

УДК 517.8

В. В. Колодний, к. т. н., доц.;

В. В. Седлецький, студ.

МЕТОДИ ГЕНЕРАЦІЇ ДЕВІАНТНО-НОРМАЛЬНИХ РОЗПОДІЛІВ ЙМОВІРНОСТЕЙ

В роботі [1] були описані основні поняття та умови доцільності використання класу девіантно-нормальних розподілів (ДНР) в імітаційному моделюванні, а також розроблений узагальнений алгоритм генерації ДНР. Ця стаття розвиває теоретичні результати, які були викладені раніше [1], і присвячена вона одному з можливих підходів до практичної реалізації комп'ютерних генераторів ДНР, що мають певні задані властивості.

Нагадаємо, що девіантний розподіл (ДР) – це близький до відомого типового (базового) закону розподіл ймовірностей, який має певні визначені суттєві особливості (відмінності від базового

закону), що називаються девіантностями.

Якщо за базовий виступає нормальний розподіл (БНР), то відповідний ДР називається девіантно-нормальним розподілом.

Девіантні розподіли, як правило, не виражаються аналітичними формулами для щільності ймовірностей (диференціальної функції розподілу). ДР отримують за допомогою комп'ютера методами імітаційного моделювання, тобто будь-який ДР – це емпіричний розподіл, який отримано певним алгоритмічним перетворенням базових випадкових чисел $R \in (0; 1)$.

Спочатку базові випадкові числа R за відомими методами перетворюються в базові розподіли (БР), а потім БР певним чином корегуються з урахуванням заданих девіантностей (рис. 1).

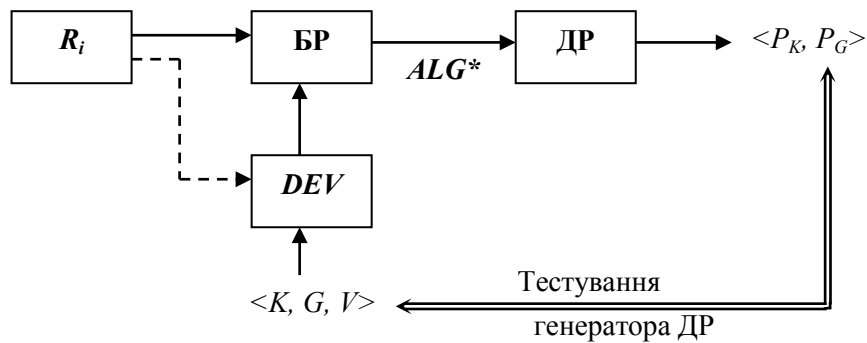


Рис. 1

Опис девіантностей має вигляд кортежу

$$DEV: \langle K, G, V \rangle,$$

де K – це множина числових елементів опису (параметрів); G – множина графічних елементів опису; V – множина вербальних елементів опису.

В загальному випадку описані девіантності не обов'язково однозначно визначають ДР, тобто можливе існування сімейства ДР із заданими властивостями. Множина алгоритмів-реалізаторів ДР ALG_{DEV} є нескінченною, тому постає задача вибору конкретного алгоритму

$$ALG^* \in \{ALG_{DEV}\}.$$

Комп'ютерна програма, що реалізує ALG^* , називається генератором ДР. Опис генератора ДР має вигляд кортежу

$$GEN: \langle BR, ALG^*, X_n^N \rangle,$$

де X_n^N — N вибірок обсягу n (вихід генератора).

Після генерації достатньої кількості значень випадкової величини, що має заданий ДР, необхідно провести тестування генератора ДР, тобто виконати перевірку адекватності генератора вимогам до ДР. Згенерований ДР характеризується скінченною множиною вихідних параметрів P , які можна розглядати або як числові значення $p_i \in \{PK\}$, або як графічні особливості гістограм $p_i \in \{PG\}$, тобто

$$P = P_K \cup P_G.$$

Тестування генераторів ДР полягає в перевірці відповідності (несуперечливості) множин отриманих параметрів PK, PG та множин елементів опису девіантностей K, G, V .

