НОВЫЙ КРИТЕРИЙ ДЛЯ ОЦЕНКИ КАЧЕСТВА КОМПЬЮТЕРНЫХ ГЕНЕРАТОРОВ НОРМАЛЬНО РАСПРЕДЕЛЕННЫХ СЛУЧАЙНЫХ ЧИСЕЛ

Тянев Димитър Стоянов

Кафедра "Вычислительная техника", ТУ-Варна, (Болгария)

dstyanev@ms.ieee.bg ; dstyanev@mbox.digsys.bg

Abstract

Тянев Д.С. A new criterion for quality assessment of computer generators of random numbers with normal distribution. *As a criterion for assessment of computer generators of random numbers with standard normal distribution, a limit is offered, corresponding to an estimated level of risk. A method for calculation of the criterion is being proposed, as well as the results of its application.*

Введение

Компьютерное моделирование и метод Монте-Карло основываются на численной модели некоторой базовой случайной величины (БСВ). Экспериментальное моделирование обычно исследует поведение объектов в условиях заранее заданных значений параметров закона распределения случайной величины. К сожалению, их практически невозможно достичь, что является объективной реальностью, сущность которой проанализирована в [1]. Чаще всего используют базовую случайную величину X, которая реализует стандартный нормальный закон $N_x(0,1)$. На практике однако получаемые с помощью БСВ статистические выборки имеют выборочное среднее значение $\overline{X} \neq 0$ и среднее квадратическое отклонение $\overline{\sigma_x} \neq 1$. Как показано в [2], оценки этих параметров в условиях малых выборок являются в значительной степени случайными. Компьютерные алгоритмы генерации БСВ исходят из теоретических положений [3,4], которые достигаются приближенно, когда объем генерируемых выборок стремится к бесконечности, что является практически непостижимым условием. Вот почему задача об "исправлении" генерируемой выборки является актуальной. Для экспериментального использования выборки ее вероятностные параметры должны точно достигать заданных значений. Эта задача решается с помощью преобразования X в Y, аналогично представленному в [1], которое в одномерном случае будет иметь вид:

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

где n - объем выборки.

1. *Некоторые естественные оценки*

Как в многомерном, так и в одномерном случае, после поправки, параметры закона распределения могут быть достигнуты с весьма высокой точностью. Однако это еще не обеспечивает полное соответствие закону распределения преобразованной выборки. Необходимо исследовать центральные моменты более высокого порядка, которыми определяются дополнительные характеристики, такие как, например, асимметрия и эксцесс: $A = \frac{m_3}{\sigma^3}$; $E = \frac{m_4}{\sigma^4} - 3$. Для распределения, подчиняющегося нормальному закону, эти две характеристики должны быть равными нулю. В это же время они являются самыми естественными оценками для данного эмпиричного распределения. Даже когда значения параметров достигнуты, эти характеристики могут выявить существенные дефекты. Когда рассеивание, например, асимметрично выборочное среднее будет отличаться от значения моды. Вот почему асимметрию и эксцесс воспринимают как оценки формы распределения. Методические основы получения количественных и качественных оценок этих характеристик можно найти, например в [5].

2. Риск как источник оценки

На практике в условиях со многими неизвестными часто исходят из понятия о приемлемой потере (скажем денежной), которая воспринимается как допустимое значение (граница) риска предпринимаемых действий - ε. Риск определяется как вероятность данного события реализовать определенные значения. В конечном счете, это означает, что необходимо вычислить площадь фигуры, ограниченной кривой закона распределения плотности и границей g, как показано на рисунке 1.

