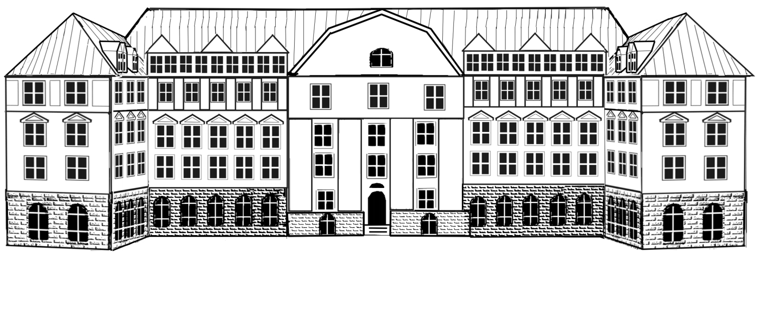
СТАТИСТИЧЕСКИЕ МЕТОДЫ ОЦЕНИВАНИЯ И ПРОВЕРКИ ГИПОТЕЗ



Пермь 2019

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»

СТАТИСТИЧЕСКИЕ МЕТОДЫ ОЦЕНИВАНИЯ И ПРОВЕРКИ ГИПОТЕЗ

Межвузовский сборник научных трудов
Вып. 29

ИЗДАТЕЛЬСКИЙ ЦЕНТР
"Perm University Press"

Пермь 2019

УДК 519.2
ББК 22.171
С781

С781 **Статистические** методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. исслед. ун-т. – Пермь, 2019. – Вып. 29. – 118 с.

ISBN 978-5-7944-3403-3 (вып. 29)
ISBN 978-5-7944-1523-0

Статьи, включенные в настоящий сборник (вып. 29), посвящены решению различных задач теоретической и прикладной математической статистики. Исследован ряд задач статистической проверки гипотез, непараметрической статистики, асимптотической теории вероятностей и математической статистики. Обсуждаются проблемы, связанные с построением и применением некоторых вероятностно-статистических моделей.

Сборник рассчитан на широкий круг специалистов по теории вероятностей и математической статистике. Он будет полезен студентам и аспирантам высших учебных заведений.

УДК 519.2
ББК 22.171

Печатается по решению ученого совета
механико-математического факультета
Пермского государственного национального исследовательского университета

ISBN 978-5-7944-3403-3 (вып. 29)
ISBN 978-5-7944-1523-0

От редакционной коллегии

Настоящий выпуск тематического сборника "Статистические методы оценивания и проверки гипотез издаваемого Пермским государственным университетом, является 29-м. Предшествующие выпуски издавались в следующие годы: вып. 1 — 1978, вып. 2 — 1980, вып. 3 — 1982, ып. 4 — 1984, вып. 5 — 1986, вып. 6 — 1988, вып. 7 — 1990, вып. 8 — 1991, вып. 9 — 1993, вып. 10 — 1995, вып. 11 — 1996, вып. 12 — 1998, вып. 13 —1999, вып. 14 — 2000, вып. 15 — 2001, вып. 16 — 2002, вып. 17 — 2003, вып. 18 — 2005, вып. 19 — 2006, вып. 20 — 2007, вып. 21 — 2008, вып. 22 — 2010, вып. 23 — 2011, вып. 24 — 2012, вып. 25 — 2013, вып. 26 — 2015, вып. 27 — 2016, вып. 28 — 2018.

Сборники 1978–1988 гг. переведены в журнале "Journal of Soviet Mathematics"(1987: Vol. 39, № 2; Vol. 39, № 4; 1988: Vol. 40, № 2; Vol. 41, № 1; 1991: Vol. 53, № 6; Vol. 56, № 3); сборники 1990–2005 гг. – в журнале "Journal of Mathematical Sciences"(1990: Vol. 75, № 1; 1991: Vol. 75, № 2; 1993: Vol. 81, № 4; 1995: Vol. 88, № 6; 1996: Vol. 103, № 4; 1998: Vol. 103, № 5; 1999: Vol. 126, № 1; 2000: Vol. 119, № 3; 2001: Vol. 127, № 4; 2002, 2003: Vol. 189, № 6; 2005: Vol. 205, № 1; 2006: Vol. 220, № 6; Vol. 228, № 5; 2007: Vol. 221, № 4; 2008: Vol. 227, № 2).

