УДК 501

ВИКОРИСТАННЯ НЕРІВНОСТІ ЧЕБИШЕВА ДЛЯ ПЕРЕВІРКИ НЕЗАЛЕЖНОСТІ СТАТИСТИЧНИХ ТЕСТІВ

А. О. Литвин[1], Л. В. Ковальчук[1]

[1] *Національний технічний університет України «Київський політехнічний інститут»*

Анотація

В роботі представлені результати дослідження методики перевірки незалежності статистичних тестів, що застосовуються для оцінки якостей генератора випадкових послідовностей. Під час дослідження розглядався відомий в сучасній криптографії метод, базований на стандартний граничних розподілах. Були представлені шляхи покращення відомого способу з метою максимально зменшити радіус послідовностей, оскільки в базовій методиці важко оцінити збіжність і треба багато послідовностей, близько одного мільйона.

Ключові слова: генератор випадкових (псевдовипадкових) послідовностей, незалежність тестів, нерівність Чебишева

Вступ

В криптографії важливою є сукупність статистичних критеріїв (тестів), призначена для перевірки аналізованої послідовності гіпотез про незалежність і рівномірність її елементів. Кожний тест складається із обчислення по аналізованій послідовності деякої статистики, що має відомий розподіл для випадкової послідовності, і використання критеріїв згоди. Генератори псевдовипадкових чисел (ГПЧ) і псевдовипадкових послідовностей (ГПВП) широко використовується в багатьох сферах вимірювальної техніки. При цьому вимоги до їх технічних характеристик відрізняються залежно від мети їхнього застосування. Ми розглядаємо застосування, що відносяться до сфери захисту інформації. Необхідною умовою стійкості криптосистеми є наявність певних криптографічних властивостей у генератора, що в ній використовується. До цих властивостей відносяться: рівноймовірність та незалежність символів вихідної послідовності; непередбачуваність; великий період (більший за 280 біт) та деякі інші. У зв'язку з цим перед розробником або користувачем будь-якої криптосистеми постають питання як про оцінку якості генератора, так і про оцінку якості його окремих послідовностей. Мета роботи полягає у детальному вивченні методики перевірки незалежності статистичних тестів саме за допомогою використання нерівності Чебишева, визначення якою має бути послідовність, щоб ми могли користуватись даною методикою і досягали практичного результату. Першочергово досліджувана методика базувалась на теоремі, основою якої була центральна гранична теорема, але за тією методикою потрібна була велика кількість послідовностей.

1. Методика формування набору тестів для перевірки якості ГВП (ГПВП)

На практиці часто виникають два тісно пов'язаних між собою питання. Перше: як за поліноміальний час отримати послідовність чисел, що має певний набір властивостей, достатній для того, щоб її можна було вважати випадковою, тобто отриманою як реалізацію послідовності незалежних випадкових величин з рівномірним розподілом, що приймають значення у певному алфавіті? І друге питання: як визначити, чи можна деяку фіксовану послідовність вважати випадковою? Надалі ми розглядатимемо лише випадкові величини з дискретним розподілом, тобто будемо вважати, для всіх t випадкові величини x_t приймають значення на множині $D = 0, 1, ...N - 1$, що називається алфавітом послідовності. За найбільш поширеним у криптографії визначенням, випадкова послідовність (ВП) – це послідовність незалежних (у сукупності) випадкових величин з рівномірним розподілом, тобто: $P(x_t = i) = N^{-1}$, де N – об'єм алфавіту D. Пристрій, що реалізує ВП, називається генератором ВП. Саме реалізації ВП є тими об'єктами, що цікавлять нас. Одна з задач криптографії, що часто виникає на практиці, полягає у тому, щоб за конкретною реалізацією визначити, чи буде вона реалізацією ВП (випадкова послідовність). Тести A та B назвемо незалежними, якщо індикатори ξ_A, ξ_B статистично незалежні. Іншими словами, незалежність тестів означає, що результат застосування тесту A для послідовності не залежить від результату застосування тесту B. Аналогічно можна означити набір незалежних тестів. Тести з набору $\{A_i\}$ назвемо незалежними, якщо незалежні у сукупності відповідні індикатори[1].

Позначимо через ζ_A, ζ_B статистики, які обчислюються в тестах A і B, K_A, K_B – критичні області тестів, відповідно. Тоді розподіл величин ξ_A, ξ_B наступний:

$$P((\xi)_A = 1) = P((\zeta)_A \notin K_A),$$
$$P((\xi)_A = 0) = P((\zeta)_A \in K_A),$$
$$P((\xi)_B = 1) = P((\zeta)_B \notin K_B),$$
$$P((\xi)_B = 0) = P((\zeta)_B \in K_B).$$

Тому для незалежності ξ_A, ξ_B необхідно і достатньо, щоб були незалежними події $\{\xi_A \in K_A\}$ та $\{\xi_B \in K_B\}$. Тобто, незалежність тестів еквівалентна незалежності подій, що полягають у попаданні статистик у критичні області. З кожним тестом пов'язані два параметри: α – ймовірність помилки 1-го роду – ймовірність відхилити випадкову послідовність, та β – ймовірність помилки 2-го роду – ймовірність прийняти послідовність, що не є випадковою.

