

Интернет-журнал «Наукоедение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 9, №3 (2017) <http://naukovedenie.ru/vol9-3.php>

URL статьи: <http://naukovedenie.ru/PDF/03TVN317.pdf>

Статья опубликована 12.05.2017

Ссылка для цитирования этой статьи:

Тимошенко С.И. Использование смешанного семейства распределений Пирсона и Джонсона для расчета вероятностных характеристик // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №3 (2017) <http://naukovedenie.ru/PDF/03TVN317.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

УДК 004.62

Тимошенко Сергей Иванович

ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Россия, Екатеринбург¹
Старший научный сотрудник
Доцент ИРИТ-РТФ
Кандидат технических наук
E-mail: s.i.timoshenko@urfu.ru

Использование смешанного семейства распределений Пирсона и Джонсона для расчета вероятностных характеристик

Аннотация. Основой для проведения исследования послужило отсутствие в библиотеках статистических функций с открытым кодом семейств распределений Пирсона и Джонсона. Как правило, они нужны при построении программного обеспечения для моделирования и анализа технических систем.

Несмотря на то, что распределения Пирсона и Джонсона известны давно, их практическое использование сталкивается с вычислительными трудностями. Для семейства распределений Пирсона – это отсутствие выражения через известные функции интеграла от IV типа распределения, что затрудняет поиск нормировочного коэффициента. Для семейства распределений Джонсона основной недостаток – отсутствие у распределения S_B аналитических выражений для моментов, пригодных для практических расчетов. Это затрудняет подгонку экспериментальных данных к этому виду распределения.

Проблема решена на основе предложенного смешанного семейства распределений Пирсона и Джонсона, в которое вошли распределения Пирсона (кроме IV и V типов) и распределение S_U Джонсона (их заменяющее).

Автором получены следующие результаты:

1. для аппроксимации распределений и расчета вероятностных характеристик предложено использовать смешанное семейство распределений Пирсона и Джонсона;
2. для распределения S_U Джонсона найдены выражения для связи параметров распределения через моменты; предложен эффективный алгоритм расчета параметров распределения;

¹ 620002, г. Екатеринбург, ул. Мира 32

3. получены соотношения для вероятностей превышения порога случайной величиной из смешанного семейства, выражающиеся через известные функции.

Ключевые слова: распределение Пирсона; распределение Джонсона; статистические функции; вероятность; открытый код

Введение

Семейства распределений Пирсона [1] и Джонсона [2] уже многие десятилетия используются в задачах аппроксимации эмпирических распределений, расчета и оценки вероятностных характеристик [3]. Не ослабевает интерес к ним и в наши дни. Спектр применений достаточно широк: в задачах анализа изображений [4, 5], в моделировании радиолокационных сигналов [6, 7], в оценке рисков [8, 9], в страховом деле [10], в регрессионном анализе [11], в оценке надежности [12], а также во многих других прикладных задачах теории вероятностей и математической статистики [13-17]. Есть работы и по модификации семейств распределений [18, 19]. Семейства распределений Пирсона и Джонсона сравнимы по своим аппроксимирующим возможностям.

Семейства распределений включены в статистические пакеты (см., например, Statistica²), в которых обычно решаются задачи подгонки экспериментальных данных к конкретным распределениям. При построении программного обеспечения для моделирования и анализа технических систем, как правило, требуются библиотеки статистических функций с открытым кодом. Однако полного набора функций для семейств распределений Пирсона и Джонсона в таких библиотеках нет (см., например, библиотеку Colt³ для Java или библиотеки языка R⁴). Поэтому актуальной задачей является создание базы для разработки таких библиотек.

При изучении алгоритмов расчета параметров распределений Пирсона и Джонсона [3, 20] можно отметить, что каждому из семейств присущи свои недостатки, затрудняющие их использование. Для семейства распределений Пирсона – это отсутствие выражения через известные функции интеграла от IV типа распределения [21], что затрудняет поиск нормировочного коэффициента.

Для семейства распределений Джонсона основной недостаток – отсутствие у распределения S_V аналитических выражений для моментов, пригодных для практических расчетов. Это затрудняет подгонку экспериментальных данных к этому виду распределения.

Перечисленные недостатки могут быть скомпенсированы, если вместо использования какого-либо одного семейства применять смешанное семейство распределений Пирсона и Джонсона [22]. В смешанное семейство входят распределения Пирсона (кроме типов IV и V) и распределение S_U Джонсона [23], заменяющее IV и V тип Пирсона (см. рис. 1). Распределение S_U Джонсона перекрывает V тип, который не включен в смешанное семейство,

² Анализ процессов. Негауссовские распределения [Электронный ресурс] // Statistica. Электронный учебник по статистике. Компания StatSoft Russia. - URL: <http://statsoft.ru/home/textbook/default.htm> (дата обращения: 21.02.2017).

³ CERN. Colt. Open Source Libraries for High Performance Scientific and Technical Computing in Java [Electronic resource] // Data Science and Technology. - URL: <https://dst.lbl.gov/ACSSoftware/colt/api/index.html> (circulation date: 22.02.2017).

⁴ An Introduction to R. Programming Environment for Data Analysis and Graphics [Electronic resource] // The Comprehensive R Archive Network. The R Manuals. - URL: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf> (circulation date: 16.03.2017).

и частично VI тип распределений Пирсона. Оси β_1 и β_2 на рис. 1 представляют из себя переменные Пирсона и определяются равенствами [21]:

$$\beta_1 = \mu_3^2 \mu_2^{-3} = \gamma_1^2, \quad \beta_2 = \mu_4 \mu_2^{-2} = \gamma_2 + 3, \quad (1)$$

где: γ_1 и γ_2 – соответственно коэффициенты асимметрии и эксцесса, μ_i – i -й центральный момент распределения вероятностей.

Прямая $\beta_2 = \beta_1 + 1$ отделяет критическую область. Для любого распределения должно выполняться неравенство $\beta_2 \geq \beta_1 + 1$ [21]. Таким образом, в критической области распределений не существует.

