ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И АВТОМАТИЗАЦИЯ

УДК 004.942

С.Б. Приходько, Л.Н. Макарова

АНАЛИТИЧЕСКАЯ ЗАВИСИМОСТЬ ДЛЯ ВЫБОРА РАСПРЕДЕЛЕНИЯ ДЖОНСОНА СЕМЕЙСТВА S_L

Отримано нелінійну залежність ексцесу від асиметрії А для діапазону $A^2 \in [0; 27]$ для вибору розподілу Джонсона сімейства S_L.

Постановка проблемы. В настоящее время для подбора аналитической модели закона распределения экспериментальных данных широкое применение получили семейства распределений Джонсона [1, 2, 3]. Для автоматизации функциональных задач управления и анализа автоматизированных систем обработки информации при выборе конкретного семейства распределений Джонсона требуются соответствующие аналитические зависимости.

Анализ последних исследований и публикаций. На сегодняшний день для выбора конкретного семейства распределения Джонсона используют либо известную диаграмму в плоскости эксцесс – асимметрия в квадрате ε-A^2 [4, 5, 6], либо аналитическую зависимость $\varepsilon(A^2)$ [7].

В плоскости ε-A^2 присутствуют две линии, которые и определяют конкретное семейство распределений Джонсона (см. рис 1). Общепринятым считается обозначение координатных осей этой плоскости в виде (β_1, β_2) или (A^2, ε). Будем придерживаться обозначения координатных осей (A^2, ε).

Если точка с координатами (A^2, ε) находится около линии S_L, то можно использовать распределение из семейства S_L. Если точка с координатами (A^2, ε) находится выше линии S_L, то можно использовать распределение из семейства S_U, а если ниже линии S_L до линии критической области – то распределение из семейства S_B. Если же точка попадает в критическую область, то использовать для аппроксимации семейства распределений Джонсона нельзя.

Диаграммы, приведенные в [4, 5, 6], приводятся для диапазона $A^2 \in [0; 4]$, а существующая линейная аппроксимация зависимости $\varepsilon(A^2)$ для линии S_L, которая приведена в [7], дает неплохие результаты в диапазоне $A^2 \in [0; 5]$:

$$\varepsilon = 3 \cdot (1 + 0{,}641\mathrm{A}^2)\,. \tag{1}$$

Однако в ряде случаев встречаются такие выборки эмпирических данных, для которых оценки асимметрии в квадрате A^2 значительно превышают значение 5. Так, для данных времени наработки между отказами устройств терминальной сети встречаются выборки с оценкой A^2=24,8, что требует нахождения зависимости для соответствующего диапазона A^2.

Приведенная линейная зависимость (1) при больших значениях A^2 дает существенные ошибки, которые влияют на выбор семейства распределения Джонсона. Поэтому для автоматизации выбора конкретного семейства распределений Джонсона при обработке большого количества эмпирических данных возникает необходимость в построении аналитической зависимости $\varepsilon(A^2)$ для диапазона $A^2 \in [0; 27]$.

Целью данной статьи является нахождение аналитической зависимости эксцесса от асимметрии в квадрате в диапазоне значений $A^2 \in [0; 27]$ для автоматизации выбора конкретного распределения не только из семейства S_L, но и из семейств S_B и S_U распределений Джонсона.

Изложение основного материала. Семейства распределений Джонсона получены путем преобразования нормированной нормально распределенной случайной величины (СВ) с математическим ожиданием M=0 и дисперсией D=1. В общем случае преобразование имеет вид [8]:

$$z = \gamma + \eta q(x, \varphi, \lambda); \eta > 0; -\infty < \gamma < \infty; \lambda > 0; -\infty < \varphi < \infty,$$

где q – нелинейная функция; γ, η, λ, φ – параметры распределения, причем γ и η – параметры формы, λ – параметр масштаба, φ – параметр смещения; x – СВ, которая нормализуется; z – нормированная, нормально распределенная СВ.

