Бутаев М.М., Тарасов А.А. Расчет времени обработки запроса в информационной системе. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей XVIII Междунар. научно-техн. конф. – Пенза: ПДЗ, 2018. – С. 5-9.

УДК 004.9:519.2

РАСЧЁТ ВРЕМЕНИ ОБРАБОТКИ ЗАПРОСА В ИНФОРМАЦИОННОЙ СИСТЕМЕ

М.М. Бутаев, А.А. Тарасов

ANALYTICAL EVALUATION THE PROCESSING TIME OF A QUERY IN AN INFORMATION SYSTEM

M.M. Butaev, A.A. Tarasov

Аннотация. Предложено для случайных величин, определённых только на положительной оси координат, использовать конечно-интер-вальные законы распределения, например бета-распределение. Выведены формулы плотности и функции вероятности для сумм двух, трёх и четырёх независимых равномерно распределённых случайных величин.

Ключевые слова: плотность вероятности, вероятностные характеристики информационных систем, критерий Колмогорова-Смирнова.

Abstract. It is proposed for random variables, determined only on the positive coordinate axis, to use finite-interval distribution laws, for example, beta distribution. Density formulas and probability functions for sums of two, three and four independent uniformly distributed random variables are derived.

Keywords: probability density, probabilistic characteristics of information systems, Kolmogorov-Smirnov test.

На начальных этапах проектирования информационных систем для оценки временных характеристик обычно используется аппарат теории вероятности. Продолжительность отдельных этапов обработки информации моделируется случайными величинами. Равномерный закон распределения продолжительности операции в наибольшей степени подходит при минимуме данных, так как определяется только минимальной и максимальной продолжительностью этапа обработки.

Оценка продолжительности обработки информации последовательной цепочкой элементов системы, очевидно, сводится к суммированию продолжительности обработки информации в каждом элементе. Математической моделью такого действия является определение вероятностных параметров суммы независимых случайных величин. Обычно на практике сумма равномерно распределённых величин оценивается только математическим ожиданием и дисперсией. Для более глубокой оценки временных характеристик необходима знать функцию плотности вероятности $\varphi(x)$ и/или функцию распределения $\Phi(x)$ непрерывной случайной величины.

Пусть случайные величины x_i $i \in 1,\dots, n$ независимы и равномерно распределены на интервалах $x_i \in [0; \beta_i]$, $\beta_i > 0$. Для упрощения дальнейших аналитических преобразований случайные величины упорядочены по возрастанию β_i. Верхние границы упорядоченных случайных величин обозначены переменными b_i. Условие упорядочения: $b_i \le b_{i+1}$, $i \in 1,\dots, n - 1$. Например, наибольшие значения случайной составляющей продолжительности первого этапа $\beta_1 = 3$, второго этапа $\beta_2 = 5$ и третьего $\beta_3 = 4$ переобозначаются $b_1 = 3$, $b_2 = 4$, $b_3 = 5$.

Плотность вероятности значений случайной составляющей i-го слагаемого цепочки:

$$\varphi_i(x) = \begin{cases} 0, x < 0; \\ \frac{1}{b_i}, x \le b_i; \\ 0, \text{иначе.} \end{cases} \qquad (1)$$

Используя непосредственный способ, функция плотности суммы двух случайных величин $\varphi_{12}(x)$ определяется через подстановку (1) в интеграл свёртки [1]. Сформулированная задача решена для частного случая b_i = a $i \in 1,\ldots, n$ [2]. Функции типа (1) являются обобщёнными, потому плотность сумм случайных величин описывается кусочно-полиномиальными функциями. Выполнив преобразования по каждому интервалу, получена функция плотности распределения суммы двух равномерно распределённых случайных величин:

$$\varphi_{12}(x) = \frac{1}{b_1 b_2} \begin{cases} x, 0 \le x \le b_1; \\ b_1, b_1 < x \le b_2; \\ -x + b_1 + b_2, b_2 < x \le b_1 + b_2; \\ 0, \text{иначе.} \end{cases}$$

Аналогично выводится функция полной вероятности:

$$\Phi_{12}(x) = \frac{1}{2 b_1 b_2} \begin{cases} x^2, 0 \le x \le b_1; \\ 2 b_1 x - b_1^2, b_1 < x \le b_2; \\ -x^2 + 2(b_1 + b_2)x - b_1^2 - b_2^2, b_2 < x \le b_1 + b_2; \\ b_1 b_2, x > b_1 + b_2. \end{cases}$$

Такие же преобразования выполнены для вывода функции плотности суммы трёх и четырёх равномерно распределённых случайных величин при $b_3 \le b_1 + b_2$ и $b_4 \le b_1 + b_2 + b_3$, которые описываются кусочно-полиномиальными функциями на 7 и 15 интервалах соответственно.