Распределения III типа Пирсона находятся на прямой $\beta_2 = (3\beta_1 + 6)/2$ [21]. Распределения V типа Пирсона – на кривой [20]:

$$\beta_2 = \frac{6(\beta_1 + 4)}{32 - \beta_1} [6 + \sqrt{\beta_1 + 4}] - 3. \quad (2)$$

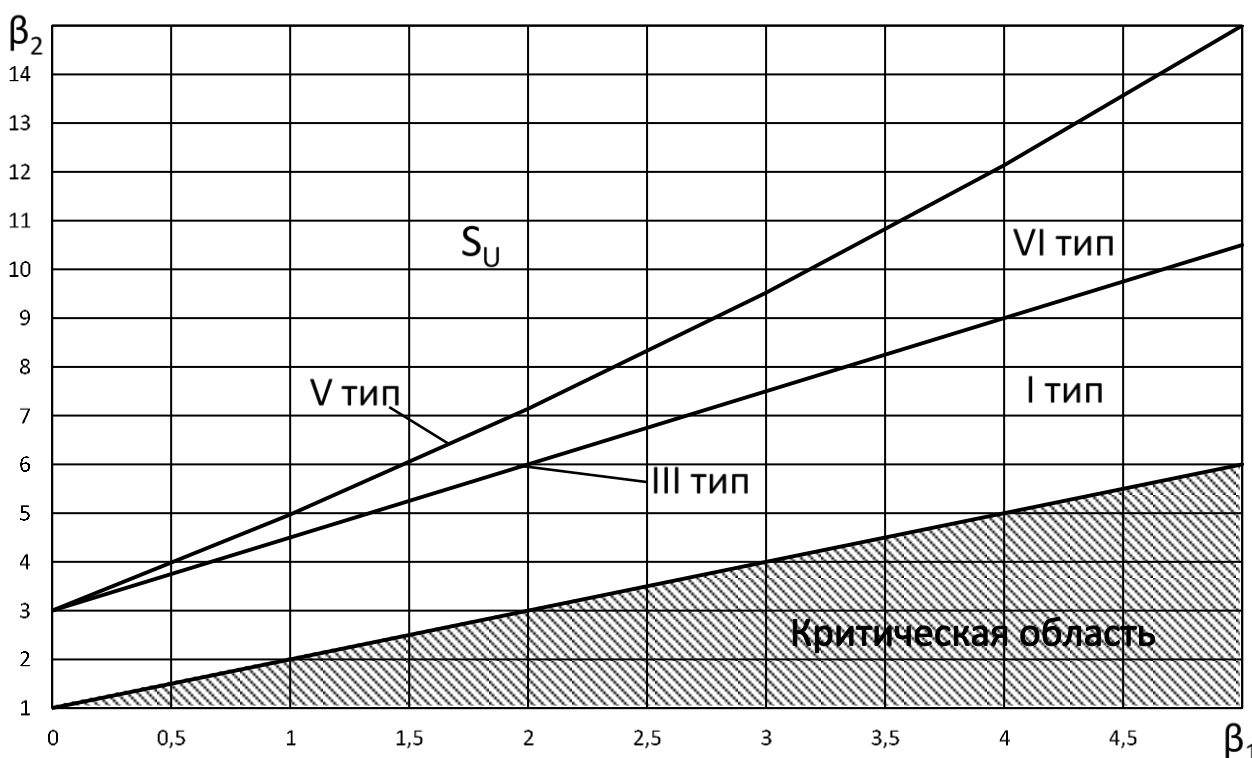


Рисунок 1. Диаграмма распределений на плоскости переменных Пирсона (линии построены автором по приведенным выше уравнениям)

Семейство распределений Пирсона

Семейство распределений Пирсона задается дифференциальным уравнением [21]:

$$\frac{\partial p(x)}{\partial x} = \frac{x - a}{b_0 + b_1 x + b_2 x^2}, \quad (3)$$

где: a и b_i – параметры распределения, $p(x)$ – плотность вероятностей.

Для центрированных распределений (математическое ожидание $m=0$) параметры имеют следующую связь с центральными моментами $\mu_i, i = \overline{2, 4}$:

$$a = b_1, \quad b_0 = \frac{c_0}{d}, \quad b_1 = \frac{c_1}{d}, \quad b_2 = \frac{c_2}{d}, \quad (4)$$

где:

$$\begin{aligned} c_0 &= -\mu_2(4\mu_2\mu_4 - 3\mu_3^2), & c_1 &= -\mu_3(\mu_4 + 3\mu_2^2), \\ c_2 &= -2\mu_2\mu_4 + 6\mu_2^3 + 3\mu_3^2, & d &= 10\mu_2\mu_4 - 18\mu_2^3 - 12\mu_3^2. \end{aligned} \quad (5)$$

В зависимости от корней $x_{1,2} = -\frac{b_1}{2b_2}(1 \pm \sqrt{1 - k^{-1}})$, $k = \frac{b_1^2}{4b_0b_2}$ уравнения

$$b_0 + b_1x + b_2x^2 = 0, \quad (6)$$

решение дифференциального уравнения (3) приводит к трем основным типам распределений:

1. тип I (бета-распределение первого рода), если корни уравнения (6) вещественны и имеют разные знаки ($k < 0$);
2. тип IV, если корни комплексные ($0 < k < 1$);
3. тип VI (бета-распределение II рода), если корни действительны и одинаковы по знаку ($k > 1$).

По существу, этим охватываются все возможные случаи. Тем не менее, граничные и частные случаи типов I, IV и VI выделяют особо. Поэтому в итоге различают 12 типов распределений Пирсона.

Для большинства практических расчетов можно ограничиться учетом в смешанном семействе лишь S_U распределения Джонсона и распределений Пирсона типа I, VI и III, которое разделяет первые два типа. Кратко остановимся на характеристике распределений, включенных в смешанное семейство.

Тип I (бета-распределение)

Для типа I Пирсона корни x_1 и x_2 действительны и различны по знаку, $k < 0$:

$$p(x) = \frac{(x - x_1)^g (x_2 - x)^h}{B(g + 1, h + 1)(x_2 - x_1)^{g+h+1}}, \quad x_1 \leq x \leq x_2, \quad (7)$$

где: $B()$ – бета-функция [24]. Распределение существует при $g > -1$, $h > -1$, которые определяются уравнениями:

$$g = \frac{b_1 - x_1}{b_2(x_2 - x_1)}, \quad h = \frac{x_2 - b_1}{b_2(x_2 - x_1)}. \quad (8)$$

При $g > 0$, $h > 0$ распределение унимодально. При $-1 < g < 0$, $-1 < h < 0$ – выпукло вниз, т.е. имеет U-образную форму. При g и h разных знаков – J-образно.

