

АНАЛИТИЧЕСКАЯ ЗАВИСИМОСТЬ ДЛЯ ВЫБОРА СЕМЕЙСТВА РАСПРЕДЕЛЕНИЙ ДЖОНСОНА

УДК 004.942:519.25

ПРИХОДЬКО Сергей Борисович

д.т.н., профессор, заведующий кафедрой программного обеспечения автоматизированных систем,
Национальный университет кораблестроения имени адмирала Макарова,

Научные интересы: математическое моделирование случайных величин и процессов в информационных технологиях.
e-mail: sergiy.prykhodko@nuos.edu.ua.

МАКАРОВА Лидия Николаевна

к.т.н., доцент кафедры программного обеспечения автоматизированных систем,
Национальный университет кораблестроения имени адмирала Макарова,

Научные интересы: математическое моделирование случайных величин и процессов в информационных технологиях.
e-mail: lidia.makarova@nuos.edu.ua.

ПРИХОДЬКО Андрей Сергеевич

студент, Национальный университет кораблестроения имени адмирала Макарова.

Научные интересы: математическое моделирование случайных величин и процессов в информационных технологиях.

ПОСТАНОВКА ПРОБЛЕМЫ

Для выбора аналитической модели закона распределения эмпирических данных широко распространены семейства распределений Джонсона [1, 2]. Для автоматизации выбора конкретного семейства распределений Джонсона в составе информационных технологий и автоматизированных систем обработки информации необходимо иметь аналитическую зависимость эксцесса ε от асимметрии в квадрате A^2 . При решении задач выбора аналитических моделей закона распределения времени наработки между отказами и времени восстановления работоспособности устройств терминальной сети использовалась нелинейная зависимость эксцесса от асимметрии в квадрате для диапазона значений $A^2 \in [0; 27]$ [3-5]. Однако при подборе ана-

литических моделей закона распределения случайных величин в области инженерии программного обеспечения (например, длительности и трудоемкости выполнения программных проектов) встречаются эмпирические данные, для которых оценка A^2 принимает большие значения, вплоть до 40 [6]. Поэтому существует необходимость построения аналитической зависимости эксцесса ε от асимметрии $\varepsilon(A^2)$ для диапазона значений $A^2 \in [0; 40]$.

АНАЛИЗ ПОСЛЕДНИХ ИССЛЕДОВАНИЙ И ПУБЛИКАЦИЙ

В плоскости ε - A^2 присутствуют две линии, которые определяют конкретное семейство распределений Джонсона (см. рис. 1). Семейство распределений Джонсона S_L возможно использовать, когда точка с ко-

ординатами (A^2, ε) розполагається вблизи лінії S_L . Семейство розподелений Джонсона S_U применяется тогда, когда точка с координатами (A^2, ε) розполагається выше лінії S_L . Семейство розподелений Джонсона S_B возможно использовать, когда точка с координатами (A^2, ε) розполагається между лінією S_L и лінією критической области. В случае, когда точка с координатами (A^2, ε) попадает в критическую область, применять семейства розподелений Джонсона для подбора аналитической модели закона розподелений эмпирических данных нельзя.

Для выбора конкретного семейства розподелений Джонсона применяются

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

Приведенные аналитические зависимости (1) и (2) при увеличении значения A^2 более 5 и 27 соответственно дают существенные ошибки, которые влияют на выбор конкретного семейства розподелений Джонсона. Поэтому возникает необходимость в построении аналитической зависимости $\varepsilon(A^2)$ для диапазона $A^2 \in [0; 40]$ для использования в составе информационных технологий и автоматизированных систем обработки информации, а также оценки адекватности построенной зависимости при ее использовании для малых значений A^2 .

Целью данной статьи является построение аналитической зависимости эксцесса от асимметрии в квадрате для диапазона значений $A^2 \in [0; 40]$ для автоматизации выбора конкретного семейства розподелений Джонсона и оценка адекватности построенной зависимости.

