Лабораторна робота № 1

Побудова оцінок та довірчих інтервалів

Нехай ω_1 та ω_2 – це незалежні рівномірно розподілені на $[0,1]$ випадкові величини (в.в.). Пара незалежних в.в. (ξ_1,ξ_2), які мають стандартний нормальний розподіл (тобто $N(0,1)$), генерується за допомогою перетворення:

$$\xi_1 = \sqrt{-2\ln\omega_1}\,\sin(2\pi\omega_2), \qquad \xi_2 = \sqrt{-2\ln\omega_1}\,\cos(2\pi\omega_2)$$

(в.в. $N(0,1)$ можна генерувати і за допомогою вбудованого в комп'ютер генератора).

Позначимо $a = \mathbf{M}\xi_i = 0,\ \sigma^2 = \mathbf{D}\xi_i = 1$.

Нехай спостерігається вибірка $\overline{X} = (X_1,\dots,X_n)$, де $X_i \sim N(0,1)$.

Завдання 1. Побудувати довірчий інтервал для:

a) математичного сподівання a у припущенні, що спостерігаються в.в. $\{X_i\}$, які мають нормальний розподіл, але дисперсія σ^2 невідома;

b) математичного сподівання a у припущенні, що спостерігаються в.в. $\{X_i\}$, розподіл яких невідомий.

c) дисперсії σ^2 у припущенні, що спостерігаються в.в. $\{X_i\}$, які мають нормальний розподіл.

Всі довірчі інтервали будуються із достовірністю $1-\gamma = 0.99$ для $n = 100$, $n = 10\,000$ та $n = 1\,000\,000$. В усіх цих випадках дослідити, чи потрапляють математичне сподівання та дисперсія у побудовані довірчі інтервали, а також оцінити, як змінюється довжина довірчого інтервалу при збільшенні n. Інакше кажучи, виводити на друк:

- кількість виконаних реалізацій;
- отриману оцінку;
- побудований довірчий інтервал;
- довжину довірчого інтервалу.

Зауваження. Формули для побудови оцінок та довірчих інтервалів див. лекцію 3. Для випадку b) краще використовувати незміщену оцінку дисперсії.

Завдання 2: обчислення ймовірності трьома способами із дослідженням швидкості збіжності. Потрібно обчислити наступну ймовірність:

$$Q_m = \mathbf{P}\{\eta > \xi_1 + \ldots + \xi_m\},$$

де m – це деяке натуральне число (параметр), а в.в. η та $\{\xi_i\}$ є незалежними та мають наступні розподіли:

$$G(u) = \mathbf{P}\{\eta < u\} = \begin{cases} 1 - \dfrac{1}{1+u}, & u > 0, \\ 0, & u \le 0, \end{cases} \qquad \mathbf{M}\eta = \int_0^\infty [1 - G(u)]\, du = \int_0^\infty \frac{1}{1+u}\, du = \infty,$$

$$F(u) = \mathbf{P}\{\xi_i < u\} = \begin{cases} 1 - e^{-u}, & u > 0, \\ 0, & u \le 0, \end{cases} \qquad \mathbf{M}\xi_i = 1, \qquad \xi_i \sim Exp(1).$$

Позначимо:

$$A_m(u) = \mathbf{P}\{\xi_1 + \ldots + \xi_m < u\} = 1 - \sum_{k=0}^{m-1} \frac{u^k}{k!} e^{-u} = \sum_{k=m}^{\infty} \frac{u^k}{k!} e^{-u}, \quad u \ge 0, \quad m \ge 1.$$

Це розподіл Ерланга з параметрами m та 1. Відповідна щільність розподілу має вигляд:

$$a_m(u) = \frac{u^{m-1}}{(m-1)!} e^{-u}, \quad u \ge 0.$$

Тоді

$$Q_m = \mathbf{P}\{\eta > \xi_1 + \ldots + \xi_m\} = \int_0^\infty \frac{1}{1+u} \cdot \frac{u^{m-1}}{(m-1)!} e^{-u}\, du.$$

Легко бачити, що цей інтеграл в явному вигляді не береться. Будемо обчислювати його методом Монте-Карло.

Зауваження 1. Нехай $\omega_1, \omega_2, \ldots$ – послідовність незалежних рівномірно розподілених на відрізку $[0,1]$ в.в. (послідовність псевдовипадкових чисел, яку отримуємо генератором випадкових чисел). Тоді

$$\eta = G^{-1}(\omega) = \frac{1}{\omega} - 1, \qquad \xi_i = F^{-1}(1-\omega_i) = -\ln \omega_i$$

(в.в. $1-\omega_i$ та ω_i мають однаковий розподіл).