Тип III (гамма-распределение)

Тип III Пирсона получается при $k = \infty$ ($b_2 = 0$):

$$p(x) = \frac{(-(x + b_0b_1^{-1})b_1^{-1})^{-1 - b_0b_1^{-2}} \exp((x + b_0b_1^{-1})b_1^{-1})}{|b_1| \Gamma(-b_0b_1^{-2})}, \quad (9)$$

где: $x = b_0 b_1^{-1} \leq x < \infty$ при $b_1 < 0$ ($\mu_3 > 0$) и $-\infty < x \leq -b_0 b_1^{-1} = x_2$ при $b_1 > 0$ ($\mu_3 < 0$), $\Gamma()$ – гамма-функция [24]. Для распределения выполняются равенства:

$$b_2 = 0, \quad b_0 = -\mu_2, \quad a = b_1 = -\frac{\mu_3}{\mu_2}. \quad (10)$$

Распределения существуют как с положительной, так и отрицательной асимметрией (получаются путем зеркального отображения).

Тип VI (бета-распределение II рода)

Тип VI Пирсона соответствует ситуации, когда корни x_1 и x_2 действительны и одинаковы по знаку, т.е. $k > 1$. В зависимости от знака μ_3 различают два случая:

а) $\mu_3 > 0, x_2 < x < \infty$:

$$p(x) = \frac{(x - x_1)^g (x - x_2)^h}{(x_2 - x_1)^{g+h+1} B(h+1, -g-h-1)}, \quad (11)$$

б) $\mu_3 < 0, -\infty < x < x_1$:

$$p(x) = \frac{(x_1 - x)^g (x_2 - x)^h}{(x_2 - x_1)^{g+h+1} B(-g-h-1, g+1)}, \quad (12)$$

где: $B()$ – бета-функция. Распределение (11) – (12) существует лишь при $b_2 < -0,2$.

S_U распределение Джонсона

Плотность вероятностей $p(x)$ распределения Джонсона типа S_U имеет вид [23]:

$$p(x) = \frac{\eta}{\sqrt{2\pi}} \frac{1}{\sqrt{(x-\varepsilon)^2 + \lambda^2}} \exp \left[-\frac{1}{2} \left(\gamma + \eta \ln \left\{ \frac{x-\varepsilon}{\lambda} + \left[\left(\frac{x-\varepsilon}{\lambda} \right)^2 + 1 \right]^{\frac{1}{2}} \right\} \right)^2 \right], \quad (13)$$

где: x – случайная переменная, $-\infty < x < \infty$; $\eta, \lambda, \varepsilon, \gamma$ – параметры распределения, причем

$$\eta > 0, \lambda > 0, -\infty < \varepsilon < \infty, -\infty < \gamma < \infty.$$

Известна [23] связь первых четырех моментов с параметрами распределения:

$$m = \varepsilon - \lambda \omega sh \Omega, \quad (14)$$

$$\mu_2 = \frac{\lambda^2}{2} [\omega^4 ch 2\Omega - \omega^2 ch 2\Omega + \omega^2 - 1], \quad (15)$$

$$\mu_3 = -\frac{\lambda^3}{4} [(\omega^9 - 3\omega^5 + 2\omega^3) sh 3\Omega + 3(\omega^5 - 2\omega^3 + \omega) sh \Omega], \quad (16)$$

$$\mu_4 = \frac{\lambda^4}{8} \left[(\omega^{16} - 4\omega^{10} + 6\omega^6 - 3\omega^4) ch 4\Omega + 4(\omega^{10} - 3\omega^6 + 2\omega^4) ch 2\Omega + \right. \\ \left. + 3(2\omega^6 - 3\omega^4 + 1) \right], \quad (17)$$

где: m – математическое ожидание (первый начальный момент); μ_2, μ_3, μ_4 – соответственно второй, третий и четвертый центральный момент; $\omega = \exp(\frac{1}{2\eta^2})$, $\Omega = \gamma/\eta$; $sh()$, $ch()$ – соответственно гиперболический синус и косинус.

Параметры распределения (13) могут быть найдены в результате решения системы трансцендентных уравнений (14) – (17). К сожалению, не существует общих аналитических методов решения подобных систем. Так как не удается получить приближительной оценки параметров распределения (1), то затруднено применение численных методов.

Попробуем свести решение системы уравнений к решению одного уравнения. Допустим, что распределение (13) центрировано (т.е. $m=0$), тогда из (14) следует, что $\varepsilon = \lambda\omega sh\Omega$ и

$$sh\Omega = \frac{\varepsilon}{\lambda\omega}. \quad (18)$$

Подставим (18) в (15) и учтем, что [25]

$$ch2\Omega = 2sh^2\Omega + 1, \quad (19)$$

тогда

$$\mu_2 = (\omega^2 - 1) + (\omega^4 - 1)\frac{\lambda^2}{2}. \quad (20)$$

Учитывая, что по определению (13) $\lambda > 0$, из (20) следует

$$\lambda = \sqrt{2(\mu_2 - (\omega^2 - 1)\varepsilon^2)/(\omega^4 - 1)}. \quad (21)$$

Подставим (18), (21) в (16) и учтем, что [25]

$$sh3\Omega = 3sh\Omega + 4sh^3\Omega, \quad (22)$$

тогда

$$\mu_3 = \frac{1}{2}(\omega^2 - 1)^3\varepsilon^3 - \frac{3}{2}\mu_2(\omega^4 - 1)\varepsilon. \quad (23)$$

Преобразуем (23) к приведенному виду кубического уравнения [26]:

$$\varepsilon^3 - 3\mu_2\frac{(\omega^2+1)}{(\omega^2-1)^2}\varepsilon - \frac{2\mu_3}{(\omega^2-1)} = 0. \quad (24)$$

При $m = 0$ и допустимых значениях λ и ω знаки у ε и μ_3 должны быть противоположны (см. (16) и (18)). Это необходимо учесть при выборе решения уравнения (24). В зависимости от знака дискриминанта

$$\Delta = (\mu_3^2 - \mu_2^3(\omega^2 + 1)^3)/(\omega^2 - 1)^6, \quad (25)$$

уравнение (24) будет иметь либо один ($\Delta > 0$), либо три действительных корня ($\Delta < 0$).

