

НОВ КРИТЕРИЙ ЗА КАЧЕСТВЕНА ОЦЕНКА НА КОМПЮТЪРНИ ГЕНЕРАТОРИ НА СЛУЧАЙНИ ЧИСЛА С НОРМАЛНО РАЗПРЕДЕЛЕНИЕ

Димитър Ст. Тянев

УДК 519.23. Като критерий за оценка на компютърни генератори на случайни числа със стандартно нормално разпределение се предлага границата, съответстваща на зададено ниво на риска. Предложен е метод за изчисление на критерия и резултати от неговото приложение.

I. Увод

Компютърното моделиране и методът Монте-Карло се основават на компютърен модел на базова случайна величина (БСВ). Експерименталното моделиране изследва поведението на обектите в условията на предварително зададени стойности на параметрите на закона, който реализира случайната величина. За съжаление практически е невъзможно да бъдат постигнати тези стойности, което е една обективна реалност, чиято същност е анализирана в [1]. Най-често се използва базова случайна величина X , която реализира стандартен нормален закон $N_x(0,1)$. На практика обаче получаваните с помощта БСВ статистически извадки имат средна стойност $\bar{X} \neq 0$ и средно квадратично отклонение $\overline{\sigma_x} \neq 1$. Както е показано в [2], оценките на тези параметри при малки по обем извадки са в значителна степен случайни. Компютърните алгоритми за реализация на БСВ се основават на известни теоретични положения [3, 4], които се постигат с приближение, когато обемът на генерираните извадки се стреми към безкрайност – непостижимо в практиката условие. Ето защо е актуална задачата за поправка на генерираните извадки така, че те да постигнат зададените параметри до употребата им в експерименталното изследване. Задачата се решава чрез преобразуване на случайната величина X в случайна величина Y от типа изложен в [1], което в едномерния случай има вида:

$$(1) \quad y_i = \frac{x_i - \bar{X}}{\sigma_x}; \quad i = \overline{1, n}$$

където с n е означен обемът на извадката.

II. Някои естествени оценки

Както в многомерния, така и в едномерния случай, след поправката параметрите на закона на разпределение могат да бъдат постигнати със завидна точност. Това обаче още не гарантира, че преобразуваната извадка съответствува на закона. Ето защо неговата теория дефинира и изследва централните моменти на разпределението от по-висок ред, чрез които се определят допълнителни негови характеристики, като например коефициентите на асиметрия и

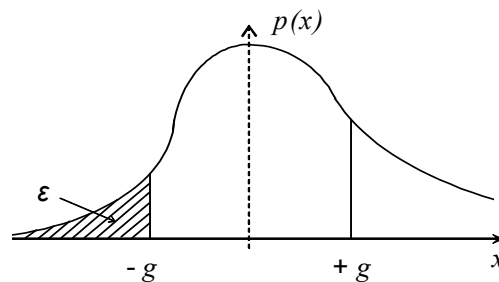
ексцес: $A = \frac{m_3}{\sigma^3}$; $E = \frac{m_4}{\sigma^4} - 3$. За разпределение, подчиняващо се на нормалния закон, тези

две характеристики трябва да бъдат равни на нула. В същото време те са едни от най-естествените оценки за едно емпирично разпределение. Дори чрез споменатото преобразуване на извадката да са постигнати отнапред зададените стойности за параметрите на разпределението, споменатите по-горе характеристики могат да изявят съществени дефекти. В случай на асиметрично разсейване например, извадковата средна стойност ще се различава от модата на разпределението. Ето защо коефициентите на асиметрия и ексцес могат да се възприемат като оценки за формата на разпределението. Методическата основа за

получаване на количествени и качествени оценки чрез тези характеристики може да бъде намерена например в [5].

III. Рискът като източник за оценка

В редица практически проблеми, при действия предприемани в условия с много неизвестни, се изхожда от някаква оценка на риска (или на грешката) ε . Рискът се определя като вероятност, с която едно събитие може да реализира определени стойности. В крайна сметка това означава, че трябва да се определя площта, затворена от закона за разпределение на плътността и граница g , както е илюстрирано на фигура 1.



Фиг. 1 Стойност и граница на риска

В случаите, когато генерираните за експеримента данни трябва да съответствуват в най-висока степен на нормалния закон на разпределение, тази вероятност може да се въведе в качеството си на критерий за оценка на това съответствие. Ясно е, че колкото по-близка е изчислената стойност на грешката до теоретичната стойност, толкова по-силно е съответствието между емпиричното и теоретичното разпределение. При това тази оценка характеризира закона не в една единствена точка, а интегрално по цялата дължина на интервала, в който попада кривата на закона. В случай на нормален закон лявата едностранна и дясната едностранна грешки съвпадат. Ако това положение се установи за едно емпирично разпределение, това ще означава, че поведението на функцията на закона вляво от границата $(-g)$ и вдясно от границата $(+g)$ (виж фигура 1) е огледално и може да представлява една силна оценка за качеството на компютърния генератор. Силата на оценката се състои в това, че тя се отнася за малко-вероятните области на разпределението, които се формират от относително малък брой реализации във всяка отделно изследвана извадка.

На практика по-често възниква обратната задача – определяне на границата при зададено от потребителя приемливо ниво на риска.

IV. Метод за изчисляване на границата

Определянето на едностранната границата g , която съответствува на зададения риск ε , в практиката означава, че изчисленията ще се провеждат в условията на неизвестен закон за разпределение и неизвестни негови параметри. Единственото на което може да се разчита в този случай е емпиричното разпределение на плътността на нормираните относителни честоти ω_j при деление на интервала с равномерна стъпка h :

$$(2) \quad P_j(x_{j-1} < X < x_j) = \omega_j \quad j = \overline{1, l}$$

където стъпката h представлява разстоянието между два съседни възела: $x_j - x_{j-1} = h$.