Зауваження 2. Загальна схема обчислення ймовірності Q_m виглядає наступним чином. Нехай $\hat{q}_1, \hat{q}_2, \dots$ – незміщені оцінки ймовірності Q_m, які отримано тим чи іншим методом. Незміщена оцінки ймовірності Q_m та вибіркова дисперсія обчислюються за формулами:

$$\hat{Q}_m^{(n)} = \frac{1}{n}\sum_{i=1}^{n}\hat{q}_i, \qquad \left[\hat{\sigma}_m^{(n)}\right]^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}\hat{q}_i^2 - n\left[\hat{Q}_m^{(n)}\right]^2\right)$$

(n – кількість реалізацій).

Кількість реалізацій n^* алгоритму, які потрібно здійснити для обчислення ймовірності Q_m із заданою *достовірністю* $1-\gamma$ та *відносною похибкою* ε обчислюється за формулою:

$$n^* = \min\left\{ n \geq n_0 : \ n \geq \frac{z_\gamma^2 \left[\hat{\sigma}_m^{(n)}\right]^2}{\varepsilon^2 \left[\hat{Q}_m^{(n)}\right]^2} \right\},$$

де n_0 – початкова кількість реалізацій, яка потрібна для “стабілізації” дисперсії, а z_γ – це коефіцієнт, який знаходиться з рівняння $2\Phi(z) = 1-\gamma$ ($\Phi(z)$ – функція Лапласа). Для кожного з наведених нижче методів обчислення n_0 приймає своє значення.

В усіх наведених вище випадках обчислення вести із достовірністю 0.99 та відносною похибкою 1%, тобто $z_\gamma = 2.575$ і $\varepsilon = 0.01$. Розглядаються три можливі значення параметра m: 1, 10, 100, 1000 та 10000. Потрібно виконати наступні завдання.

A. При кожному $m = 1;\ 10;\ 100;\ 1000;\ 10000$ обчислити точне значення ймовірності Q_m (для перевірки коректності алгоритмів моделювання).

B. Стандартний метод Монте-Карло (*метод 1*):

$Q_m = \mathbf{M} I(\eta > \xi_1 + \ldots + \xi_m)$, тобто $\hat{q}_i = I(\eta^{(i)} > \xi_1^{(i)} + \ldots + \xi_m^{(i)})$, де $I(\cdot)$ – індикаторна функція.

C. *Метод 2:*

$$Q_m = \mathbf{P}\{\eta > \xi_1 + \ldots + \xi_m\} = \int_0^\infty [1 - G(u)]\, a_m(u)\, du = \int_0^\infty [1 - G(u)]\, dA_m(u) =$$

$$= \mathbf{M}[1 - G(\xi_1 + \ldots + \xi_m)] = \mathbf{M}\frac{1}{1 + \xi_1 + \ldots + \xi_m},$$

тобто $\hat{q}_i = \dfrac{1}{1 + \xi_1^{(i)} + \ldots + \xi_m^{(i)}}$.

D. *Метод 4* (випадок, коли $m > 1$):

$$Q_m = \mathbf{P}\{\eta > \xi_1 + \ldots + \xi_m\} = \int_0^\infty \frac{1}{1+u} \cdot \frac{u^{m-1}}{(m-1)!} e^{-u}\, du = \frac{1}{m-1} \int_0^\infty \frac{u}{1+u} \cdot \frac{u^{m-2}}{(m-2)!} e^{-u}\, du =$$

$$= \frac{1}{m-1} \int_0^\infty \frac{u}{1+u}\, dA_{m-1}(u) = \frac{1}{m-1} \mathbf{M} \frac{\xi_1 + \ldots + \xi_{m-1}}{1 + \xi_1 + \ldots + \xi_{m-1}},$$

тобто $\hat{q}_i = \dfrac{1}{m-1} \cdot \dfrac{\xi_1^{(i)} + \ldots + \xi_{m-1}^{(i)}}{1 + \xi_1^{(i)} + \ldots + \xi_{m-1}^{(i)}}$.

Виводити на друк:

- оцінку;
- вибіркову дисперсію;
- довірчий інтервал;
- кількість виконаних реалізацій для побудови оцінки із достовірністю 0.99 та відносною похибкою 1%.