Рассмотрим каждое из возможных решений. При $\Delta > 0$, воспользуясь формулой Кардано [26], получим

$$\varepsilon = \left(\sqrt[3]{\mu_3 + \sqrt{\mu_3^2 - \mu_2^3(\omega^2 + 1)^3}} + \sqrt[3]{\mu_3 - \sqrt{\mu_3^2 - \mu_2^3(\omega^2 + 1)^3}} \right) / (\omega^2 - 1). \quad (26)$$

В (26) знаки у ε и μ_3 совпадают. Следовательно, (26) не может быть решением системы уравнений (14) – (17). Воспользовавшись третьим способом [26] решения приведенных уравнений, при $\Delta < 0$ получим

$$\varepsilon = 2\text{sign}(\mu_3)\sqrt{\mu_2(\omega^2 + 1)}\cos(z) / (\omega^2 - 1). \quad (27)$$

Возможны три решения, отличающиеся значением z , которое может быть равным $\frac{\varphi}{3}$, $\frac{\varphi}{3} + \frac{2\pi}{3}$ и $\frac{\varphi}{3} + \frac{4\pi}{3}$, где:

$$\varphi = \arccos\left(\frac{|\mu_3|}{(\mu_2(\omega^2 + 1))^{\frac{3}{2}}}\right) = \arccos\left(\frac{\sqrt{\beta_1}}{(\omega^2 + 1)^{\frac{3}{2}}}\right). \quad (28)$$

Выражение, стоящее под знаком функции в (28), положительно, поэтому $0 \leq \varphi \leq \frac{\pi}{2}$. С учетом этого, области допустимых значений для z , определяющих три возможных решения, графически изображены на рис. 2 (кривые 1, 2, 3 соответствуют $z = \frac{\varphi}{3}$, $z = \frac{\varphi}{3} + \frac{2\pi}{3}$ и $z = \frac{\varphi}{3} + \frac{4\pi}{3}$).

Из анализа (27) и рис. 2 следует, что первое решение ($z = \frac{\varphi}{3}$) не удовлетворяет системе уравнений (14) – (17), так как в этом случае знаки у ε и μ_3 совпадают.

Из (16) следует также, что $sh\Omega = 0$ и $sh3\Omega = 0$, если $\mu_3 = 0$. Тогда для $m=0$ следует, что и $\varepsilon = 0$. Этой проверке удовлетворяет лишь третье решение ($z = \frac{\varphi}{3} + \frac{4\pi}{3}$), которое и определяет истинное решение (27), удовлетворяющее системе уравнений (14) – (17). Таким образом

$$\varepsilon = 2sign(\mu_3)\sqrt{\mu_2(\omega^2 + 1)} \cos\left(\frac{\varphi}{3} + \frac{4\pi}{3}\right) / (\omega^2 - 1). \quad (29)$$

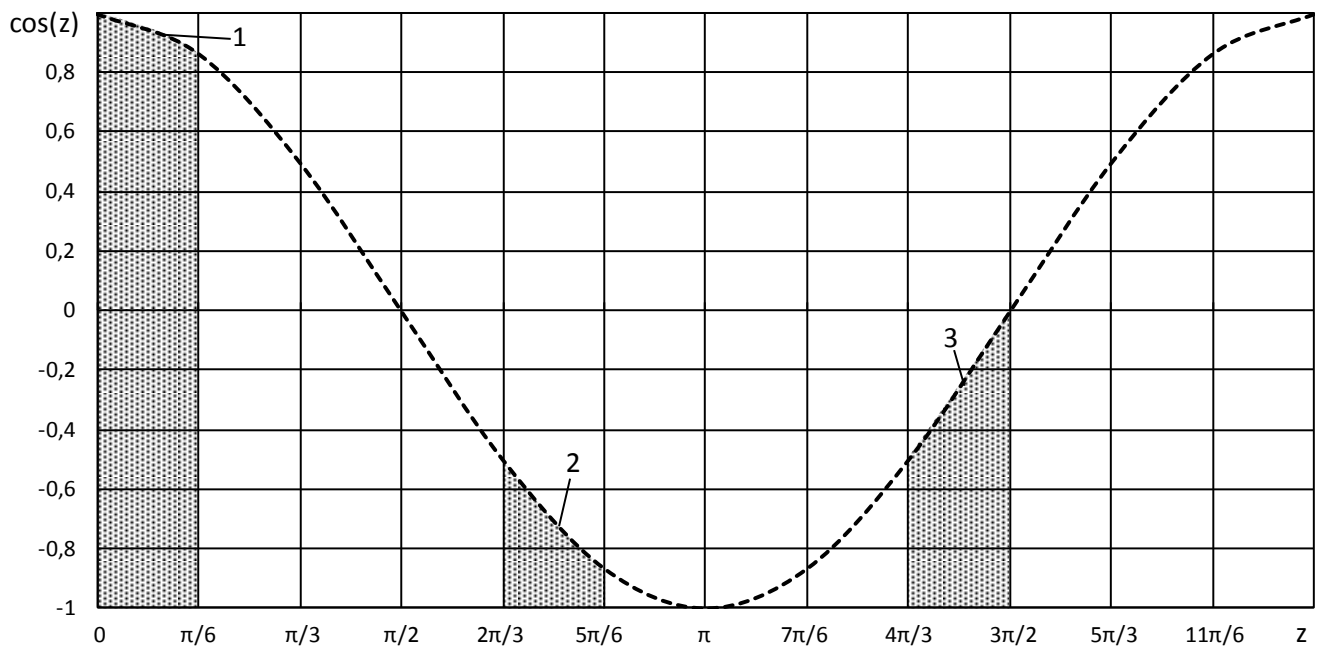


Рисунок 2. К выбору корня кубического уравнения (разработано автором для иллюстрации поиска решений уравнения (27))

Область определения для ω^2 может быть найдена из условия $\Delta < 0$. Тогда из (25) следует:

$$\omega^2 > \sqrt[3]{\mu_3^2 \mu_2^{-3}} - 1 = \sqrt[3]{\beta_1} - 1. \quad (30)$$

Кроме того, из обозначения $\omega = \exp\left(\frac{1}{2\eta^2}\right)$, которое было введено для системы уравнений (14) – (17), следует, что $\omega > 1$ и, лишь при $\eta \rightarrow \infty$, $\omega \rightarrow 1$.

