УДК 624.042

А.В. Махінько, к.т.н. доц.,
Полтавський національний технічний університет
імені Юрія Кондратюка

СПІВВІДНОШЕННЯ СТАТИСТИЧНИХ КРИТЕРІЇВ УЗГОДЖЕНОСТІ ТА ІМІТАЦІЙНОЇ ПРОЦЕДУРИ МОНТЕ-КАРЛО В ЗАДАЧАХ НАДІЙНОСТІ КОНСТРУКЦІЙ

Наводяться міркування про можливість і доцільність застосування статистичних критеріїв узгодженості для аналізу експериментальних вибірок, отриманих на основі імітаційної процедури Монте-Карло.

Приводятся некоторые соображения касательно возможности и целесообразности применения статистических критериев согласия для анализа экспериментальных выборок, полученных на основе имитационной процедуры Монте-Карло.

In this article possibility and expediency of application of consent statistical criteria for the analysis of the experimental data obtained on the basis of imitating procedure of Monte-Carlo is discussed.

Ключові слова: *статистичні методи, критерії узгодженості, метод Монте-Карло, експериментальні вибірки, випадкова величина.*

Постановка проблеми. В умовах сьогодення майже всі теоретичні та експериментальні дослідження вимагають обов’язкового залучення електронно-обчислювальної техніки. За її допомогою розв’язуються задачі, рішення яких не можна було уявити собі ще 20 − 30 років тому. Головним чином, це пояснюється здатністю комп’ютерів швидко і без випадкових помилок обробляти великі масиви інформації за малий проміжок часу.

Одним із цифрових алгоритмів, котрий часто залучають науковці до своїх досліджень, є імітаційна процедура Монте-Карло, в основу якої покладений нерекурентний ітераційний алгоритм прямого програмування. Перевагою процедури Монте-Карло є те, що вона не вимагає глибоких теоретичних знань дослідника в галузі теорії ймовірності й математичної статистики, а ефективність її застосування визначається лише потужністю комп’ютера користувача. Однак простота застосування даної процедури не поширюється на аналіз отриманих за її допомогою результатів. У цьому місці необхідна певна обережність і розуміння методологічних основ тих методів, що застосовуються для аналізу даних, одержаних за допомогою процедури Монте-Карло. Одними з таких методів аналізу в математичній статистиці є критерії узгодженості, які для наявних експериментальних даних дозволяють обґрунтувати чи спростувати певну теоретичну модель, що до них приміряється. Надалі ця теоретична модель приймається в якості еталонного відображення характеристик дослідної величини і на її основі формуються ієрархічні структури вищого рівня для ще однієї глибокої теорії у певній галузі науки чи техніки.

Таким чином, хибні висновки на етапі теоретичного обґрунтування експериментальних даних за допомогою критеріїв узгодженості можуть призводити до створення неправильних математичних моделей як окре-мих величин, так і цілих теорій, в яких вони фігурують. Ось чому питання співвідношення статистичних критеріїв узгодженості та імітаційної процедури Монте-Карло є завданням важливим і актуальним.

Аналіз останніх досліджень і виділення не розв’язаних раніше частин загальної проблеми. Методам статистичного моделювання у різних галузях науки присвячено доволі багато публікацій, серед яких найбільш відомими є роботи Д. Бендата, В.В. Бикова, Н.П. Бусленка, А. Пірсола [2 − 4]. Процедура Монте-Карло, яка є одним з основних методів статистичного моделювання, детально розглянута в роботах С.М. Єрмакова та І.М. Соболя [8, 9, 11]. Матеріал щодо критеріїв узгодженості можна знайти в роботах [1, 5 − 7, 10].

Структура цих робіт має ту особливість, що методи статистичного моделювання розглядаються окремо від методів статистичної обробки змодельованих даних, і, навпаки, у будь-якому посібнику з математичної статистики методи аналізу застосовуються до вже наявних даних. Яким чином були отримані ці дані, який чинник випадковості в них присутній, скільки серій дослідів було проведено над випадковою величиною, чи можна у даному випадку застосовувати критерії узгодженості – на це математична статистика відповіді не дає, як не дають відповіді й методи статистичного моделювання.

Формулювання цілей статті. У даній роботі автор ставив за мету дати відповідь на такі запитання, як можливість теоретичного обґрунтування даних імітаційного моделювання критеріями узгодженості та правомірність подальшого застосування таких теоретичних моделей на практиці.

Виклад основного матеріалу. Будемо вважати, що для розв’язання деякої задачі необхідно знайти закон розподілу випадкової величини $\tilde{Y}$, безпосередні експериментальні дослідження якої за незалежних від нас причин неможливі. Проте достеменно відомо, що величина $\tilde{Y}$ пов’язана строгою функціональною залежністю $\tilde{Y} = f(\tilde{X})$ із випадковою величиною $\tilde{X}$, для експериментального вимірювання якої немає ніяких перешкод.