ИЗЛОЖЕНИЕ ОСНОВНОГО МАТЕРИАЛА

Семейства розподелений Джонсона получены путем преобразования норми-

рованных нормально розподеленной случайной величины. В общем виде преобразование имеет вид [11]:

$$\varepsilon = 3 \cdot (1 + 0,641A^2); \quad (1)$$

либо нелинейная зависимость $\varepsilon(A^2)$, которая построена для диапазона $A^2 \in [0; 27]$ [5]:

где z – нормированная нормально розподеленная случайная величина; $\gamma, \eta, \lambda, \varphi$ – параметры розподелений, причем γ и η – параметры формы, λ – параметр масштаба, φ – параметр смещения; x – случайная величина, которая нормализуется; h – нелинейная функция конкретного семейства розподелений Джонсона:

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

где z – нормированная нормально розподеленная случайная величина; $\gamma, \eta, \lambda, \varphi$ – параметры розподелений, причем γ и η – параметры формы, λ – параметр масштаба, φ – параметр смещения; x – случайная величина, которая нормализуется; h – нелинейная функция конкретного семейства розподелений Джонсона:

$$h_1(x, \varphi, \lambda) = \ln(\tilde{x}), x > \varphi; \\ h_2(x, \varphi, \lambda) = \ln\left(\frac{\tilde{x}}{1 - \tilde{x}}\right), \varphi < x < \varphi + \lambda; \\ h_3(x, \varphi, \lambda) = \text{Arsh}(\tilde{x}), -\infty \leq x \leq +\infty.$$

Семейству функций h_1 соответствует логарифмически нормальное распре-

ление S_L Джонсона, семейству функций h_2 соответствует семейство распределений S_B Джонсона, семейству функций h_3 соответствует семейство распределений S_U Джонсона, $\bar{x} = \frac{x - \varphi}{\lambda}$.

Верхняя граница критической области при использовании аналитического способа выбора конкретного семейства распределений Джонсона задается следующим уравнением [10]:

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

Линия S_L является нелинейной и зависит от количества значений в выборке. Зависимость $\varepsilon(A^2)$ для линии S_L возможно построить, используя систему моментных уравнений, приведенную в [11]:

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

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 систему параметрических уравнений, которая и определяет зависимость $\varepsilon(A^2)$ [11]:

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

Пары значений (A^2, ε) , полученные из решения системы уравнений (3) с помощью задания значений ω , приведены в табл. 1. Аналитическая зависимость $\varepsilon(A^2)$ для диапазона значений $A^2 \in [0; 40]$ построена на основе метода наименьших квадратов по значениям, приведенным в табл. 1, и имеет вид:

$$\varepsilon(A^2) = 3,59 \cdot 10^{-6} A^8 - 4,8805 \cdot 10^{-4} A^6 + 4,1655 \cdot 10^{-2} A^4 + 1,8203 A^2 + 2,9658. \quad (4)$$

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

Для оценки адекватности построенной аналитической зависимости $\varepsilon(A^2)$ используются следующие критерии:

- сумма квадратов отклонений $\sigma = \sum_{i=1}^n (\varepsilon_i - \varepsilon_{it})^2$;
 - сумма относительных погрешностей $\delta = \sum_{i=1}^n \left| \frac{\varepsilon_i - \varepsilon_{it}}{\varepsilon_{it}} \right|$;
 - максимальная абсолютная погрешность $\Delta\varepsilon_{\max}$;
 - максимальная относительная погрешность $\delta\varepsilon_{\max}$;
- где ε_i – значение ε , рассчитанное по формуле (4), ε_{it} – значение ε , взятое из табл. 1.

Значения указанных выше критериев для диапазона значений $A^2 \in [0; 40]$ приведены в табл. 2, а для диапазона $A^2 \in [0; 5]$ – в табл. 3.