Подставим (18), (21) в (17), учтем (19), и что

$$ch4\Omega = 8sh^4\Omega + 8sh^2\Omega + 1, \quad (31)$$

тогда

$$\mu_4 = \frac{1}{2}(-\omega^{12} + 2\omega^{10} - \omega^8 + 4\omega^6 - 11\omega^4 + 10\omega^2 - 3)\varepsilon^4 + \quad (32)$$

$$+(\omega^{10} + \omega^8 - 4\omega^4 - \omega^2 + 3)\mu_2\varepsilon^2 + \frac{1}{2}(\omega^8 + 2\omega^4 + 3)\mu_2^2.$$

Пусть

$$\varepsilon_{ц} = 2\text{sign}(\mu_3)\sqrt{\mu_2(\omega^2 + 1)} \cos\left(\frac{\varphi}{3} + \frac{4\pi}{3}\right), \quad (33)$$

тогда из (29) следует, что

$$\varepsilon = \varepsilon_{ц}/(\omega^2 - 1). \quad (34)$$

Подставим (34) в (32). Перенеся все члены (32) в левую часть уравнения и разделив многочлены при ε^4 и ε^2 на соответствующие степени от многочлена $(\omega^2 - 1)$, получим:

$$\mu_4 + \frac{1}{2}(\omega^4 + 2\omega^2 + 3)\varepsilon_{ц}^4 - (\omega^6 + 3\omega^4 + 5\omega^2 + 3)\varepsilon_{ц}^2 - \frac{1}{2}(\omega^8 + 2\omega^4 + 3)\mu_2^2 = 0. \quad (35)$$

Порядок уравнения может быть понижен вдвое, так как в нем присутствуют только четные степени переменной ω . В общем случае аналитически разрешить уравнение (35) относительно переменной ω не удастся. Однако численное его решение при введенных ограничениях на ω может быть проведено более эффективно, по сравнению с численным решением исходной системы уравнений (14) – (17).

Повысить эффективность численного решения можно с использованием производной от функции $f(\vartheta)$, расположенной в левой части уравнения (35), где: $\vartheta = \omega^2$:

$$f(\vartheta) = \mu_4 + \frac{1}{2}(\vartheta^2 + 2\vartheta + 3)\varepsilon_{ц}^4 - (\vartheta^3 + 3\vartheta^2 + 5\vartheta + 3)\varepsilon_{ц}^2 - \frac{1}{2}(\vartheta^4 + 2\vartheta^2 + 3)\mu_2^2. \quad (36)$$

Найдем производную $f(\vartheta)$ от функции (36) по ϑ . Для этого разобьем функцию $f(\vartheta)$ на три части: $f(\vartheta) = f_1(\vartheta) + f_2(\vartheta) + f_3(\vartheta)$, где:

$$f_1(\vartheta) = \mu_4 + \frac{1}{2}(\vartheta^2 + 2\vartheta + 3)\varepsilon_{ц}^4 = \quad (37)$$

$$= \mu_4 + 8(\vartheta^4 + 4\vartheta^3 + 8\vartheta^2 + 8\vartheta + 3) \cos^4\left(\frac{\arccos(\sqrt{\beta_1}(\vartheta+1)^{-\frac{3}{2}})}{3} + \frac{4\pi}{3}\right);$$

$$f_2(\vartheta) = -4\mu_2(\vartheta^4 + 4\vartheta^3 + 8\vartheta^2 + 8\vartheta + 3) \cos^2\left(\frac{\arccos(\sqrt{\beta_1}(\vartheta+1)^{-\frac{3}{2}})}{3} + \frac{4\pi}{3}\right); \quad (38)$$

$$f_3(\vartheta) = -\mu_2^2(\vartheta^4 + 2\vartheta^2 + 3)/2. \quad (39)$$

Взяв производную от каждой из частей (37) – (39) функции $f(\vartheta)$, получим

$$f'(\vartheta) = 16\mu_2(\vartheta^3 + 3\vartheta^2 + 4\vartheta + 2) \cos^2\left(\frac{\arccos(\sqrt{\beta_1}(\vartheta + 1)^{-\frac{3}{2}})}{3} + \frac{4\pi}{3}\right) \times \quad (40)$$

$$\times \left(2\mu_2 \cos^2\left(\frac{\arccos(\sqrt{\beta_1}(\vartheta + 1)^{-\frac{3}{2}})}{3} + \frac{4\pi}{3}\right) - 1\right) + 2\mu_2\sqrt{\beta_1}(\vartheta + 1)^{-\frac{3}{2}} \times$$

$$\begin{aligned} & \times (\vartheta^3 + 3\vartheta^2 + 5\vartheta + 3) \sin \left(2 \left(\frac{\arccos \left(\sqrt{\beta_1} (\vartheta + 1)^{-\frac{3}{2}} \right)}{3} + \frac{4\pi}{3} \right) \right) (1 - \beta_1 (\vartheta + 1)^{-3})^{-\frac{1}{2}} \times \\ & \times \left(1 - 4\mu_2 \cos^2 \left(\frac{\arccos \left(\sqrt{\beta_1} (\vartheta + 1)^{-\frac{3}{2}} \right)}{3} + \frac{4\pi}{3} \right) \right) - 2\mu_2^2 (\vartheta^3 + \vartheta). \end{aligned}$$

После нахождения из (35) ω , параметры распределения могут быть последовательно получены из уравнений (29) – ε и (21) – λ . Так как $\omega = \exp\left(\frac{1}{2\eta^2}\right)$, то $\eta = (2\ln\omega)^{\frac{1}{2}}$. Учитывая ранее введенное обозначение $\Omega = \gamma/\eta$ и формулу (18), а также найденные параметры $\eta, \lambda, \varepsilon$, можно найти γ :

$$\gamma = \eta \operatorname{arsh} \left(\frac{\varepsilon}{\lambda\omega} \right) = \eta \ln \left(\frac{\varepsilon}{\lambda\omega} + \sqrt{\left(\frac{\varepsilon}{\lambda\omega} \right)^2 + 1} \right). \quad (41)$$

Так как параметры найдены для математического ожидания $m = 0$, то в общем случае $m \neq 0$ в конце расчетов придется сделать поправку только для параметра ε с учетом формулы (14), т.е. к найденному по формуле (29) значению ε прибавить m .