Задача легко вирішується, якщо $f(\bullet)$ є лінійною функцією свого аргументу, й ускладнюється, якщо $f(\bullet)$ має чітко виражену нелінійну або трансцендентну структуру. На практиці найбільш часто зустрічається саме останній варіант, який і розглядається далі.

Припустімо, що функція $f(\bullet)$ настільки відрізняється від прямої пропорційності, що застосування методу лінеаризації [5, 6] для отримання вибіркового середнього та середньоквадратичного відхилення призводить до суттєвої похибки, а вираз для закону розподілу величини $\tilde{Y}$ має тільки складну інтегральну форму. Для уникнення цих складностей логічно застосувати процедуру Монте-Карло, суть якої полягає в наступному:

$$Y_i = f(X_i), \quad i = 1, 2 \ldots N, \tag{1}$$

де N – об'єм вибірки, або кількість розіграшів величини $\tilde{X}$.

Такі статистичні характеристики, як вибіркове середнє $\overline{Y}$, середньоквадратичне відхилення $\tilde{Y}$, асиметрія A_Y та ексцес A_E знаходяться досить просто для величини $\tilde{Y}$, чого не можна сказати про закон її розподілу.

Для теоретичного обґрунтування закону розподілу випадкової величини $\tilde{Y}$, по-перше, робиться візуальне оцінювання гістограми її роз-поділу, по-друге, підбирається теоретична крива розподілу, яка найкраще описує дослідну гістограму, по-третє, за допомогою критеріїв узгодженості обґрунтовується чи спростовується гіпотеза про можливість опису закону розподілу випадкової величини $\tilde{Y}$ даною теоретичною кривою.

Для обґрунтування теоретичного закону розподілу найбільш часто застосовують критерії Пірсона, Колмогорова та Крамера – Мізеса – Смірнова [5 − 7]. Ідея першого з них полягає у визначенні суми квадратів відхилень між теоретичними ймовірностями p_j та частотами p_j^*, за якими спостерігають і котрі взяті з деяким ваговим коефіцієнтом $c_j = 1 / p_j$. Сам критерій записується у вигляді

$$\chi^2 = N \sum_{j=1}^{k} \frac{(p_j^* - p_j)^2}{p_j} \le \chi_{kp}^2, \tag{2}$$

де $j = 1, 2 \ldots k$ – порядковий номер розряду статистичного ряду;

χ_{kp}^2 – критичне значення критерію при заданому рівні значимості.

Основна ідея критеріїв Колмогорова, Крамера – Мізеса − Смірнова і аналогічних їм полягає у вимірі відстані D_λ між функцією емпіричного $F^*(y)$ та функцією теоретичного $F(y)$ розподілу. Відрізняються ці критерії видом відстані у просторі функцій розподілу, а загальна формульна інтерпретація дається у вигляді

$$D_\lambda = \max\{F^*(y) - F(y)\}, \qquad \lambda = D_\lambda \sqrt{N}. \tag{3}$$

При цьому для узгодженості теоретичного та емпіричного розподілів величина λ повинна бути якомога меншою.

На перший погляд описана процедура перевірки узгодженості теоретичного й експериментального розподілу логічна і правомірна. Однак це не відповідає дійсності.

Випадкова величина $\tilde{X}$ має строге теоретичне обґрунтування і, як наслідок, теоретичний закон розподілу. Це, у свою чергу, обумовлює те, що випадкова величина $\tilde{Y}$ також має теоретичний закон розподілу. Інша річ, що він не виражається через елементарні функції або його побудова неможлива навіть при застосуванні чисельного інтегрування. Однак закон розподілу величини $\tilde{Y}$ об'єктивно існує! Коли у даній ситуації ми застосовуємо критерії узгодженості, то повністю заперечуємо існування наперед відомого закону розподілу величини $\tilde{Y}$ і «приміряємо» на неї інший закон, який, на нашу суб'єктивну думку, добре описує емпіричні дані. При цьому до уваги не береться сама ідея критеріїв узгодженості, котрі міру відмінності між теоретичним та емпіричним розподілом пояснюють наявністю випадкових причин, до яких належать спосіб групування даних, об'єм експериментальної вибірки, спосіб перевірки складних гіпотез і т.п. Якщо теоретична крива дійсно адекватно відтворює статистичний розкид деякої випадкової величини, то вплив випадкових факторів можна звести до нуля, збільшуючи обсяг вибірки, тобто спрямовуючи $N \to \infty$. У цьому випадку у формулах (2), (3) $\chi^2 \to 0$ при $\chi^2_{kp}(при\ \ N = const) = const$; $D_\lambda \to 0$, як наслідок, $\lambda \to 0$.