Джонсон предложил три семейства функций q:

$$q_1(x,\varphi,\lambda) = \ln\left(\frac{x-\varphi}{\lambda}\right), x > \varphi;$$

$$q_2(x,\varphi,\lambda) = \ln\left(\frac{x-\varphi}{\lambda+\varphi-x}\right), \varphi < x < \varphi+\lambda;$$

$$q_3(x,\varphi,\lambda) = \mathrm{Arsh}\left(\frac{x-\varphi}{\lambda}\right), -\infty \le x \le +\infty.$$

Семейству функций q_1 соответствует логарифмически нормальное распределение S_L Джонсона, семейству функций q_2 соответствует семейство распределений S_B Джонсона, семейству функций q_3 соответствует семейство распределений S_U Джонсона.

При использовании аналитических зависимостей верхняя граница критической области, является прямой линией и может быть задана следующим уравнением [7]:

$$\varepsilon = A^2 + 1.$$

Линия S_L является нелинейной и зависит от количества значений в выборке. Построим зависимость $\varepsilon(A^2)$ для линии S_L, используя следующую систему моментных уравнения [8]:

$$\begin{cases} \alpha_1 = \omega\rho; \\ \mu_2 = \omega^2\rho^2(\omega^2-1); \\ \mu_3 = \omega^3\rho^3(\omega^2-1)^2(\omega^2+2); \\ \mu_4 = \omega^4\rho^4(\omega^2-1)^2(\omega^8+2\omega^6+3\omega^4-3), \end{cases}$$

где $\omega = \exp(\frac{1}{2}\eta^2), \rho = \exp(-\frac{\gamma}{\eta})$.

Учитывая, что

$$A^2 = \frac{\mu_3{}^2}{\mu_2{}^3}, \varepsilon = \frac{\mu_4}{\mu_2{}^2},$$

получаем параметрические уравнения, которые определяют зависимость $\varepsilon(A^2)$ и имеют вид [8]:

$$\begin{aligned} A^2 &= \omega^6 + 3\omega^4 - 4; \\ \varepsilon &= \omega^8 + 2\omega^6 + 3\omega^4 - 3. \end{aligned} \qquad (2)$$

Пары значений (A^2, ε), полученные из решения параметрических уравнений (2) с помощью задания ω, приведены в табл. 1. На основе метода наименьших квадратов по значениям, приведенным в табл. 1, получаем аналитическую зависимость $\varepsilon(A^2)$:

$$e(A^2) = 7{,}2315 \quad 10 \quad A^8 \quad 6{,}9860 \quad 10 \quad A^6 + 4{,}5460 \quad 10 \quad A^4 + 1{,}7979A^2 + 2{,}9891. \qquad (3)$$

Таблица 1

Расчетные значения для построения зависимости ε(A^2)

ω	A^2	ε	ω	A^2	ε
1,0000	0,0000	3,0000	1,3000	9,3951	23,3792
1,0250	0,4711	3,8492	1,3250	10,6579	26,5692
1,0500	0,9866	4,8042	1,3500	12,0180	30,1038
1,0750	1,5497	5,8765	1,3750	13,4814	34,0161
1,1000	2,1639	7,0790	1,4000	15,0543	38,3418
1,1250	2,8327	8,4258	1,4250	16,7435	43,1194
1,1500	3,5601	9,9322	1,4500	18,5556	48,3906
1,1750	4,3500	11,6150	1,4750	20,4980	54,2005
1,2000	5,2068	13,4926	1,5000	22,5781	60,5977
1,2250	6,1348	15,5850	1,5250	24,8038	67,6342

1,2500	7,1389	17,9141	1,5500	27,1833	75,3666
1,2750	8,2239	20,5035			

Для оценки адекватности построения аналитической зависимости $\varepsilon(A^2)$ используем следующие критерии: сумму квадратов отклонений $\sigma = \sum_{i=1}^{n}(\varepsilon_{iт} - \varepsilon_i)^2$, сумму относительных погрешностей $\delta = \sum_{i=1}^{n}\left|\frac{\varepsilon_{iт} - \varepsilon_i}{\varepsilon_i}\right|$, максимальную абсолютную погрешность $\Delta\varepsilon_{max}$, максимальную относительную погрешность $\delta\varepsilon_{max}$, где $\varepsilon_{iт}$ – значение ε, взятое из табл. 1, ε – значение ε, рассчитанное по формуле (3).