Методика аппроксимации распределений смешанным семейством

Методика аппроксимации распределений смешанным семейством Пирсона и Джонсона заключается в следующем:

1. вычисляются четыре момента распределения: m, μ_2, μ_3, μ_4 ;
2. вычисляется значение k ;
3. по величине k определяется тип распределения;
4. по моментам рассчитываются параметры распределения.

Вычисление вероятностей превышения порога

В ряде прикладных задач требуется определить вероятность превышения порога. К таким задачам относятся задачи обнаружения сигналов. Обратной к задаче обнаружения сигналов является поиск порога для обеспечения заданной вероятности ложной тревоги. Близкой к последней задаче является определение доверительного интервала для оценки случайного параметра (см. рис. 3, где: $p(x)$ – плотность вероятностей, P – доверительная вероятность, T_1 и T_2 – границы доверительного интервала). Во всех этих задачах основой являются алгоритмы вычисления вероятностей превышения порога.

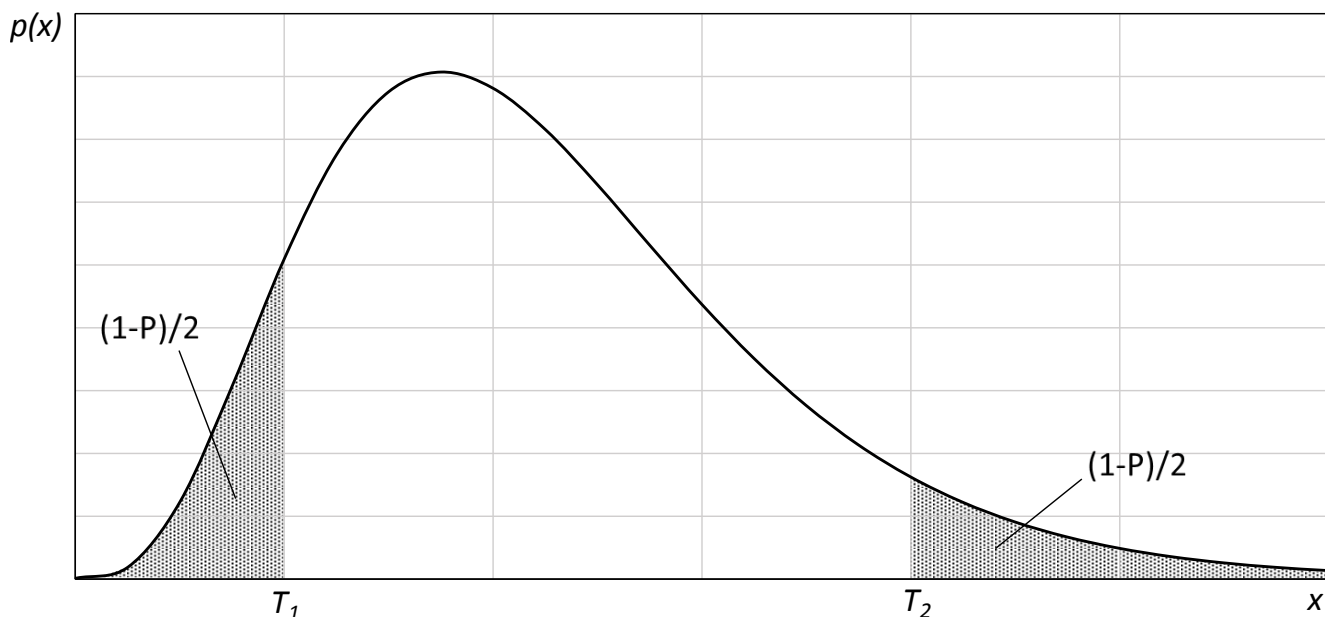


Рисунок 3. Доверительный интервал для заданной доверительной вероятности P (график построен автором для показательного-степенной плотности вероятностей с параметром 3,5)

Определим аналитические выражения для вероятностей $P(x > T)$ превышения порога T случайной величиной x с распределением из смешанного семейства. Необходимо учесть, что найденные ранее параметры определяют центрированные распределения. Кроме того, конечные формулы для вероятностей должны выражаться через известные функции.

Общее соотношение для вероятности $P(x > T)$, выраженное через плотность вероятностей $p_u(x)$ центрированного распределения, имеет вид:

$$P(x > T) = \int_{T-m}^{\infty} p_u(x) dx. \quad (42)$$

Тип I (бета-распределение)

Подставим в (42) плотность вероятностей (7). В этом случае, исходя из области определения для x , верхний предел в (42) будет равен x_2 . Осуществим замену переменной: $z = \frac{(x-x_1)}{(x_2-x_1)}$, тогда $T_N = (T - m - x_1)/(x_2 - x_1)$, $dx = (x_2 - x_1)dz$. Окончательное выражение:

$$P(x > T) = \int_{T_N}^1 B^{-1}(g + 1, h + 1) z^g (1 - z)^h dz = 1 - I_{T_N}(g + 1, h + 1), \quad (43)$$

где: $I_{T_N}()$ – отношение неполной бета-функции к полной [27] или бета-функция распределения вероятностей.

Тип III (гамма-распределение)

Подставим в (42) плотность вероятностей (9). Сделаем замену переменной: $z = -xb_1^{-1} - b_0b_1^{-2}$, тогда $T_N = -(T - m)b_1^{-1} - b_0b_1^{-2}$, $dx = -b_1 dz$. Выражение для $P(x > T)$ будет зависеть от знака μ_3 . Для $\mu_3 > 0$ верхний предел в (42) равен бесконечности, причем $b_1 = -|b_1|$, поэтому

$$P(x > T) = \int_{T_N}^{\infty} \Gamma^{-1}(-b_0 b_1^{-2}) z^{-b_0 b_1^{-2}-1} e^{-z} \partial z = 1 - I_{T_N}(-b_0 b_1^{-2}). \quad (44)$$

Для $\mu_3 < 0$ верхний предел в (42) равен $x_2 = -b_0 b_1^{-1}$, причем $b_1 = |b_1|$, поэтому

$$P(x > T) = \int_0^{T_N} \Gamma^{-1}(-b_0 b_1^{-2}) z^{-b_0 b_1^{-2}-1} e^{-z} \partial z = I_{T_N}(-b_0 b_1^{-2}). \quad (45)$$

В (44) – (45) $I_{T_N}()$ – гамма-функция распределения вероятностей.

Тип VI (бета-распределение II рода)

Вид плотности вероятностей в этом случае, как и выражение для $P(x > T)$, зависит от знака μ_3 .

