

МОДЕЛИРОВАНИЕ УСЕЧЕННЫХ РАСПРЕДЕЛЕНИЙ С ЗАДАННОЙ КВАНТИЛЬЮ НА ОСНОВЕ ПОКАЗАТЕЛЬНОГО РАСПРЕДЕЛЕНИЯ

М.С.Токмачев

Институт электронных и информационных систем НовГУ, tokm@mail.natm.ru

Разработан метод моделирования усеченных вероятностных распределений на основе преобразования некоторого базового распределения. Одним из параметров является квантиль произвольного порядка. В случае базового показательного распределения в общем виде получены выражения для плотности, функции распределения, характеристической функции и моментов.

Ключевые слова: *усечение, квантиль, плотность, функция распределения, характеристическая функция, моменты*

The method of modeling of truncated probability distributions on the base of some exponential distribution changing is worked out. One of the parameters is quantile of random order. For exponential distribution expressions for density function, probability distribution function, characteristic function and moments were worked out in standard form.

Keywords: *truncation, quantile, density function, probability distribution function, characteristic function, moments*

Введение

В [1,2] представлен метод моделирования новых, усеченных на $[a;b]$ вероятностных распределений на основе базового теоретического распределения с функцией плотности $f(x)$. В качестве параметра вводится x_α — квантиль распределения порядка α , $x_\alpha \in (a;b)$, $0 < \alpha < 1$. Вместо постоянного нормирующего множителя на каждом из двух получившихся интервалов, $[a; x_\alpha)$ и $[x_\alpha; b]$, используется неотрицательная квадратичная функция $n_i x^2 + k_i x + c_i$, $i=1,2$. В частности, эти множители могут быть как линейными ($n_i = 0$), так и постоянными ($n_i = k_i = 0$), $i=1,2$. Базовая функция $f(x)$ имеет одно из заданных распределений. При базовой функции нормального закона случай линейных множителей рассмотрен в [1], а квадратичных — в [2]. Ввиду трансформации базового распределения введением квантили x_α и квадратичных множителей преобразованная функция плотности $g(x)$ на $[a;b]$, вообще говоря, весьма существенно отличается от $f(x)$. Совокупность задаваемых «квантильных характеристик», $\alpha, x_\alpha, g(x_\alpha)$, видоизменяет структуру базового распределения. Варьирование всеми параметрами функции $g(x)$ приводит к большому разнообразию новых распределений, среди которых встречаются и ранее известные. В частности, для базовой функции стандартного нормального закона усечение на положительную полуось с введением квантили x_α приводит к преобразованию классических распределений: при множителях $k_i x$ — распределения Рэлея, при множителях $n_i x^2$ — распределения Максвелла, а также распределений Накагами и Пирсона (χ -распределение) с конкретными параметрами. Если вводимые исследователем квантили совпадают с квантилями указанных распределений, то преобразование базовой функции стандартного нормального закона приводит к этим распределениям в чистом виде.

В предельном случае, когда квадратичные множители вырождаются в константу, а значение квантили совпадает с граничным значением ($x_0 = a$ или $x_1 = b$) приходим к стандартному классическому усечению.

В отличие от [1,2] в данной работе в качестве базовой функции рассмотрена плотность показательного закона распределения с произвольным параметром. При рассматриваемой структуре плотности вероятностей при некоторых значениях параметров в данный класс распределений, в частности, входят распределение Эрланга и χ^2 -распределение Пирсона. Найдены функции распределения и характеристические функции моделируемых распределений. Представлены формулы расчета моментов произвольного порядка. Методика компьютеризирована.

Построение функции плотности распределения вероятностей

Пусть задана случайная величина Z с плотностью распределения вероятностей $f(z)$. Для построения нового распределения, распределения случайной величины X , эту функцию полагаем базовой. Область значений Z включает в себя интервал усечения $[a;b]$. Вводом x_α , квантили порядка α для нового распределения, причем $x_\alpha \in (a;b)$, разобьем этот интервал на два: $[a; x_\alpha)$, $[x_\alpha; b]$. В каждом из полученных интервалов умножим функцию плотности исходного распределения на свой квадратичный множитель. Тогда функция плотности нового распределения имеет вид

$$g(x) = \begin{cases} (n_1 x^2 + k_1 x + c_1) f(x) & \text{при } x \in [a; x_\alpha), \\ (n_2 x^2 + k_2 x + c_2) f(x) & \text{при } x \in [x_\alpha; b], \\ 0 & \text{при } x \notin [a; b]. \end{cases} \quad (1)$$

При этом $\alpha = P(X \in [a; x_\alpha))$ — параметр концентрации (автоматически получаем $P(X \in [x_\alpha; b]) = 1 - \alpha$, обозначим $1 - \alpha = \beta$).