Възможностите за определяне на границата са две:

- чрез вероятността отляво:

$$(3) \quad P(X \leq g_L) = \varepsilon;$$

- чрез вероятността отдясно:

$$(4) \quad P(X \geq g_R) = 1 - \varepsilon;$$

където g_L и g_R са най-близките възможни стойности на границата g .

Вероятностите (3) и (4) се постигат като последователно натрупващи се суми, които трябва да удовлетворят следните условия:

$$(5) \quad \sum_{j=1}^{d_L} \omega_j \geq \varepsilon; \quad j = 1, 2, 3, \dots, d_L; \quad (d_L < l);$$

$$(6) \quad \sum_{j=l}^{l-d_R} \omega_j \geq 1 - \varepsilon; \quad j = l, (l-1), (l-2), \dots, (l-d_R); \quad (d_R < l).$$

Търсената граница съществува в интервала $[g_L = d_L \cdot h; \quad g_R = (d_R + l) \cdot h]$.

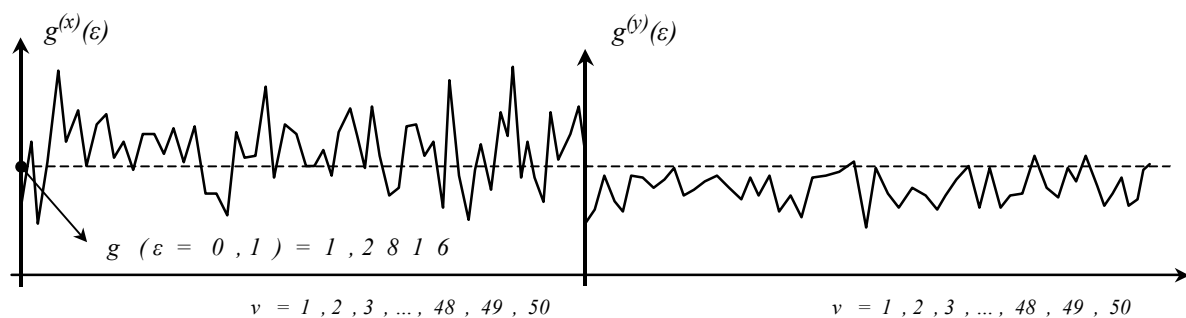
Ако се организира многократно изчисление на границите при различно деление на изследвания интервал на случайната величина ще се получи множеството реализации $\{g_L^k; g_R^k\}; \quad k = \overline{1, q}$. Тези $2 \cdot q$ на брой реализации за стойностите отляво и отдясно на границата могат да се приемат като независими събития и тогава по силата на теоремата за най-доброто приближение [6] най-добрата оценка на математическото очакване за търсената граница ще бъде:

$$(7) \quad M(g) = \frac{1}{2 \cdot q} \cdot \sum_k (g_L^k + g_R^k).$$

V. Експериментално потвърждение

Работоспособността на предложения метод за оценка на границата беше изследвана върху извадки, генерирани в съответствие със стандартния нормален закон на разпределение. Получаваните резултати бяха сравнявани с теоретичната стойност на границата, която за контрол може да бъде намерена в табличен вид в [6], при това за различни предварително избрани стойности на риска.

На фигура 2 графично е представено как се променя стойността на изчислената по предложения метод граница, при зададено ниво на грешката $\varepsilon = 0,1$. Теоретичната стойност, съответстваща на това зададено ниво е $g = 1,2816$. В лявата половина на графиката от фигура 2 са представени изчислените стойности на границата $g^{(x)}(\varepsilon)$ относно нивото на нейната теоретична стойност, за 50 извадки $v = \overline{1, 50}$, всяка с обем от $n=10000$ числа. В дясната половина на графиката са представени същите изчисления, но след като всяка от извадките е била поправена, т.е. за $g^{(y)}(\varepsilon)$ (виж [1] и (1)). Очевидно е, че направената поправка на данните е повишила степента на тяхното съответствие спрямо нормалния закон на разпределение. А това е още една пресечна точка на всички твърдения, които бяха изказани по-горе.



Фиг. 2 Стойности на изчислената граница върху извадки преди и след поправка

Получаваните стойности на границата имат случайно поведение. Отклоненията за непоправените извадки не надвишават нивото $\pm 4,5\%$ относно теоретичната стойност, като след поправка нивото на отклоненията спада двойно. Необходимо е да отбележим, че тези

резултати са получени при използване на конкретни алгоритми за равномерно и за нормално генериране. Именно тези алгоритми могат да бъдат изследвани в смисъла на предлагания тук нов критерий за оценка на тяхното качество.

Литература:

- [1]. Тянев Д. С., *Компютърно генериране на случайни вектори с гарантиране на статистическите параметри*, списание "Електротехника и електроника", брой 5-6, 1999 год., стр. 37-41.
- [2]. Венцель Е. С., *Теория вероятностей*, Москва, Издательство "Наука", 1964 год.
- [3]. Knut D. E., *The Art of Computer Programming. Seminumerical Algorithms*, vol. 2, 2nd ed., 1981, Addison Wesley, Reading, MA.
- [4]. Харин, Степанова, *Практикум на ЭВМ по математической статистике*, Минск, Издательство "Университетское", 1987 год.
- [5]. Гмурман В. Е., *Теория вероятностей и математическая статистика*, Москва, Издательство "Высшая школа", 1977 год.
- [6]. Lothar Sachs, *Statistische Auswertungs-methoden*, Springer – verlag, Berlin, Heidelberg, New York, 1972.