При $\mu_3 > 0$ подставим в (42) плотность вероятностей (11). При этом верхний предел в (42) будет равен бесконечности. Осуществим замену переменной: $z = (x_2 - x_1)/(x - x_1)$, тогда $T_N = (x_2 - x_1)/(T - m - x_1)$, $\partial x = -(x_2 - x_1)z^{-2} \partial z$, и получим

$$P(x > T) = \int_0^{T_N} B^{-1}(-g - h - 1, h + 1) z^{-g-h-2} (1 - z)^h \partial z = I_{T_N}(-g - h - 1, h + 1). \quad (46)$$

При $\mu_3 < 0$ подставим в (42) плотность вероятностей (12). Верхний предел в этом случае будет равен x_1 . Проведем замену переменной: $z = (x_2 - x_1)/(x_2 - x)$, тогда

$$T_N = (x_2 - x_1)/(x_2 - T + m), \partial x = (x_2 - x_1)z^{-2} \partial z, \text{ и получим}$$

$$P(x > T) = \int_{T_N}^1 B^{-1}(-g - h - 1, g + 1) z^{-g-h-2} (1 - z)^g \partial z = 1 - I_{T_N}(-g - h - 1, g + 1). \quad (47)$$

В (46) – (47) $I_{T_N}()$ – бета-функция распределения вероятностей.

S_U распределение

Подставим в (42) плотность вероятностей (13). Осуществим замену переменной: $z = \gamma + \eta arsh\left(\frac{x-\varepsilon}{\lambda}\right)$, тогда $T_N = \gamma + \eta arsh\left(\frac{T-m-\varepsilon}{\lambda}\right)$, $x = \lambda sh\left(\frac{z-\gamma}{\eta}\right) + \varepsilon$, $\partial x = \frac{\lambda}{\eta} ch\left(\frac{z-\gamma}{\eta}\right) \partial z$ и получим

$$P(x > T) = \frac{1}{\sqrt{2\pi}} \int_{T_N}^{\infty} e^{-\frac{1}{2}z^2} \partial z = 1 - \Phi(T_N), \quad (48)$$

где: $\Phi()$ – интеграл вероятности [21].

Заключение

Для аппроксимации распределений и расчета вероятностных характеристик предложено использовать смешанное семейство распределений Пирсона и Джонсона, в которое вошли распределения Пирсона (кроме IV и V типов) и распределение S_U Джонсона. Смешанное семейство распределений эквивалентно по аппроксимирующим возможностям семействам Пирсона или Джонсона, однако позволяет избежать вычислительных трудностей определения параметров распределений IV типа Пирсона и S_B Джонсона. Смешанное семейство опробовано на различных статистических данных, в частности на данных нагрузочного тестирования интернет-сервисов [28].

Для распределения S_U Джонсона найдены выражения для связи параметров распределения через моменты. Предложен эффективный алгоритм расчета параметров распределения.

Получены соотношения для вероятностей превышения порога случайной величиной из смешанного семейства, выражающиеся через известные функции (бета-функцию, гамма-функцию и интеграл вероятности).

Найденные выражения для смешанного семейства распределений Пирсона и Джонсона могут стать базой для разработки библиотек статистических функций с открытым кодом.

ЛИТЕРАТУРА

1. Pearson, K. Contributions to the Mathematical Theory of Evolution. II. Skew Variations in Homogeneous Material [Text] / K. Pearson // Philosophical Transactions of the Royal Society of London. Ser. A. – 1895. – Vol. 186. – P. 343-414.
2. Johnson, N.L. Systems of Frequency Curves Generated by Methods of Translation [Text] / N.L. Johnson // Biometrika. – 1949. – Vol. 36, № 1/2. – P. 149-176.
3. Джонсон, Н.Л. Одномерные непрерывные распределения [Текст]: в 2 ч. Ч. 1 / Н.Л. Джонсон, С. Коц, Н. Балакришнан. – М.: БИНОМ. Лаборатория знаний, 2010. – 703 с.
4. Буркатовская, Ю.Б. Применение распределений Джонсона к задаче классификации аэрокосмических изображений [Текст] / Ю.Б. Буркатовская, Н.Г. Марков, А.С. Морозов, А.П. Серых // Известия Томского политехнического университета. – 2007. – Т. 311, № 5. – С. 76-80.
5. Воронов, И.В. Применение универсального семейства распределений Пирсона для аппроксимации распределения значений вектора псевдоградиента при совмещении изображений [Текст] / И.В. Воронов // Радиоэлектронная техника. – 2015. – № 2(8). – С. 123-127.
6. Пат. 2187888 Российская Федерация, МПК⁷ G 01 S 7/40. Способ имитации радиолокационных сигналов радиолокационных систем навигации летательных аппаратов [Текст] / Кольцов Ю.В., Белинский А.В.; заявитель и патентообладатель Государственная корпорация по атомной энергии “Росатом” и ФГУП ФНПЦ “Научно-исследовательский институт измерительных систем им. Ю.Е. Седакова”. – № 2015108240/07; заявл. 10.03.2015; опубл. 10.06.2016, Бюл. № 16. – 17 с.
7. Карпов, И.Г. Модернизация распределений Пирсона для аппроксимации экспериментальных распределений радиолокационных сигналов [Текст] / И.Г. Карпов, Е.А. Галкин // Известия вузов. Радиоэлектроника. – 2004. – Т. 47, № 11. – С. 52-61.
8. Лебедев, С.А. Оценка рыночных рисков ценных бумаг на основе универсальных семейств распределений [Текст] / С.А. Лебедев // Сибирская финансовая школа. – 2008. – № 2(67). – С. 87-90.
9. Naguez, N. Dynamic Portfolio Insurance Strategies: Risk Management under Johnson Distributions [Text] / N. Naguez // Annals of Operations Research. – 06 February 2016. – P. 1-25. – DOI 10.1007/s10479-016-2121-8.