Таблиця 1

Расчетные значения для построения зависимости $\varepsilon(A^2)$ для диапазона значений $A^2 \in [0; 40]$

ω	A^2	ε	ω	A^2	ε
1,000	0,0000	3,0000	1,350	12,0180	30,1038
1,025	0,4711	3,8492	1,375	13,4814	34,0161
1,050	0,9866	4,8042	1,400	15,0543	38,3418
1,075	1,5497	5,8765	1,425	16,7435	43,1194
1,100	2,1639	7,0790	1,450	18,5556	48,3906
1,125	2,8327	8,4258	1,475	20,4980	54,2005
1,150	3,5601	9,9322	1,500	22,5781	60,5977
1,175	4,3500	11,6150	1,525	24,8038	67,6342
1,200	5,2068	13,4926	1,550	27,1833	75,3666
1,225	6,1348	15,5850	1,575	29,7250	83,8551
1,250	7,1389	17,9141	1,600	32,4380	93,1649
1,275	8,2239	20,5035	1,625	35,3315	103,3657
1,300	9,3951	23,3792	1,650	38,4152	114,5322
1,325	10,6579	26,5692	1,675	41,6992	126,7447

Таблиця 2

Значения σ , δ , $\Delta\varepsilon_{\max}$, $\delta\varepsilon_{\max}$ для оценки адекватности аналитической зависимости $\varepsilon(A^2)$ в диапазоне $A^2 \in [0; 40]$

Аналитическая зависимость	σ	δ	$\Delta\varepsilon_{\max}$	$\delta\varepsilon_{\max}$
Формула (4)	0,0069	0,0322	0,0342	0,0114
Формула (2)	2,8847	0,0343	1,4728	0,0116
Формула (1)	6990,4751	3,9141	43,5573	0,3437

Таблиця 3

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

Аналитическая зависимость	σ	δ	$\Delta\varepsilon_{\max}$	$\delta\varepsilon_{\max}$
Формула (4)	0,0031	0,0270	0,0342	0,0114
Формула (2)	0,0003	0,0080	0,0109	0,0036
Формула (1)	0,3300	0,1317	0,4799	0,0356

По результатам, приведенным в табл. 2, можно сделать вывод об адекватности построенной нелинейной зависимости (4) расчетным данным для диапазона $A^2 \in [0; 40]$. Анализ данных, приведенных в табл. 3, позволяет сделать вывод, что ту или иную зависимость рекомендуется применять для того диапазона, для которого она была построена, т.е. зависимость (2) дает лучшие результаты для

диапазона $A^2 \in [0; 27]$, а зависимость (4) – для диапазона $A^2 \in [0; 40]$.

Нелинейная зависимость, построенная по формуле (4) и линейная зависимость, построенная по формуле (1), для диапазона $A^2 \in [0; 40]$ отображены в одной координатной плоскости $\varepsilon-A^2$ на рис. 1. Графическое представление зависимостей (2) и (1) в одной координатной плоскости приведено в [5].