Що буде спостерігатися, коли об'єктивно існуючий теоретичний закон розподілу $f(Y)$ величини $\tilde{Y}$ намагатися замінити за допомогою критеріїв узгодженості псевдотеоретичним $f_T(Y)$? Якщо функція $f(Y)$ виражається в явному вигляді або процедура чисельного інтегрування не викликає значних труднощів, то для випадку $N \to \infty$ різниця ординат Δ експериментального полігону розподілу і теоретичного закону (різниця ординат Δ' відповідних функцій розподілу) завжди прямує до нуля

$$\Delta = f^*(y_j) - f(y_j) \to 0, \qquad \Delta' = F^*(y) - F(y) \to 0. \qquad \textbf{(4)}$$

Якщо для опису емпіричних даних застосовується псевдотеоретичний розподіл, то різниця ординат експериментального полігону і теоретичного закону прямуватиме не до нуля, а до деякої постійної величини, яка визначатиметься ступенем близькості псевдотеоретичного розподілу до теоретичного. У математичному вигляді це можна записати як

$$a_j = f(y_j) - f_T(y_j) = \lim_{N \to \infty} [f^*(y_j) - f_T(y_j)], \qquad \textbf{(5)}$$

$$b_j = F(y_j) - F_T(y_j) = \lim_{N \to \infty} [F^*(y_j) - F_T(y_j)]. \qquad \textbf{(6)}$$

Згідно з формулою (3) при $N \to \infty$ величина D_λ прямуватиме не до нуля, а до деякої константи $B = \max\{b_j\}$, що обумовлюватиме прямування величини λ до нескінченності. Наслідком цього буде висновок, що незалежно від ступеня близькості кривих $f^*(Y)$ і $f_T(Y)$ застосування останньої для опису експериментальних даних не можна обґрунтувати з позицій статистичних критеріїв узгодженості. Аналогічного висновку можна дійти і на основі критерію Пірсона. Для цього формулу (2) достатньо подати у вигляді, аналогічному виразам (5), (6),

$$d_j = (p_j - p_{T,j})^2 = \lim_{N \to \infty} [(p_j^* - p_{T,j})^2],$$

$$\chi^2 = N \lim_{N \to \infty} \sum_{j=1}^{k} \frac{d_j}{p_{T,j}} = \infty. \qquad \textbf{(7)}$$

Зроблений висновок зовсім не означає, що криву $f_T(Y)$ не можна застосовувати для опису експериментальних даних, він тільки наголошує на тому, що обґрунтовувати її використання за допомогою статистичних критеріїв узгодженості є некоректним і просто абсурдним. Навіть якщо при порівняно невеликих об'ємах емпіричних даних N критерії узгодже-ності дають такий бажаний позитивний результат.

Висновки. Таким чином, у статті розглянуті випадки, коли не можна користуватися критеріями узгодженості для обґрунтування чи спростуван-ня теоретичних законів з емпіричними даними, отриманими за допомогою імітаційної процедури Монте-Карло. Розуміння і чітке розмежування цих випадків дуже важливе для практики, тому що може призводити до прийняття хибних або відкидання правильних рішень. Узагалі методи статистичного моделювання не слід вважати за панацею від нашої неспро-можності розв'язати задачу теоретичним шляхом, вони можуть допомогти її розв'язанню, але аналіз результату завжди буде на совісті дослідника.

Література

1.Аугусти Г. Вероятностные методы в строительном проектировании / Г. Аугусти, А. Баратта, Ф. Кашиати. – М.: Стройиздат, 1988. – 584 с.

2.Бендат Д., Пирсол А. Измерение и анализ случайных процессов / Д. Бендат, А. Пирсол. – М.: Мир, 1974. – 464 с.

3.Бусленко Н.П. Метод статистического моделирования / Н.П. Бусленко. – М.: Статистика, 1970. – 112 с.

4.Быков В.В. Цифровое моделирование в статистической радиотехнике / В.В. Быков. – М.: Советское радио, 1971. – 328 с.

5.Вентцель Е.С. Теория вероятностей: учеб. для вузов / Е.С. Вентцель. – М.: Высш. шк., 2001. – 575 с.

6.Вентцель Е.С., Овчаров Л.А. Теория вероятностей и её инженерные приложения / Е.С. Вентцель, Л.А. Овчаров. – М.: Наука. Гл. ред. физ.-мат. лит., 1988. – 480 с.

7.Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для вузов / В.Е. Гмурман. – М.: высш. шк., 1979. – 400 с.

8.Ермаков С.М. Метод Монте-Карло и смежные вопросы / С.М. Ерма-ков. – М.: Наука, 1975. – 471 с.

9.Ермаков С.М., Михайлов Г.А. Курс статистического моделирования / С.М. Ермаков, Г.А. Михайлов. – М.: Наука, 1976. – 235 с.

10. Жлуктенко В.І., Наконечний С.І. Теорія ймовірностей і математична статистика / В.І. Жлуктенко, С.І. Наконечний. − К.: КНЕУ, 2000. − 304 с.

11. Соболь И.М. Численные методы Монте-Карло / И.М. Соболь. − М.: Наука, 1973. − 73с.

Надійшла до редакції 17.04.2009