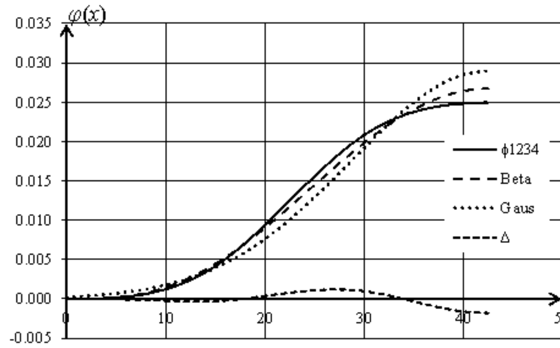
Полученные формулы применены для вычислений различных децилей. При вероятности завершения обработки информации, равной 0,1, в цепочке из четырёх элементов, имеющих только равномерно распределённые случайные составляющие продолжительности в секундах на интервалах [0;15], [0;15], [0;15] и [0;40], продолжительность обработки равна $x_{0,1}$ = 24,25 с. Для оценки рисков несвоевременного выполнения обработки данных используются вычисления дециля. Например, для вычисления продолжительности обработки информации в такой же цепочке при уровне риска 10% численно решается уравнение $\Phi_{1234}(x) = 1 - 0,1$, что даёт $x_{0,9}$ = 60,75 с. Для вычисления децилей использованы традиционные методы численного решения нелинейных уравнений.

Формулы для плотности и полной вероятности при большем количестве случайных величин $n > 4$ становятся громоздкими, но достаточно просто реализуемыми в программах для ЭВМ.

При промежуточных значениях n, когда точные формулы, выведенные для малых значений n, слишком сложны, а приближенные формулы, полученные из условия больших значений n (предельные теоремы теории вероятности), не адекватны, предлагается использовать аппроксимацию функции плотности вероятности. Функцию плотности, определённую на конечном интервале, целесообразно аппроксимировать конечно-интер-вальной функцией, например обобщённым бета-распределением. Параметры формы бета-распределения для симметричных в интервале функций равны $u = v$. Параметры формы предлагается определять с помощью аппарата проверки статистических гипотез.

Проверка гипотезы на отсутствие различия между значениями $\varphi_{123}(x)$ и бета-распределением для $n = 3$ $b_1 = b_2 = b_3 = 15$ даёт $u = v = 4$. В соответствии с критерием Колмогорова-Смирнова при максимальной разности между функциями, равной 0,0185, точном уровне значимости (P-value), равном 1,0, критическом уровне, равном 0,14117 для уровня значимости 0,05, вычисленные значения плотности $\varphi_{123}(x)$ можно считать бета-распре-делением. Наибольшая относительная погрешность аппроксимации $\Delta \approx 4\%$. Погрешность аппроксимации возрастает при $b_3 \le b_1 + b_2$, поскольку возникает участок с постоянным значением функции.

Проверка гипотезы на отсутствие различия между значениями $\varphi_{1234}(x)$ и бета-распределением для $n = 4$ $b_1 = b_2 = b_3 = 15$ и $b_4 = 40$ даёт $u = v = 4.3$. В соответствии с критерием Колмогорова-Смирнова при максимальной разности между функциями, равной 0,01556, точном уровне значимости (P-value), равном 1,0, критическом уровне равном 0,10385 для уровня значимости 0,05 вычисленные значения плотности $\varphi_{1234}(x)$ можно считать бета-распределением. Наибольшая относительная погрешность аппроксимации $\Delta \approx 7{,}1\%$ при $x = 0{,}5(b_1 + b_2 + b_3 + b_4)$. Погрешность аппроксимации возрастает при $b_4 \le b_1 + b_2 + b_3$. Полученные результаты проиллюстрированы на рисунке, на котором также нанесён график нормального распределения.



Графики плотности $\varphi_{1234}(x)$, бета-распределения (Beta) с $u = v = 4.3$, нормального распределения (Gaus) и относительной погрешности аппроксимации Δ для $n = 4$, $b_1 = b_2 = b_3 = 15$ и $b_4 = 40$ при $0 \le x \le 42{,}5$

Для случайных величин, определённых только на положительной оси координат, предложено применять конечно-интервальные законы распределения. Широко используемое нормальное распределение искажает результаты, и его можно использовать только как начальное приближение. В современных условиях доступны большие вычислительные ресурсы, поэтому предлагается использовать более сложные функции распределения случайных величин для более адекватного определения их характеристик.

Выведенные формулы для сумм двух, трёх и четырёх независимых равномерно распределённых случайных величин позволяют достаточно просто реализовать их в виде программы ЭВМ, обеспечивающей точность вычислений при небольших затратах процессорного времени.

Библиографический список

1. Вадзинский Р.Н. Справочник по вероятностным распределениям. СПб.: Наука, 2001. 295 с.
2. Феллер В. Введение в теорию вероятностей и её приложения. Том 2. М.: Мир, 1967. 746 с.

Бутаев Михаил Матвеевич
АО «НПП «Рубин»,
г. Пенза, Россия
E-mail: nts@npp-rubin.ru

Butaev M.M.
SC «SIE «Rubin»,
Penza, Russia

Тарасов Андрей Анатольевич
АО «НПП «Рубин»,
г. Пенза, Россия
E-mail: mail@npp-rubin.ru

Tarasov A.A.
SC «SIE «Rubin»,
Penza, Russia