Клас нормальних розподілів є найбільш дослідженим в теорії ймовірностей та математичній статистиці, що обумовлює широке застосування генераторів нормальних розподілів в імітаційному моделюванні [2, 3, 4].

В реальності розподіли ймовірностей часто відрізняються від теоретичних нормальних, залишаючись при цьому майже нормальними, нормальноподібними [5]. Саме в цих випадках доцільно застосовувати генератори ДР з визначеними девіантностями [1].

Однією з найпоширеніших девіантностей нормальних розподілів є розтягнутість хвостових частин [5]. Девіантності цього виду можуть бути досить легко описані експертами-статистиками, але, як правило, не враховуються при побудові імітаційних моделей. Такі методичні похибки можуть призводити до втрати адекватності розроблених моделей та до зниження достовірності результатів моделювання.

Нормальні розподіли, що мають девіантності типу «розтягнутість хвостів», можуть бути згенеровані різноманітними способами. В цій статті розглядається один з ефективних підходів до генерації ДНР з такою девіантністю, заснований на генерації сумішей розподілів.

Сутність цього підходу полягає в змішуванні БНР та іншого (додаткового) розподілу в певній пропорції. Вид та параметри додаткового розподілу, а також пропорція суміші (ймовірність генерації додаткового розподілу p) залежать від конкретного опису девіантності DEV.

Розглянемо такий опис девіантності нормального розподілу

$$DEV: \langle K, G, V \rangle, V = \{v_1\},$$

де v_1 – вербальний опис девіантності: «розтягнутий правий хвіст».

Визначимо деякі з елементів множини числових параметрів опису $k_i \in K$.

Відомо, що базові нормальні розподіли повністю визначаються двома параметрами: математичним сподіванням M і середньоквадратичним відхиленням σ .

З метою спрощення задачі генерації ДНР доцільно розглядати стандартизований БНР з параметрами $M = 0$, $\sigma = 1$, а потім переходити до розподілів з довільними параметрами $-\infty < M < +\infty$, $\sigma > 0$ за формулою

$$\eta = M + \sigma \eta_{ст.}$$

Для теоретичних нормальних розподілів завжди коефіцієнт асиметрії $As = 0$, а коефіцієнт ексцесу $Ex = 3$. Тому ці числові параметри разом з іншими можна розглядати як діагностичні з тестуванням БНР та ДНР:

$$As \in K; \quad Ex \in K.$$

Очевидно, що девіантність «розтягнутість правого хвоста» впливає на перелічені числові параметри. Для відповідних стандартизованих ДНР виконуються такі співвідношення:

$$M > 0; \quad \sigma > 1; \quad As > 0; \quad Ex > 3.$$

На жаль, сукупність цих числових параметрів не може достатньою мірою характеризувати ступінь розтягнутості хвоста, тому пропонується ще один параметр $\chi \in K$ – доля правого хвоста в загальній гістограмі ДНР.

Правим хвостом стандартизованого БНР логічно вважати частину графіку щільності розподілу правіше точки перегину $x = 1$. Виходячи з цього, можна обчислити значення параметра χ_0 – долю правого хвоста для стандартизованого БНР

$$\chi_0 = \left(\int_1^{\infty} \frac{1}{2\pi} e^{-\frac{x^2}{2}} \right) / \left(\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{x^2}{2}} \right) = 0,5 - \Phi(1,0) = 0,5 - 0,34134 = 0,15866,$$

де $\Phi(1,0)$ – табличне значення інтегралу ймовірностей для точки $x = 1$.

Таким чином, якщо $\chi > \chi_0$, то це свідчить про розтягнутість правого хвоста згенерованого ДНР: чим більше значення χ , тим більше ступінь розтягнутості хвостової частини розподілу.

Параметр $\chi \in K$ є основним параметром, що використовується у чисельному описі девіантності «розтягнутість правого хвоста». Прикладами таких описів можуть бути висловлювання статистиків-експертів на зразок: «Правий хвіст розподілу займає не менше 18 % площі під гісторгамою» або «В правий хвіст іде приблизно п'ята частина всього розподілу». Зрозуміло, що для першого прикладу $\chi_1 \geq 0,18$, а для другого $\chi_2 \approx 0,2$.