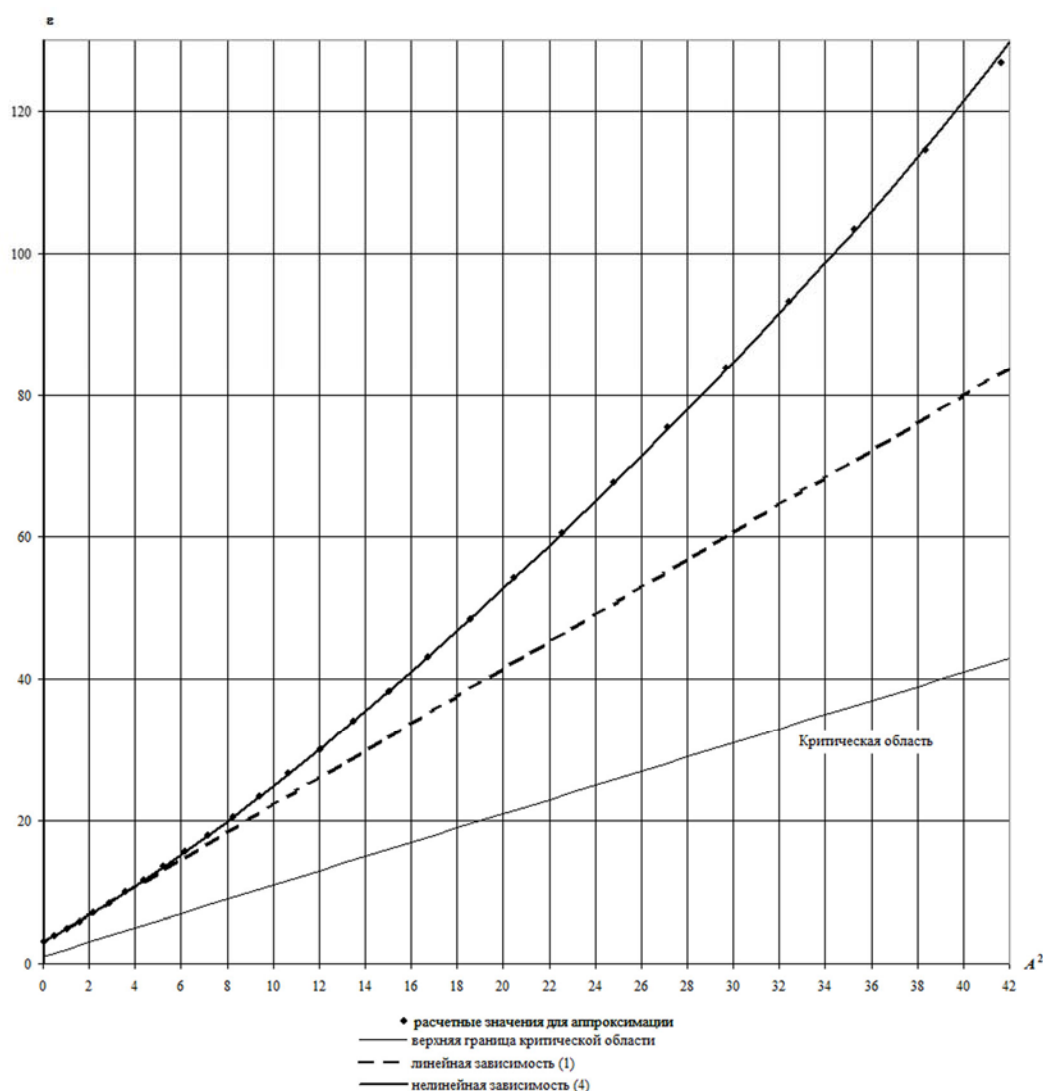


Рисунок 1. Зависимость $\varepsilon(A^2)$ в плоскости эксцесс - асимметрия в квадрате для диапазона значений $A^2 \in [0; 40]$

Из анализа рисунка 1 можно сделать вывод, что представленные зависимости практически совпадают в диапазоне малых значений асимметрии в квадрате до значения $A^2=5$, однако по мере увеличения значений асимметрии в квадрате их значения все больше расходятся. При значении $A^2=40$ расхождение между зависимостью, найденной по формуле (1) и зависимостью, найденной по формуле (4), составляет 41,49 или 51,91%, что подтверждает невозможность использования линейной зависимости (1) для диапазона $A^2 \in [0; 40]$.

ВЫВОДЫ

Впервые построена нелинейная зависимость эксцесса от асимметрии в квадрате для диапазона $A^2 \in [0; 40]$ на основе метода наименьших квадратов, что позволяет автоматизировать выбор семейства распределений Джонсона вплоть до значения $A^2=40$.

Усовершенствована программа для построения нелинейной зависимости эксцесса от асимметрии на языке программирования VBA. Расчет с использованием этой программы показал ее работоспособность.



