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 $\overline{X} = (X_1, \ldots, X_n)$, 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;
- ширину довірчого інтервалу.

Зауваження. Формули для побудови оцінок та довірчих інтервалів див. лекцію 4. Для випадку b) краще використовувати незміщену оцінку дисперсії.

Завдання 2: обчислення ймовірності чотирьма способами із дослідженням швидкості збіжності, тобто кількості реалізацій, витрачених різними алгоритмами на побудову оцінки із заданими достовірністю та відносною похибкою. Потрібно обчислити наступну ймовірність:

$$Q(\alpha) = \mathbf{P}\{\xi_\alpha < \eta\},$$

де ξ_α та η – незалежні в.в., які мають функції розподілу (ф.р.) $F_\alpha(x)$ та $G(x)$ відповідно. Припустимо, що $\xi_\alpha \ge 0$ для будь-яких значень параметра α та $\eta \ge 0$ з ймовірністю 1 і існують щільності $f_\alpha(u) = F'_\alpha(u),\ u \ge 0,$ та $g(u) = G'(u),\ u \ge 0$. Виберемо наступні ф.р.:

$$F_\alpha(u) = 1 - e^{-(\alpha u)^4}, \quad G(u) = 1 - e^{-u^2}, \quad u \ge 0\,.$$

Тоді ймовірність $Q(\alpha)$ обчислюється за формулою:

$$Q(\alpha) = \mathbf{P}\{\xi_\alpha < \eta\} = \int_0^\infty F_\alpha(u)\,dG(u) = \int_0^\infty \left(1 - e^{-(\alpha u)^4}\right) 2u\,e^{-u^2}\,du \underset{\alpha\to 0}{\sim} 2\alpha^4 \int_0^\infty u^5 e^{-u^2}\,du =$$

$$\left\| v = u^2,\ dv = 2u\,du \right\| = \alpha^4 \int_0^\infty v^2 e^{-v}\,dv = \alpha^4\,\Gamma(3) = 2\alpha^4, \qquad (1)$$

де $\Gamma(\beta) = \int_0^\infty v^{\beta-1} e^{-v}\,dv,\quad \beta > 0$, – гамма-функція.

Зауваження 1. Нехай $\omega, \omega_1, \omega_2, \ldots$ – послідовність незалежних рівномірно розподілених на відрізку $[0,1]$ в.в. (послідовність псевдовипадкових чисел). Тоді моделювання в.в. ξ_α та η відбувається за формулами: $\xi_\alpha = F_\alpha^{-1}(\omega)$ і $\eta = G^{-1}(\omega)$ (за умови, що ці обернені функції існують). Тобто

$$\xi_\alpha^{(i)} = \frac{1}{\alpha}\left(-\ln\omega_i\right)^{\frac{1}{4}}, \quad \eta_i = \left(-\ln\omega_i\right)^{\frac{1}{2}}, \quad i = 1, 2, \ldots$$

(враховано, що в.в. $1-\omega_i$ та ω_i мають один і той же рівномірний на $[0,1]$ розподіл).

Зауваження 2. Загальна схема обчислення ймовірності $Q(\alpha)$ виглядає наступним чином. Нехай $\hat{q}_1, \hat{q}_2, \ldots$ – незміщені оцінки ймовірності $Q(\alpha)$. Позначимо

$$\hat{Q}_n(\alpha) = \frac{1}{n}\sum_{i=1}^n \hat{q}_i, \quad \hat{\sigma}_n^2(\alpha) = \frac{1}{n-1}\left(\sum_{i=1}^n \hat{q}_i^{\,2} - n\,\hat{Q}_n^2(\alpha)\right)$$

– незміщена оцінка дисперсії.

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

$$n^*(\alpha) = \min\left\{ n \ge n_0(\alpha):\ n \ge \frac{z_\gamma^2\ \hat{\sigma}_n^2(\alpha)}{\varepsilon^2\ \hat{Q}_n^2(\alpha)} \right\},$$

де $n_0(\alpha)$ – початкова кількість реалізацій, яка потрібна для "стабілізації" дисперсії, а z_γ – це коефіцієнт, який знаходиться з рівняння $2\Phi(z) = 1-\gamma$ ($\Phi(z)$ – це функція Лапласа).