У разі вербального опису цієї ж девіантності експерти використовують, як правило, такі зна-

чення лінгвістичної змінної «ступінь розтягнутості правого хвоста» [5]:

трохи, більш, суттєво, значно, дуже.

Для того, щоб «об'єктивізувати» такі нечіткі оцінки, можна застосувати пробну генерацію ДНР з різними значеннями χ . Після отримання та вивчення гістограм ДНР експерт характеризує девіантність вербально, тобто виконує переклад

$$P_G \Rightarrow V.$$

Аналізуючи разом з експертом значення параметра $\chi \in P_K$, можна зробити переклад відповідного вербального опису в чисельний:

$$V \Rightarrow K.$$

Таким чином, в процесі формування графічно-вербальних та чисельно-вербальних описів девіантностей відбувається налаштування генератора ДНР на знання конкретного експерта, тобто кожному значенню лінгвістичної змінної ставиться у відповідність область числових значень параметра χ .

Серед методів, заснованих на генерації сумішей різних розподілів, можна виділити два різновиди: метод часткового домішування і метод суміші-композиції.

Спочатку опишемо алгоритм генерації ДНР за допомогою методу часткового домішування.

1. Згенерувати базове випадкове число $R \in (0; 1)$ та порівняти його із значенням ймовірності генерації додаткового розподілу ρ .
2. Якщо $R < \rho$, то перейти до п. 3, інакше перейти до п. 5.
3. Згенерувати випадкове число x_i за додатковим законом розподілу суміші.
4. Перейти до п. 6.
5. Згенерувати випадкове число x_i за стандартизованим БНР.
6. Порівняти загальну кількість згенерованих за заданим ДНР випадкових чисел із заданою.
7. Якщо кількість згенерованих за заданим ДНР випадкових чисел недостатня, то перейти до п. 1, інакше перейти до п. 8.
8. Кінець.

В результаті виконання цього алгоритму формується вибірка X_n випадкових чисел, що мають заданий ДНР.

Тепер розглянемо алгоритм генерації ДНР за допомогою *методу суміші-композиції*.

1. Згенерувати базове випадкове число $R \in (0; 1)$ та порівняти його із значенням χ .
2. Якщо $R < \chi$, то перейти до п. 3, інакше перейти до п. 5.
3. Згенерувати випадкове число $x_i > 1$ за додатковим законом розподілу суміші (генерація хвостової частини ДНР).
4. Перейти до п. 6.
5. Згенерувати випадкове число $x_i \leq 1$ за базовим стандартизованим нормальним розподілом (генерація зрізаного справа БНР).
6. Порівняти загальну кількість згенерованих за заданим ДНР випадкових чисел із заданою.
7. Якщо кількість згенерованих за заданим ДНР випадкових чисел недостатня, то перейти до п. 1, інакше перейти до п. 8.
8. Кінець.

В результаті виконання цього алгоритму отримуємо потрібну кількість випадкових чисел (вибірку X_n), розподілених за заданим ДНР. Слід підкреслити, що в методі суміші-композиції основна та хвостова частини ДНР завжди генеруються за різними законами розподілу.

Для дослідження запропонованих методів та відповідних алгоритмів була розроблена комп'ютерна програма, яка дозволила промоделювати роботу генератора ДНР, отримати діагностичні параметри та оцінити ступінь адекватності генераторів. Окремі результати роботи програми з використанням методу часткового домішування наведені на рис. 2—5, а з використанням методу суміші-композиції – на рис. 6, 7.

Кожному пікселю гистограми відповідає одне значення випадкової величини χ_i , що має заданий ДНР. Загальний обсяг вибірки $n = 50000$. Для полегшення аналізу графічних особливостей отриманих гистограм P_G побудовані теоретичні криві щільності розподілу стандартизованого БНР в межах $-3 \leq x < 3$. Поруч з гистограмами наведені значення основних числових параметрів $p_i \in P_K$.