Коэффициенты n_i, k_i, c_i , ($i=1, 2$) в формуле (1) однозначно определяются как решения систем линейных алгебраических уравнений

$$\begin{cases} \int_a^{x_a} (n_1 x^2 + k_1 x + c_1) f(x) dx = \alpha, \\ (n_1 x_a^2 + k_1 x_a + c_1) f(x_a) = g(x_a), \\ (n_1 a^2 + k_1 a + c_1) f(a) = g(a); \\ \int_{x_a}^b (n_2 x^2 + k_2 x + c_2) f(x) dx = \beta, \\ (n_2 x_a^2 + k_2 x_a + c_2) f(x_a) = g(x_a), \\ (n_2 b^2 + k_2 b + c_2) f(b) = g(b). \end{cases} \quad (2)$$

Условие нормировки выполнено, так как $\int_{-\infty}^{+\infty} g(x) dx = \int_a^b g(x) dx = \alpha + \beta = 1$. Вторые уравнения в системах (2) устанавливают непрерывность функции $g(x)$ в точке x_a с заданной ординатой $g(x_a)$. Заметим, что, вообще говоря, непрерывность $g(x)$ в точке x_a не является обязательным условием распределения, и правые части вторых уравнений систем (2) можно задавать различными. Третьи уравнения систем формируют граничные условия. В частных случаях, например $k_i=0, c_i=0$ или $n_i=0$, уменьшается количество уравнений в (2), и вычисления упрощаются: в каждой из систем при двух неизвестных коэффициентах остаются два первых уравнения, при одном неизвестном — лишь одно первое уравнение.

Интегрирование в первом уравнении каждой из систем (2) приводит к линейному алгебраическому уравнению относительно n_i, k_i, c_i , ($i=1, 2$). При $a \neq x_a \neq b$ решение систем единственно. Заметим, что для произвольно заданной совокупности параметров не всегда следует вероятностное решение систем (2). По определению плотности для функции $g(x)$ также необходимо гарантировать неотрицательность введенных квадратичных множителей:

$$\begin{cases} (n_1 x^2 + k_1 x + c_1) \geq 0 \text{ при } x \in [a; x_a], \\ (n_2 x^2 + k_2 x + c_2) \geq 0 \text{ при } x \in [x_a; b]. \end{cases} \quad (3)$$

Лемма 1. Если для квадратного трехчлена $Q(x) = nx^2 + kx + c$ справедливы соотношения $Q(A) = 0$, $Q(B) = 0$, $n < 0$, то $Q(x)$ неотрицателен на $[A; B]$.

Лемма 2. Если квадратный трехчлен $Q(x) = nx^2 + kx + c$ не имеет корней на интервале $(A; B)$, $Q(A) \geq 0$, $Q(B) > 0$ (или $Q(A) > 0$, $Q(B) \geq 0$), то $Q(x)$ неотрицателен на $[A; B]$.

Поскольку на границах указанных в (3) интервалов значения трехчленов неотрицательны, $(g(a) \geq 0, g(x_a) \geq 0, g(b) \geq 0, f(a) > 0, f(x_a) > 0,$

$f(b) > 0$, что следует из определения плотности), то согласно леммам 1, 2 для выполнения (3) достаточно условия отсутствия корней каждого из трехчленов на соответствующем интервале, а при нулевых граничных условиях — отрицательности старшего коэффициента. Это контрольное условие вводится в компьютерную программу и выбраковывает совокупности коэффициентов, решений систем (2), приводящих к отрицательным значениям функции $g(x)$ на интервале.

Таким образом, с заданием значений $a, b, \alpha, x_a, g(a), g(b), g(x_a)$ функция плотности $g(x)$ для класса моделируемых распределений оказывается полностью определенной.