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 α: 1, 0.3 та 0.1. Потрібно виконати наступні завдання.

A. При кожному $\alpha = 1;\ 0.3;\ 0.1$ обчислити точне значення ймовірності $Q(\alpha)$ (виключно для тестування всіх чотирьох алгоритмів).

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

$Q(\alpha) = \mathbf{M}\, I(\xi_\alpha < \eta)$, тобто $\hat{q}_i = I(\xi_\alpha^{(i)} < \eta_i)$, де $I(\cdot)$ – індикаторна функція.

C. *Метод 2:* $Q(\alpha) = \int\limits_0^\infty [1-G(u)]\,dF_\alpha(u) = \mathbf{M}[1-G(\xi_\alpha)]$, тобто

$$\hat{q}_i = 1 - G(\xi_\alpha^{(i)}) = e^{-\left[\xi_\alpha^{(i)}\right]^2}.$$

D. *Метод 3:* $Q(\alpha) = \int\limits_0^\infty F_\alpha(u)\,dG(u) = \mathbf{M}\,F_\alpha(\eta)$, тобто

$$\hat{q}_i = F_\alpha(\eta_i) = 1 - e^{-(\alpha\,\eta_i)^4}.$$

E. *Метод 4:*

$$Q(\alpha) = \int\limits_0^\infty F_\alpha(u)\,dG(u) = \int\limits_0^\infty F(u)\,g(u)\,du = \int\limits_0^\infty F_\alpha(u)\ \frac{g(u)}{h_\alpha(u)}\,h_\alpha(u)\,du = \mathbf{M}\left[F_\alpha(\gamma_\alpha)\ \frac{g(\gamma_\alpha)}{h(\gamma_\alpha)}\right],$$

де γ_α – невід'ємна в.в. із щільністю $h_\alpha(u),\ u \ge 0$. Використовуючи співвідношення (1), маємо

$$F_\alpha(u)\,g(u) = \left(1 - e^{-(\alpha\,u)^4}\right) 2u\,e^{-u^2} \underset{\alpha\to 0}{\sim} 2\alpha^4 u^5\,e^{-u^2}.$$

Із співвідношення (1) випливає, що $\int\limits_0^\infty u^5\, e^{-u^2}\, du = 1$. Тому як щільність $h_\alpha(u),\ u \ge 0$, раціонально вибрати $h_\alpha(u) = u^5\, e^{-u^2},\ u \ge 0$. Легко показати, що саме таку щільність має в.в. $\gamma_\alpha = \sqrt{\theta^{(1)} + \theta^{(2)} + \theta^{(3)}}$, де $\theta^{(1)}$, $\theta^{(2)}$ та $\theta^{(3)}$ – незалежні в.в., які мають показниковий розподіл з параметром 1, тобто $\theta^{(i)} = -\ln \omega_i$ (у нашому випадку в.в. γ_α та щільність розподілу $h_\alpha(u)$ від параметра α не залежать). Звідси випливає, що

$$\hat{q}_i = \left[1 - e^{-(\alpha\beta_i)^4}\right] \frac{2\beta_i\, e^{-\beta_i^2}}{\beta_i^5\, e^{-\beta_i^2}} = \frac{2}{\beta_i^4}\left[1 - e^{-(\alpha\beta_i)^4}\right],$$

де $\beta_i = \sqrt{\theta_i^{(1)} + \theta_i^{(2)} + \theta_i^{(3)}}$, причому $\theta_i^{(1)} = -\ln \omega_{3i-2},\ \theta_i^{(2)} = -\ln \omega_{3i-1},\ \theta_i^{(3)} = -\ln \omega_{3i}$, $i = 1, 2, \ldots$.

Для кожного значення параметра α для кожного з чотирьох методів виводити на друк:

- точне значення $Q(\alpha)$;
- оцінку;
- вибіркову дисперсію;
- довірчий інтервал;
- кількість виконаних реалізацій для побудови оцінки із достовірністю 0.99 та відносною похибкою 1%.