Значения указанных выше критериев для двух зависимостей – нелинейной полиномиальной и линейной – в диапазоне $A^2 \in [0; 27]$ приведены в табл. 2, а для диапазона $A^2 \in [0; 5]$ – в табл. 3.

Аппроксимация методом наименьших квадратов и все расчеты были выполнены с помощью программы, разработанной на языке программирования Visual Basic for Application (VBA). Расчет с использованием этой программы показал ее работоспособность и адекватность.

Отобразим нелинейную зависимость, найденную по формуле (3), и линейную зависимость, полученную из [7], в одной плоскости (см. рис. 1).

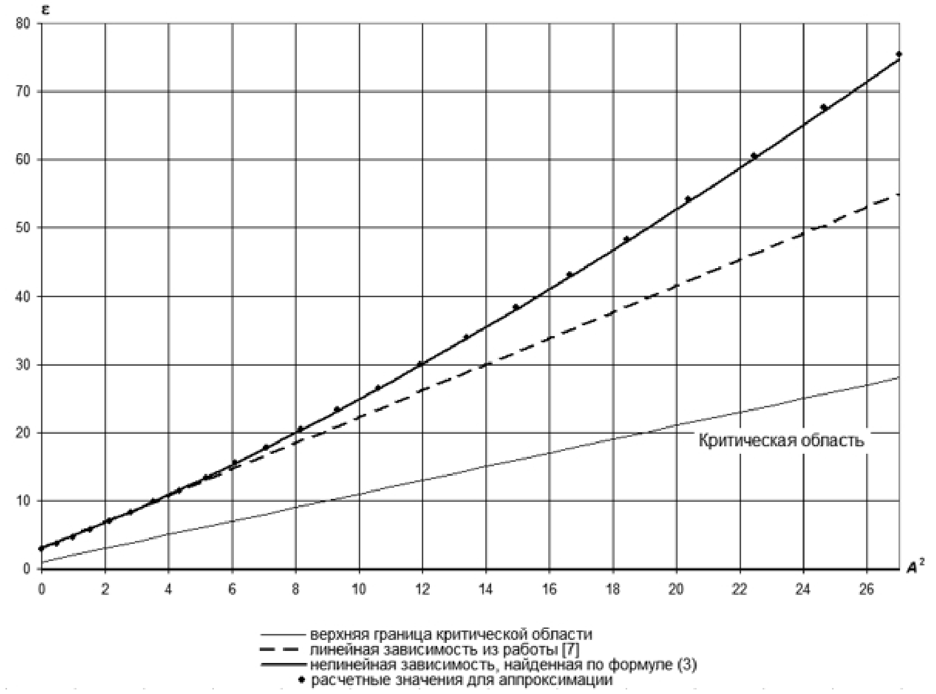

Рис. 1. Зависимость ε (A^2) в плоскости эксцесс - асимметрия в квадрате

Как видно из рисунка, указанные зависимости практически совпадают между собой в диапазоне малых значений асимметрии в квадрате до значения A^2=5, однако при увеличении значений асимметрии в квадрате их значения все больше расходятся. При значении A^2=27 расхождение составляет 19,8443 или 36,13%, что подтверждает невозможность использования линейной зависимости, приведенной в [7] для диапазона $A^2 \in [0; 27]$.

Таблица 2

Значения σ, δ, $\Delta\varepsilon_{max}$, $\delta\varepsilon_{max}$ для оценки адекватности аналитической зависимости $\varepsilon(A^2)$ в диапазоне $A^2 \in [0; 27]$

Нелинейная зависимость, найденная по формуле (3)				Линейная зависимость, полученная из [7]			
σ	δ	$\Delta\varepsilon_{max}$	$\delta\varepsilon_{max}$	σ	δ	$\Delta\varepsilon_{max}$	$\delta\varepsilon_{max}$
$7\cdot10^{-4}$	$1{,}07\cdot10^{-2}$	$1{,}09\cdot10^{-2}$	$3{,}6\cdot10^{-3}$	1290,2240	2,8729	20,0931	0,3635