10. Смагина, С.Р. Расчет тарифных ставок с использованием кривых Джонсона при страховании автотранспорта [Текст] / С.Р. Смагина, Г.М. Кошкин // Известия высших учебных заведений. Физика. – 2015. – Т. 58, № 11-2. – С. 297-302.
11. Тимофеев В.С. Оценивание параметров регрессионных зависимостей с использованием кривых Пирсона [Текст] / В.С. Тимофеев // Науч. вестн. Новосибир. гос. техн. Ун-та. – 2009. – № 4 (37). – С. 57-66.
12. Игнатенко, В.И. Применение распределений Джонсона и Пирсона в задачах технической эксплуатации автомобилей [Текст] / В.И. Игнатенко // Сборник научных трудов конференции «Проблемы современной науки». – Липецк: Издательство ЛГТУ, 2016. – С. 63-66.
13. Битюкова, В.В. Применение универсальных семейств распределений Пирсона для моделирования загруженности кабинетов лечебно-профилактических учреждений [Текст] / В.В. Битюкова, А.А. Хвостов, Д.И. Ребриков // Вестник Тамбовского государственного технического университета. – 2008. – Т. 14, № 1. – С. 202-208.
14. Свиркина, Н.М. Метод кривых Пирсона в анализе баллов ЕГЭ абитуриентов, поступивших в вуз [Текст] / Н.М. Свиркина // Процессы управления и устойчивость. – 2014. – Т. 1, № 1. – С. 500-505.
15. Mateus, A. Fitting Johnson's S_B Distribution to Forest Tree Diameter [Text] / A. Mateus, M. Tome // Advances in Regression, Survival Analysis, Extreme Values, Markov Processes and Other Statistical Applications. Studies in Theoretical and Applied Statistics. – Berlin: Springer-Verlag, 2013. – P. 289-296. – DOI: 10.1007/978-3-642-34904-1_30.
16. Заикин, П.В. Аппроксимация эмпирических функций полиномами высших порядков [Текст] / П.В. Заикин, М.А. Погореловский, В.С. Микшина // Вестник кибернетики. – 2015. – № 4. – С. 129-134.
17. Губарев, В.В. Идентификация эмпирических распределений [Текст] / В.В. Губарев // Вестник Самарского государственного аэрокосмического университета. – 2011. – № 6(30). – С. 205-215.
18. Карпов, И.Г. Модифицированные распределения Джонсона и их применение для аппроксимации законов распределения экспериментальных данных [Текст] / И.Г. Карпов, Ю.Т. Зырянов, А.Н. Грибков // Известия Томского политехнического университета. – 2013. – Т. 322, № 2. – С. 46-50.
19. Карпов, И.Г. Методика идентификации параметров модифицированных распределений Пирсона с использованием степенных, экспоненциальных и логарифмических моментов [Текст] / И.Г. Карпов, Ю.Т. Зырянов // Фундаментальные и прикладные проблемы техники и технологии. – 2015. – № 3 (311). – С. 117-127.
20. Бостанджиян, Б.А. Распределение Пирсона, Джонсона, Вейбулла и обратное нормальное. Оценивание их параметров [Текст] / Б.А. Бостанджиян. – Черноголовка: Редакционно-издательский отдел ИПХФ РАН, 2009. – 240 с.
21. Тихонов, В.И. Статистическая радиотехника [Текст] / В.И. Тихонов. – 2-е изд., перераб. и доп. – М.: Радио и связь, 1982. – 624 с.
22. Тимошенко, С.И. Использование семейств кривых Джонсона и Пирсона в задачах аппроксимации распределений, расчета и оценки вероятностных

- характеристик [Текст] / С.И. Тимошенко // УПИ. – Свердловск, 1986. – 59 с. – Библиогр.: С. 47-48. – Деп. в ВИНТИ 30.07.86, № 6190-В86.
23. Тимошенко, С.И. Расчет параметров распределения Джонсона типа S_U по моментам [Текст] / С.И. Тимошенко // Научные труды международной научно-практической конференции «СВЯЗЬ-ПРОМ 2009». – Екатеринбург: УрТИСИ ГОУ ВПО «СибГУТИ», 2009. – С. 26-29.
 24. Градштейн, И.С. Таблицы интегралов, сумм, рядов и произведений [Текст] / И.С. Градштейн, И.М. Рыжик. – М.: Наука, 1971. – 1108 с.
 25. Двайт, Г.Б. Таблицы интегралов и другие математические формулы [Текст] / Г.Б. Двайт. – М.: Наука, 1978. – 224 с.
 26. Бронштейн, И.Н. Справочник по математике для инженеров и учащихся втузов [Текст] / И.Н. Бронштейн, К.А. Семендяев. – Изд. перераб. под ред. Г. Гроше и В. Циглера; пер. с нем. – М.: Наука, 1980. – 976 с.
 27. Корн, Г. Справочник по математике для научных работников и инженеров [Текст] / Г. Корн, Т. Корн. – М.: Наука, 1977. – 831 с.
 28. Суханов, В.И. Исследование визуализации данных с применением открытых геоинформационных сервисов [Текст] / В.И. Суханов, С.И. Тимошенко, Р.М. Чернин // Информационные системы и технологии. – 2011. – № 6(68). – С. 139-144.

Timoshenko Sergey Ivanovich

Ural federal university, Russia, Ekaterinburg

E-mail: s.i.timoshenko@urfu.ru

Using a mixed family of Pearson and Johnson distributions to calculate the probability characteristics

Abstract. The basis for the study was the absence in libraries of statistical functions with open code families of distributions of Pearson and Johnson. Typically, they are needed when building software for modeling and analysis of technical systems.

Despite the fact that the distributions of Pearson and Johnson are known for a long time, their practical use is faced with computational difficulties. For the family of Pearson distributions, this is the absence of an expression in terms of known functions of the integral of the IV type of distribution, which makes it difficult to find the normalization coefficient. For the family of Johnson distributions the main drawback is the lack of analytic expressions for the distribution of S_B for moments suitable for practical calculations. This makes it difficult to fit the experimental data to this type of distribution.

The problem was solved on the basis of the proposed mixed family of Pearson and Johnson distributions, which included the Pearson distributions (except for IV and V types) and the Johnson S_U distribution (their substitute).

The author obtained the following results:

1. for the approximation of distributions and the calculation of probabilistic characteristics, it is proposed to use a mixed family of Pearson and Johnson distributions;
2. for the Johnson S_U distribution, expressions are found for the relationship of the distribution parameters through moments; an efficient algorithm for calculating the distribution parameters;
3. relations are obtained for the probability of exceeding the threshold by a random variable from a mixed family, expressed in terms of known functions.

Keywords: Pearson distribution; Johnson distribution; statistical functions; probability; open code