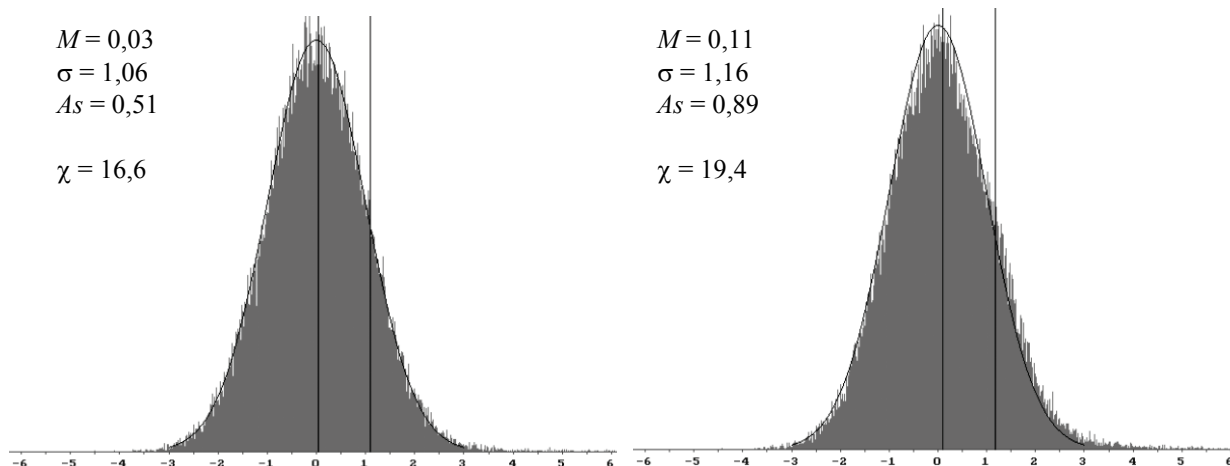


Рис. 2

Рис. 3

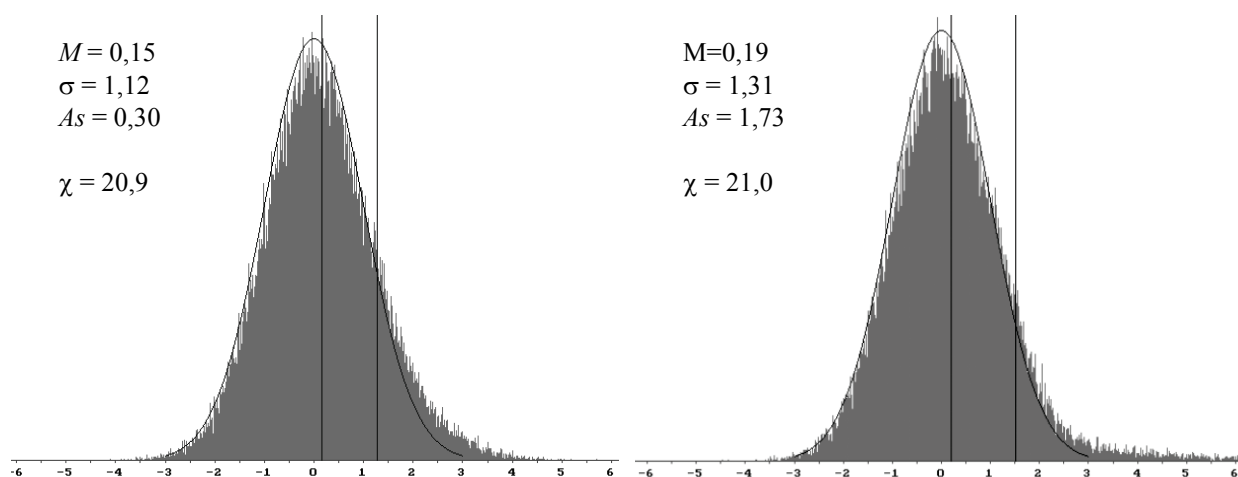


Рис. 4

Рис. 5

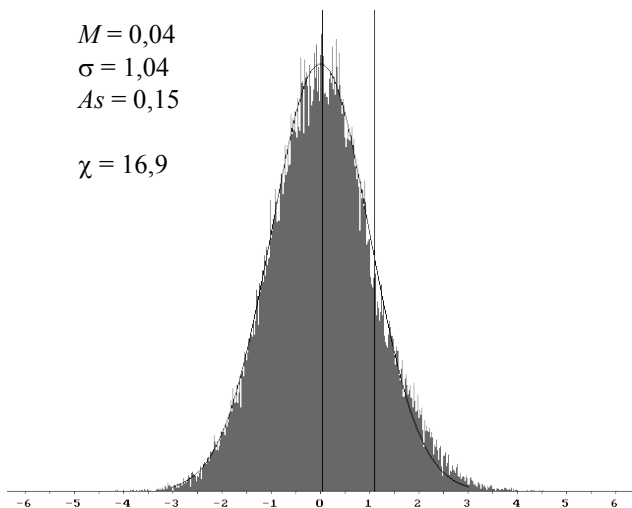


Рис. 6

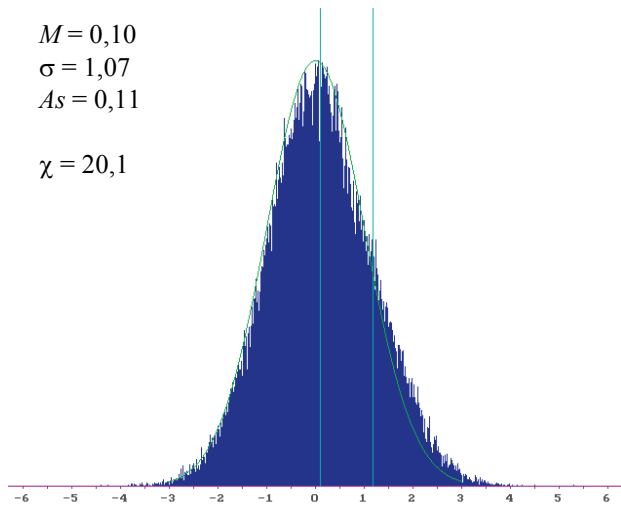


Рис. 7

Висновки

1. Запропоновані два методи генерації девіантно-нормальних розподілів: метод часткового домішування та метод суміші-композиції.
2. Розроблені відповідні алгоритми генерації ДНР для випадку розповсюдженої девіантності типу «розтягнутість правого хвоста».
3. Запропоновані числові параметри, які дозволяють статистикам-експертам об'єктивніше описувати девіантності розподілів ймовірностей.
4. Запропонована методика тестування генераторів ДНР, яка засновується на порівняльному аналізі відповідних числових, графічних і вербальних елементів опису девіантностей та отриманих після генерації числових параметрів і гістограм.
5. Розроблена комп'ютерна програма-генератор ДНР, з допомогою якої можна моделювати та вивчати властивості запропонованих методів та алгоритмів.

СПИСОК ЛІТЕРАТУРИ

1. Колодний В. В. Застосування класу девіантно-нормальних розподілів в імітаційному моделюванні // Вісник ВПІ. — 2001. — № 2. — С. 82—85.
2. Клейнен Дж. Статистические методы в имитационном моделировании. Вып. 1. — М.: Статистика, 1978. — 221 с.
3. Клейнен Дж. Статистические методы в имитационном моделировании. Вып. 2. — М.: Статистика, 1978. — 335 с.
4. Харин Ю. С., Степанова М. Д. Практикум на ЭВМ по математической статистике. — Мн.: Изд-во «Университетское», 1987. — 304 с.
5. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. — Вып. 1. — М.: Финансы и статистика, 1982. — 317 с.

Рекомендована кафедрою інтелектуальних систем

Надійшла до редакції 15.03.04
 Рекомендована до друку 7.04.04

Колодний Володимир Володимирович — доцент кафедри інтелектуальних систем, **Седлецький Вячеслав Валерійович** — студент Інституту інформаційних технологій та комп'ютерної інженерії.
 Вінницький національний технічний університет