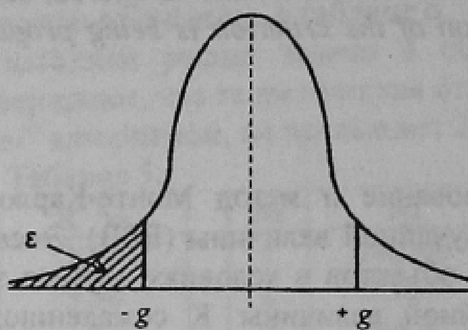

Рис. 1 Значения границы и риска

В случаях, когда генерируемые для эксперимента данные должны максимально соответствовать нормальному закону, эту вероятность можно ввести в качестве критерия оценки этого соответствия. Ясно, что чем ближе вычисленное значение ошибки к теоретическому, тем точнее соответствие между модельным и теоретическим распределениями. При этом надо понимать, что эта оценка характеризует закон не в одной единственной точке, а интегрально на протяжении всего интервала, в который попадает кривая закона. Когда он нормальный, левая и правая односторонние ошибки совпадают. Если такое положение будет установлено для данного модельного распределения, это будет означать что поведение функции закона влево от границе $(-g)$ и вправо от границе $(+g)$ зеркально (огледально), что является сильной оценкой качества компьютерного генератора. Достоинство такой оценки состоит в том, что она относиться к областям распределения с малой вероятностью, которые формируются относительно малым количеством реализаций в каждой исследуемой выборке.

В практике чаще возникает обратная задача – определить границу, когда пользователем задано приемлемое значение риска.

3. Метод вычисления границы

Определение значения границы g, соответствующей заданному риску ε в действительности означает, что вычисления будут производится в условиях неизвестного закона распределения, когда его параметры тоже неизвестны. То, на что можно рассчитывать в этом случае есть эмпирическое распределение плотности нормированных относительных частот ω_j, когда интервал реализации случайной величины делится равномерным шагом:

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

Возможно два определения границы:

- чрез вероятность слева:

$$P(X \le g_L) = \varepsilon\ ; \tag{3}$$

- чрез вероятность справа:

$$P(X \ge g_R) = 1 - \varepsilon\ ; \tag{4}$$

где g_L и g_R есть самые близкие возможные значения границы g.

Вероятности (3) и (4) достигаются как последовательно накапливающиеся суммы, которые должны удовлетворять следующим условиям:

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

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

Искомая граница существует в интервале $[g_L = d_L.h; \;\; g_R = (d_R + 1).h]$.

Если организовать многократное вычисление границ разными разбиениями исследуемого интервала случайной величины получится множество реализаций $\{g_L^k; g_R^k\}$; $k = \overline{1,q}$. Эти $2.q$ реализации могут быть приняты как независимые события и тогда на основание теоремы самого лучшего приближения [6], самая лучшая оценка математического ожидания искомой границы будет:

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

4. Экспериментальное исследование

Работоспособность предложенного метода для оценки значения границы риска была исследована выборками, сгенерированными в соответствием со стандартным нормальным законом распределения. Полученные результаты были сопоставлены с теоретическим значением границы, которое для контроля можно найти в соответствующей таблице, например в [6].

На рисунке 2 графически представлено изменение левостаронной границы, вычисленной предложенным методом, когда задана ошибка $\varepsilon = 0,1$. Теоретическое значение, которое ей соответствует, есть $g = 1,2816$ и находится оно на уровне абсциссы. В левой половине рисунки изображено изменение вычисленной границы для 50 последовательно сгенерированных выборок, каждой с объемом 10000 чисел. В правой половине рисунки изображено те же результаты, вычисленные после того, как каждая выборка была преобразована. Очевидно, что сделанная поправка при генерации выборок привела к повышению степени соответствия нормальному закону, что является еще одним подтверждением положений, приведенных выше.

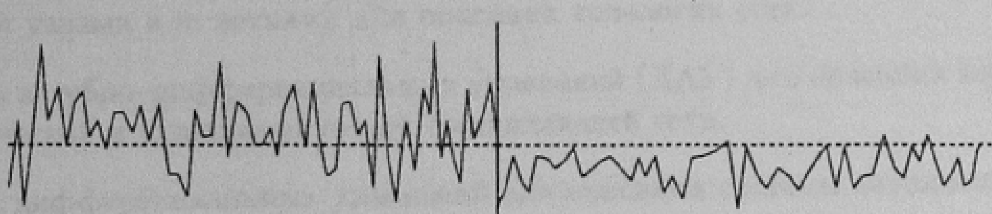

Рис. 2 Значения вычисленной границы до и после поправки выборок

Получаемые значения границы имеют случайное поведение. Отклонения границы относительно ее теоретического значения не превышают уровень ±4,5%. Однако после поправки отклонения уменьшаются в два раза. Необходимо отметить еще, что эти результаты получены при использования конкретных алгоритмов для равномерного и для нормального генераторов. Именно эти алгоритмы могут быть исследованы в смысле предлагаемого здесь критерия для оценки их 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.

Поступила в редакційну колегію 1.03.2000 р.