Базовая функция плотности показательного распределения

В качестве базовой функции введем плотность показательного закона

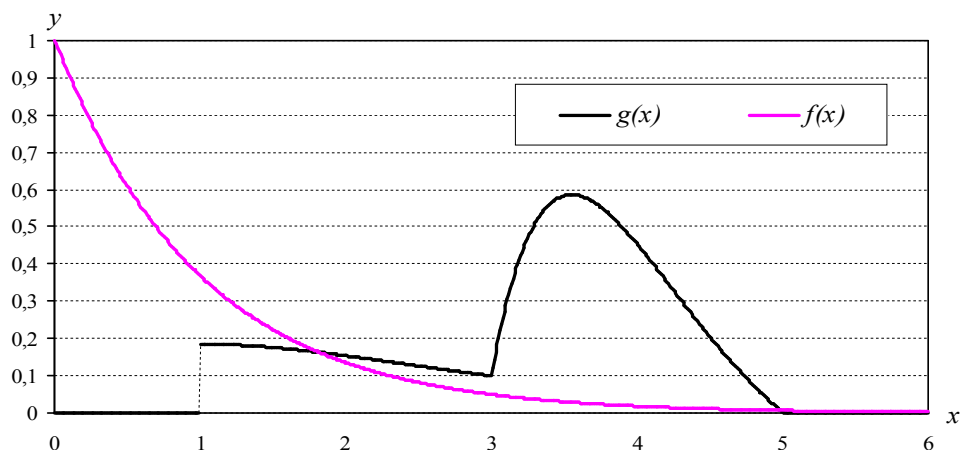
$$f(z) = \begin{cases} \lambda e^{-\lambda z} & \text{при } z \geq 0, \\ 0 & \text{при } z < 0, \end{cases}$$

где параметр $\lambda > 0$. Соответствующая функция плотности моделируемого распределения на $[a; b]$ при $0 \leq a < b \leq +\infty$ имеет вид

$$g(x) = \begin{cases} (n_1 x^2 + k_1 x + c_1) \lambda e^{-\lambda x} & \text{при } x \in [a; x_a], \\ (n_2 x^2 + k_2 x + c_2) \lambda e^{-\lambda x} & \text{при } x \in [x_a; b], \\ 0 & \text{при } x \notin [a; b]. \end{cases} \quad (4)$$

При подстановке выражения (4) в (2) коэффициенты n_i, k_i, c_i , ($i=1, 2$) вычисляются как решения систем трех линейных алгебраических уравнений:

$$\begin{cases} n_1 \left[\frac{f(a)}{\lambda^3} (\lambda^2 a^2 + 2\lambda a + 2) - \frac{f(x_a)}{\lambda^3} (\lambda^2 x_a^2 + 2\lambda x_a + 2) \right] + \\ + k_1 \left[\frac{f(a)}{\lambda^2} (\lambda a + 1) - \frac{f(x_a)}{\lambda^2} (\lambda x_a + 1) \right] + \\ + c_1 \left[\frac{1}{\lambda} (f(a) - f(x_a)) \right] = \alpha, \\ n_1 x_a^2 + k_1 x_a + c_1 = \frac{g(x_a)}{f(x_a)}, \\ (n_1 a^2 + k_1 a + c_1) = \frac{g(a)}{f(a)}; \\ n_2 \left[\frac{f(x_a)}{\lambda^3} (\lambda^2 x_a^2 + 2\lambda x_a + 2) - \right. \\ \left. - \frac{f(b)}{\lambda^3} (\lambda^2 b^2 + 2\lambda b + 2) \right] + k_2 \left[\frac{f(x_a)}{\lambda^2} (\lambda x_a + 1) - \right. \\ \left. - \frac{f(b)}{\lambda^2} (\lambda b + 1) \right] + c_2 \left[\frac{1}{\lambda} (f(x_a) - f(b)) \right] = \beta, \\ n_2 x_a^2 + k_2 x_a + c_2 = \frac{g(x_a)}{f(x_a)}, \\ n_2 b^2 + k_2 b + c_2 = \frac{g(b)}{f(b)}, \end{cases}$$



Графики функции плотности: $f(x)$ — базового показательного распределения с параметром $\lambda=1$; $g(x)$ — распределения, усеченного на $[1;5]$, с квантилью $x_{0,3}=3$

где $f(x)=\lambda e^{-\lambda x}$ — функция плотности вероятностей показательного распределения.

Характерный пример графика функции $g(x)$ при конкретных значениях параметров представлен на рис.