2. Використання нерівності Чебишева для для перевірки незалежності статистичних тестів

Нехай потрібно перевірити незалежність набору з N тестів, де і-й тест має рівень значимості α_i, $i = 1...N$. Для перевірки незалежності набору використовується ГВЧ, який має необхідні статистичні якості. Це означає, що даний генератор повинен пройти тестування набором тестів, для якого перевіряється незалежність, та результати тестування повинні бути оцінені за методикою, вказаною в. Залежність чи незалежність тестів не впливають на якість оцінки ГВЧ, вони впливають лише на час проведення тестування. Нехай отримано n послідовностей з ГВЧ (довжини послідовностей повинні бути придатними для проведення тестування).

Теорема: Покладемо $\eta_j = 1$, якщо j-та послідовність пройшла всі тести, та $\eta_j = 0$ у протилежному випадку, тобто якщо j-тою послідовністю хоча б один тест не пройдено, $j = 1, ..., n$. Позначимо через η частку послідовностей, які пройшли всі тести: $\eta = \frac{1}{n}\sum_{j=1}^{n}\eta_j$.Покладемо

$$a = \Pi_{i=1}^{N}(1-\alpha_i), \sigma^2 = \frac{1}{n}(a-a^2).$$

Тоді, якщо тести незалежні, то $P(|\eta - a| > \varepsilon) \le \frac{\sigma^2}{\varepsilon^2}$. Дана теорема дає можливість побудувати ефективну та математично обґрунтовану методику перевірки незалежності тестів, що застосовуються для оцінки якостей ГВП.

3. Методика перевірки незалежності тестів

Нехай для оцінки якостей ГВП застосовується набір з N тестів, які мають один і той же рівень значимості α. Використовуючи результати попереднього пункту, можна запропонувати наступну методику для перевірки незалежності тестів цього набору.

1) Задати β – рівень значимості для статистичної перевірки незалежності тестів цього набору (ймовірність помилки першого роду).
2) Обчислити наступні величини: $a = p = (1-\alpha)^N$, $q = 1-(1-\alpha)^N$, $\sigma^2 = \frac{pq}{n}$
3) Використовуючи послідовностей, отриманих з ГВП, обчислити величини η_j, $j = 1, ..., n$.
4) Обчислити $\eta = \frac{1}{n}\sum_{j=1}^{n}\eta_j$.
5) Обчислити $|\eta - a|$.
6) Обчислити $\varepsilon = \frac{\sigma^2}{\beta}$.

n	N	α	a	σ^2	σ	ε
100	6	0,05	0,735092	0,001947	0,044128	0,197348
200	6	0,05	0,735092	0,000974	0,031204	0,139546
300	6	0,05	0,735092	0,000649	0,025478	0,113939
400	6	0,05	0,735092	0,000487	0,022064	0,098674
500	6	0,05	0,735092	0,000389	0,019735	0,088257
600	6	0,05	0,735092	0,000325	0,018015	0,080567
700	6	0,05	0,735092	0,000278	0,016679	0,074591
800	6	0,05	0,735092	0,000243	0,015602	0,069773
900	6	0,05	0,735092	0,000216	0,014709	0,065783
1000	6	0,05	0,735092	0,000195	0,013955	0,062407

Рис. 1. Розрахунки при $N = 6$, $\alpha = 0,05$

7) Якщо $|\eta - a| < \sqrt{\frac{\sigma^2}{\alpha}}$ – гіпотеза про незалежність тестів у наборі приймається . Інакше відхиляється.

4. Побудова критеріїв для перевірки незалежності тестів

До критеріїв відносимо a-середнє, σ^2 – дисперсія. α – помилку 1-го роду (рівень довіри) задаємо $\alpha = 0,05$. N – кількість зроблених тестів. Візьмемо N=6. Отож, в нас є:

$$P(|\eta - a| > \varepsilon) \le \frac{\sigma^2}{\varepsilon^2}, P(|\eta - a| \le \varepsilon) \ge 1 - \frac{\sigma^2}{\varepsilon^2}.$$

Тоді обчислюємо критерії наступним чином:

$$a = (1-\alpha)^N, \alpha = \frac{\sigma^2}{\varepsilon^2}, \varepsilon^2 = \frac{\sigma^2}{\alpha},$$

$$\varepsilon = \sqrt{\frac{\sigma^2}{\alpha}}, \sigma^2 = \frac{1}{n}a(1-a), \sigma = \sqrt{\frac{a(1-a)}{n}},$$

де $\alpha_1, \alpha_2, \alpha_3, ..., \alpha_i, ...$ однакові і дорівнюють $\alpha = 0,05$, $n = 100, 200, 300, ..., 1000$. Результати обчислень наведені на рис. 1.

Висновки

В ході роботи ми детально зупинились на перевірці незалежності статистичних тестів, та спробували в одній з методик перевірки використати нерівність Чебишева для її реалізації. В даній роботі були спроби дослідити як і коли ми можемо користуватись запропонованим методом, чи буде оцінка менш грубою, та намагались дізнатись скільки треба послідовностей, щоб ця оцінка була практичною. Але звичайно точну кількість послідовностей визначити не вдасться. Отож, за результатами розрахунків бачимо, що методика суттєво покращилась, де n – це величина ε (межі ε) повинна бути суттєво меншою чим його математичне сподівання. Тобто дана методика є більш зручною, в той час коли при використанні ЦГТ ми не знали приблизно коли достатньо послідовностей (близько 1 млн). А в нашому випадку достатньо набагато менше послідовностей.

Перелік використаних джерел

1. Іванов М. А.,Чугунков І. В. Теорія, застосування і оцінка якості генераторів псевдовипадкових послідовностей. — М.: "КУДИЦ-ОБРАЗ", 2003. — 240 с.