собою 18 тестів для кожного зі станів: -9, -8, …, -1, 1, 2, …, 9.

9. Позначення.

n – довжина вхідної послідовності в бітах.

ε - послідовність бітів, яку необхідно протестувати ($\varepsilon = \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$).

10. Статистика тесту та граничний розподіл.

ζ - для заданого стану x, загальна кількість досягнень випадковим блуканням стану x.

Граничний розподіл статистики тесту напівнормальний (для великих n). (Якщо випадкова величина ζ має нормальний розподіл, тоді випадкова величина $|\zeta|$ має напівнормальний розподіл). Якщо послідовність випадкова, тоді статистика буде мати значення біля нуля. Якщо в послідовності дуже багато 1 або 0, тоді статистика буде приймати великі значення.

11. Опис тесту.

(13) Формуємо нову послідовність: нулі послідовності ε замінюємо на (-1), $X_i = 2 \cdot \varepsilon_i - 1$.

Наприклад, якщо $\varepsilon = 0110110101$, тоді $X = -1,1,1,-1,1,1,-1,1,-1,1$.

(14) Підраховуємо частинні суми S_i починаючи з X_1, формуємо множину $S = \{S_i\}$:

$$S_1 = X_1$$
$$S_2 = X_1 + X_2$$
$$S_3 = X_1 + X_2 + X_3$$
$$\dots$$
$$S_n = X_1 + X_2 + ... + X_n$$

Для прикладу з цього пункту маємо:

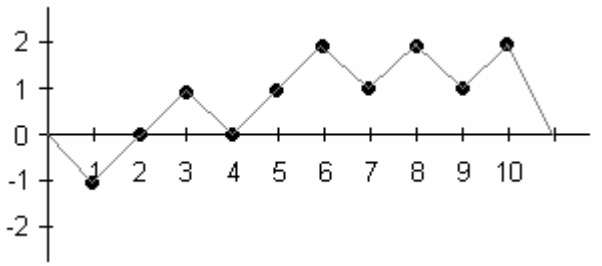
$S_1 = -1$, $S_2 = 0$, $S_3 = 1$, $S_4 = 0$, $S_5 = 1$,

$S_6 = 2$, $S_7 = 1$, $S_8 = 2$, $S_9 = 1$, $S_{10} = 2$.

Отримали множину $S = \{-1,0,1,0,1,2,1,2,1,2\}$

(15) Формуємо нову послідовність S' приєднуючи нулі перед і після послідовності S. $S' = 0, s_1, s_2, ..., s_n, 0$.

Для прикладу з цього пункту $S' = 0,-1,0,1,0,1,2,1,2,1,2,0$. Результуюче випадкове блукання зображено нижче.



(16) Для кожного з 18 станів x підраховуємо $\zeta(x)$ - загальна кількість, яку досягається стан x по всім J циклам (див. 2.15).

Для прикладу з цього пункту: $\xi(-1)=1$, $\zeta(1)=4$, $\xi(2)=3$, і для інших станів $\zeta(x)=0$

(17) Для кожного $\xi(x)$ підраховуємо $Pvalue = erfc\left(\frac{|\xi(x)-J|}{\sqrt{2\cdot J\cdot(4\cdot|x|-2)}}\right)$, де

$$erfc(z)=\frac{2}{\sqrt{\pi}}\cdot\int_{z}^{+\infty}e^{-u^2}\,du\,.$$

Для прикладу з цього пункту, для стану $x=1$:

$$Pvalue = erfc\left(\frac{|4-3|}{\sqrt{2\cdot 3\cdot(4\cdot|4|-2)}}\right)=0.877371$$

12. Вирішуюче правило (для рівня значущості 1%).

Якщо підраховане значення $Pvalue$ менше за 0.01, тоді робимо висновок, що послідовність ε не випадкова. Інакше робимо висновок, що послідовність ε випадкова.

13. Висновки та інтерпретація результатів тесту.

Так як значення $Pvalue$ для стану $x=1$, що отримане в пункті 16.4.(5) є ≥ 0.01 ($Pvalue = 0.877371$), робимо висновок, що послідовність випадкова.

Зауважимо, що великі значення $\chi^2(obs)$ свідчать про відхилення від теоретичного розподілу для заданого стану.

14. Рекомендації по вхідним розмірам.

Рекомендується, щоб кожна послідовність, що тестується, складалася як мінімум з 10^6 бітів ($n \geq 10^6$).

Григорий А. Кравцов

Digitally signed by Григорий А. Кравцов
DN: cn=Григорий А. Кравцов, ou=SMTP, dc=org
Reason: I am the author of this document
Date: 2006.01.19 06:19:32 +02'00'