График $g(x)$ на рис. представляет плотность

$g(x)=$

$$= \begin{cases} (0,11517x^2 + 0,28934x + 0,09550)e^{-x} & \text{при } x \in [1;3], \\ (-23,60072x^2 + 188,05578x - 349,76048)e^{-x} & \text{при } x \in [3;5], \\ 0 & \text{при } x \notin [1;5]. \end{cases}$$

Это функция плотности при $\alpha=0,3$; $x_\alpha=3$; $g(1)=0,5f(1)=0,5e^{-1}$; $g(x_\alpha)=2f(x_\alpha)=2e^{-3}$; $g(5)=0,5f(5)=0,5e^{-5}$. Как легко заметить, график функции $g(x)$ существенно отличается от графика базовой функции $f(x)$, а введение квантили x_α по сути разбивает распределение на два самостоятельных с весовыми коэффициентами α и $1-\alpha$.

Моменты, характеристическая функция, функция распределения

Для дальнейших вычислений используем последовательность многочленов определенной структуры:

$$P_0(x)=1; P_1(x)=x+1; P_2(x)=x^2+2x+2;$$

$$P_3(x)=x^3+3x^2+6x+6; \text{ и т.д.}$$

$$P_4(x)=x^4+4x^3+12x^2+24x+24; \text{ и т.д.}$$

В общем виде

$$P_n(x) = \sum_{m=0}^n A_n^m x^{n-m}, \quad (5)$$

где $A_n^m = \frac{n!}{(n-m)!}$ при $0 \leq m \leq n$, $n=0,1,2,\dots$. При

$x \geq 0$, начиная с $n=1$, все указанные многочлены являются положительными возрастающими функциями. При этом $P_n(0)=n!$, а последовательность,

образуемая коэффициентами многочленов $\{1,1,1,1,2,2,1,3,6,6,1,4,\dots\}$, известна и представлена, например, в энциклопедии [3], раздел A008279. Последовательность $\{P_n(1)\}$ также представлена в [3], раздел A000522.

Теорема 1. Функция распределения, соответствующая плотности вероятностей $g(x)$ типа (4), имеет вид

$$G(x) = \begin{cases} 0 & \text{при } x \in (-\infty, a), \\ -e^{-\lambda x} \left[\frac{n_1}{\lambda^2} P_2(\lambda x) + \frac{k_1}{\lambda} P_1(\lambda x) + c_1 \right] + \\ + e^{-\lambda a} \left[\frac{n_1}{\lambda^2} P_2(\lambda a) + \frac{k_1}{\lambda} P_1(\lambda a) + c_1 \right] & \text{при } x \in [a; x_\alpha), \\ \alpha - e^{-\lambda x} \left[\frac{n_2}{\lambda^2} P_2(\lambda x) + \frac{k_2}{\lambda} P_1(\lambda x) + c_2 \right] + \\ + e^{-\lambda x_\alpha} \left[\frac{n_2}{\lambda^2} P_2(\lambda x_\alpha) + \frac{k_2}{\lambda} P_1(\lambda x_\alpha) + c_2 \right] & \text{при } x \in [x_\alpha; b], \\ 1 & \text{при } x \in (b; +\infty), \end{cases}$$

где $P_n(\dots)$ — многочлены вида (5).

Теорема 2. Характеристическая функция, соответствующая плотности вероятностей $g(x)$ типа (4), имеет вид

$$\begin{aligned} \chi(t) = & \frac{\lambda}{(\lambda - it)^4} [n_1 f_z(a) P_2((\lambda - it)a)] + \\ & + (n_2 - n_1) f_z(x_\alpha) P_2((\lambda - it)x_\alpha) - n_2 f_z(b) P_2((\lambda - it)b) + \\ & + \frac{\lambda}{(\lambda - it)^3} [k_1 f_z(a) P_1((\lambda - it)a)] + \\ & + (k_2 - k_1) f_z(x_\alpha) P_1((\lambda - it)x_\alpha) - k_2 f_z(b) P_1((\lambda - it)b) + \\ & + \frac{\lambda}{(\lambda - it)^2} [c_1 f_z(a) + (c_2 - c_1) f_z(x_\alpha) - c_2 f_z(b)], \end{aligned}$$

где $f_z(x) = (\lambda - it) e^{-(\lambda - it)x}$ — комплекснозначная функция, $P_n(\dots)$ — многочлены вида (5).