Таблица 3

Значения σ, δ, $\Delta\varepsilon_{max}$, $\delta\varepsilon_{max}$ для оценки адекватности аналитической зависимости $\varepsilon(A^2)$ в диапазоне $A^2 \in [0; 5]$

Нелинейная зависимость, найденная по формуле (3)				Линейная зависимость, полученная из [7]			
σ	δ	$\Delta\varepsilon_{max}$	$\delta\varepsilon_{max}$	σ	δ	$\Delta\varepsilon_{max}$	$\delta\varepsilon_{max}$
$2{,}145\cdot10^{-5}$	$2{,}1972\cdot10^{-3}$	$1{,}09\cdot10^{-2}$	$3{,}6\cdot10^{-3}$	0,2830	0,1288	0,4799	0,0369

По результатам, приведенным в табл. 2, можно сделать вывод об адекватности зависимости, полученной по формуле (3), эмпирическим данным, а из анализа данных, приведенных в табл. 3, – о том, что даже для диапазона $A^2 \in [0; 5]$ погрешность меньше, чем при использовании линейной зависимости (1).

Выводы. Впервые получена нелинейная зависимость эксцесса от асимметрии для диапазона $A^2 \in [0; 27]$ на основе метода наименьших квадратов, что позволяет автоматизировать выбор распределения Джонсона, вплоть до значения A^2=27.

Разработана программа для нахождения нелинейной зависимости эксцесса от асимметрии на языке программирования Visual Basic for Application (VBA). Расчет с использованием этой программы показал ее работоспособность и адекватность.

В дальнейшем планируется использование полученной зависимости при решении задачи выбора аналитической модели закона распределения времени наработки между отказами устройств терминальной сети.

ЛИТЕРАТУРА:

1. Приходько, С. Б Програмне забезпечення для оцінки ризиків при управлінні програмними проектами [Текст] / С. Б. Приходько, А. В. Пухалевич // Електронний вісник НУК – Миколаїв: НУК, 2010. – №4.
2. Семенюк, С. Д. Рациональная и эффективная разработка нерудных строительных материалов (песок, гравий) в карьерах могилевской области для строительства и ремонта автодорог [Текст] / С. Д. Семенюк, С. Н. Березовский, А. Н. Терещенко // Вестник Белорусско-Российского университета: научно-методический журнал. – Могилев: ГУВПО БРУ, 2011. – № 2 (31). – С. 134-140.
3. Лебедев, С. А. Оценка рыночных рисков ценных бумаг на основе универсальных семейств распределений [Текст] / С. А. Лебедев // Сибирская финансовая школа. – Новосибирск: НОУ ВПО САФБД, 2008. – № 2. – С. 87-90.
4. Johnson N.L. System of Frequency Curves Generated by Methods of Translation // Biometrica, 1949, Vol. 36, No. 1/2 (Jun., 1949), pp.149-176.
5. Хан, Г. Статистические модели в инженерных задачах. Пер. с англ. [Текст] / Г. Хан, С. Шапиро. – М.: Мир, 1969. – 396 с.
6. Бостанджиян В.А. Распределение Пирсона, Джонсона, Вейбулла и обратное нормальное. Оценивание их параметров. [Текст] / В. А. Бостанджиян – Черноголовка: Редакционно-издательский отдел ИПХФ РАН, 2009. – 240 с.
7. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. [Текст] / А. И. Кобзарь – М.: ФИЗМАТЛИТ, 2006. – 816 с.
8. Кендалл М. Теория распределений [Текст] / М. Кендалл, А. Стьюарт. – М.: Наука, 1966. – 588 с.

ПРИХОДЬКО Сергей Борисович – к.т.н., доцент, зав. кафедрой программного обеспечения автоматизированных систем Национального университета кораблестроения им. адмирала Макарова.

Научные интересы:

– математическое моделирование случайных процессов в информационных технологиях.

МАКАРОВА Лидия Николаевна – соискатель, Национальный университет кораблестроения им. адмирала Макарова.

Научные интересы:

– математическое моделирование случайных процессов в информационных технологиях.