В дальнейшем планируется использование построенной зависимости в составе информационных технологий и автоматизированных систем обработки информации для выбора семейства рас-

пределений Джонсона в зависимости от значения оценки асимметрии в квадрате при решении задачи подбора аналитической модели закона распределения случайной величины.

ЛИТЕРАТУРА

1. Primenenie raspredelenii Dzhonsona k zadache klassifikatsii aerokosmicheskikh izobrazhenii [Tekst] / Yu.B. Burkatovskaia, N.G. Markov, A.S. Morozov [i dr.] // Izvestiia Tomskogo politekhnicheskogo universiteta. – 2007. – T. 311. № 5. – S.76-80.
2. Primenenie raspredeleniia Dzhonsona dlia opredeleniia stokhasticheskikh parametrov dugovykh замыkaniy [Tekst] / A.V. Telegin, V.G. Salnikov, Yu.M. Denchik [i dr.] // Effektivnoe i kachestvennoe snabzhenie i ispolzovanie elektroenergii: sb. dokl. 5-i mezhdunar. nauch.-prakt. konf. v ramkakh spetsializir. foruma «Expo Build Russia» (Ekaterinburg, 14 apreliia 2016 g.). – Ekaterinburg: Izdatelstvo UMTS UPI, 2016. – S.243-246.
3. Prykhodko, S.B. Vybora analiticheskoi modeli zakona raspredeleniia vremeni narabotki mezhdru otkazami ustroystv terminalnoi seti [Tekst] / S.B. Prykhodko, L.N. Makarova // Naukovi pratsi: naukovo-metodychnyi zhurnal. – Vyp. 179. T. 191. Kompiuterni tekhnologiy. – Mykolayv: Vyd-vo Chdu im. Petra Mogyly, 2012. – S.42-45.
4. Prykhodko, S.B. Doveritelnyi interval nelineynoi regressii vremeni vosstanovleniia rabotosposobnosti ustroystv terminalnoi seti [Tekst] / S.B. Prykhodko, L.N. Makarova // Vostochno-evropeyskii zhurnalпередovykh tekhnologiy. Matematika i kibernetika – prikladnye aspekty. – T. 3/4 (69). – Kharkov: CHP "Tekhnologicheskii tsentr", 2014. – S.26-31.
5. Prykhodko, S.B. Analiticheskaia zavisimost dlia vybora raspredeleniia Dzhonsona semeystva S_L [Tekst] / S.B. Prykhodko, L.N. Makarova // Vestnik KHNTU. – Kherson: KHNTU, 2012. – №2 (45). – S.101-104.
6. Praynlin E., Latha P. Software Effort Estimation Models Using Radial Basis Function Network // International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol:8, No:1, 2014, pp.258-263.
7. Johnson, N.L. System of Frequency Curves Generated by Methods of Translation // Biometrika, 1949, Vol. 36, No. 1/2 (Jun., 1949), pp.149-176.
8. Khan, G. Statisticheskie modeli v inzhenernykh zadachakh. Per. s angl. [Tekst] / G. Khan, S. Shapiro. – M.: Mir, 1969. – 396 s.
9. Bostandzhiian, V.A. Raspredelenie Pirsona, Dzhonsona, Veybulla i obratnoe normalnoe. Otsenivanie ikh parametrov. [Tekst] / V.A. Bostandzhiian – Chernogolovka: Redaktsionno-izdatelckii otdel IPKHF RAN, 2009. – 240 s.
10. Kobzap, A.I. Prikladnaia matematicheskaia statistika. Dlia inzhenerov i nauchnykh rabotnikov. [Tekst] / A.I. Kobzap – M.: FIZMATLIT, 2006. – 816 s.
11. Kendall, M. Teoriia raspredelenii [Tekst] / M. Kendall, A. Stiuart – M.: Nauka, 1966. – 588 s.

Рецензент: д.т.н., проф. Ходаков В.Е.,
Херсонский национальный технический университет.