Теорема 3. Начальные моменты произвольного порядка случайной величины X с плотностью вероятностей $g(x)$ типа (4) вычисляются по формуле

$$M(X^m) = \frac{1}{\lambda^{m+3}} [n_1 f(a) P_{m+2}(\lambda a) + (n_2 - n_1) f(x_a) P_{m+2}(\lambda x_a) - n_2 f(b) P_{m+2}(\lambda b)] + \\ + \frac{1}{\lambda^{m+2}} [k_1 f(a) P_{m+1}(\lambda a) + (k_2 - k_1) f(x_a) P_{m+1}(\lambda x_a) - k_2 f(b) P_{m+1}(\lambda b)] + \\ + \frac{1}{\lambda^{m+1}} [c_1 f(a) P_m(\lambda a) + (c_2 - c_1) f(x_a) P_m(\lambda x_a) - c_2 f(b) P_m(\lambda b)],$$

где $m=1,2,\dots$, $f(x) = \lambda e^{-\lambda x}$ — базовая функция плотности, $P_n(\dots)$ — многочлены вида (5).

Следствие. Формула для вычисления математического ожидания указанной случайной величины имеет вид

$$M(X) = \frac{1}{\lambda^4} [n_1 f(a) P_3(\lambda a) + (n_2 - n_1) f(x_a) P_3(\lambda x_a) - n_2 f(b) P_3(\lambda b)] + \\ + \frac{1}{\lambda^3} [k_1 f(a) P_2(\lambda a) + (k_2 - k_1) f(x_a) P_2(\lambda x_a) - k_2 f(b) P_2(\lambda b)] + \\ + \frac{1}{\lambda^2} [c_1 f(a) P_1(\lambda a) + (c_2 - c_1) f(x_a) P_1(\lambda x_a) - c_2 f(b) P_1(\lambda b)].$$

Заметим, что центральные моменты вычисляются по стандартным формулам связи с начальными моментами. В частности, $D(X) = M(X^2) - M^2(X)$.

Все приведенные утверждения доказаны аналитически.

Заключение

Представленный способ моделирования определяет широкий класс вероятностных распределений, в основу формирования которого положены следующие шаги:

- 1) выбор базового распределения (функция $f(x)$) с точностью до параметров;
- 2) усечение (выбор a и b);
- 3) «квантильное разбиение» на интервалы (выбор α и x_α);

4) выбор граничных значений плотности на полученных интервалах;

5) расчет коэффициентов в соотношении (1) с проверкой условия неотрицательности функции $g(x)$.

Варьирование не только параметрами группировки, но и квантилью порядка α позволяет конструировать различные нестандартные, в том числе мультимодальные, распределения с вычисляемыми характеристиками. Компьютерная реализация методики дает возможность строить распределения указанного типа при различных базовых функциях (в частности, плотности показательного распределения) и параметрах. Также можно осуществлять подгонку модельного теоретического распределения к реальным данным с проверкой соответствия по критерию согласия.

Результаты работы могут быть полезны при обработке статистических данных со сложной структурой распределения вероятностей. Подобного типа распределения зачастую встречаются в медико-социальных исследованиях, в частности в статистике заболеваемости населения в зависимости от возраста.

1. Токмачев М.С., Рязанцев П.П. Моделирование усеченных распределений // Вестник НовГУ. Сер.: Техн. науки. 2010. №55. С.34-36.
2. Токмачев М.С. Моделирование усеченных распределений с заданной квантилью // Вестник КГТУ им. А.Н.Туполева. 2011. № 2. С.157-162.
3. The On-Line Encyclopedia of Integer Sequences™ (OEIS™) — <http://www.research.att.com/~njas/sequences/> (дата обращения 15.09.2011)

Bibliography (Transliterated)

1. Tokmachev M.S., Rjazancev P.P. Modelirovanie usechennykh raspredelenij // Vestnik NovGU. Ser.: Tekhn. nauki. 2010. №55. S.34-36.
2. Tokmachev M.S. Modelirovanie usechennykh raspredelenij s zadannoj kvantil'ju // Vestnik KGTU im. A.N.Tupoleva. 2011. № 2. S.157-162.
3. The On-Line Encyclopedia of Integer Sequences™ (OEIS™) — <http://www.research.att.com/~njas/sequences/> (data obrashhenija 15.09.2011)