Это издание включает в себя статьи преподавателей и научных сотрудников Московского государственного университета имени М.В. Ломоносова, Пермского государственного национального исследовательского университета, Московского государственного технического университета им. Н.Э. Баумана, Национального университета Узбекистана имени Мирзо Улугбека, Ташкентского филиала Московского Государственного Университета имени М.В.Ломоносова и Каракалпакского государственного университета имени Бердаха.

Большинство статей посвящено решению задач теории оценивания и статистической проверки гипотез. В ряде публикаций рассмотрены асимптотические задачи теории вероятностей и математической статистики, а также вероятностно-статистические модели, имеющие важное прикладное значение.

УДК 519.24

Оценивание условной функции распределения в модели гетероскедастичной регрессии со слабозависимыми наблюдениями

А. А. Абдушукуров[1], Ф. А. Абдикаликов[2]

[1]Филиал Московского Государственного Университета имени М.В. Ломоносова, факультет прикладной математики и информатики, Узбекистан, Ташкент. a_abdushukurov@rambler.ru; (+99871) 232-28-22; Узбекистан, 100060, г.Ташкент, проспект Амира Темура 22, Филиал Московского Государственного Университета имени М.В.Ломоносова, кафедра "Прикладная математика и информатика"
[2]Каракалпакский государственный университет имени Бердаха, физико-математический факультет, Узбекистан, Республика Каракалпакстан abdikalikovf@mail.ru; (99861) 223-60-78; Узбекистан, 230112, г.Нукус, ул.Абидова 1, Каракалпакский государственный университет имени Бердаха, кафедра "Функциональный анализ, алгебра и геометрия"

Аннотация. *Статья посвящена задаче оценивания условной функции распределения в модели гетероскедастичной регрессии, в которой переменная отклика образует α-перемешивающуюся последовательность случайных величин. Исследована оценка условной функции распределения. Найдены представление для среднеквадратического отклонения оценки и оптимальная последовательность "ширины окна".*

Ключевые слова: *модель гетероскедастичной регрессии, α-перемешивание, ядерная оценка.*

Модель гетероскедастичной регрессии величины отклика Z на X определяется формулой [1]: $Z = m(X) + \sigma(X)\varepsilon$, где X — случайная ковариата, ε — случайная ошибка, не зависящая от X. Функционалы сдвига и масштаба $m(x) = \mathbf{M}(Z|X = x)$ и $\sigma^2(x) = \mathbf{D}(Z|X = x)$ соответственно означают функцию регрессии Z на X и условную функцию дисперсии, т.е. возможную гетероскедастичность. Определим условную функцию распределения (ф.р.) $F_x(t) = P(Z \leq t|X = x)$, $(t; x) \in R^+ \times D_X$, где D_X – носитель ковариаты X. По определению функционалы $m(x) = T(F_x(\cdot))$ и $\sigma(x) = S(F_x(\cdot))$ для всех $a \geq 0$ и $b \in R$ удовлетворяют равенствам (см. [1]):

$$T\left(Q_{aZ+b}(\cdot|x)\right) = aT\left(Q_Z(\cdot|x)\right) + b = am(x) + b,$$
$$S\left(Q_{aZ+b}(\cdot|x)\right) = aS\left(Q_Z(\cdot|x)\right) = a\sigma(x),$$

где $Q_{aZ+b}(t|x) = P\left(aZ + b \le t|X = x\right) = F_x\left(\frac{t-b}{a}\right)$. Для них справедливы представления

$$m(x) = \int\limits_0^1 F_x^{-1}(s)J(s)ds, \;\; \sigma^2(x) = \int\limits_0^1 \left(F_x^{-1}(s)\right)^2 J(s)ds - m^2(x), \qquad (1)$$

где $F_x^{-1}(s) = \inf\left\{y : F_x(y) \ge s\right\}$, $0 \le s \le 1$, — квантильная функция случайной величины (с.в.) Z при заданном $X = x$, а $J(s)$ — заданная функция метки такая, что $J(s) \ge 0$ и $\int\limits_0^1 J(s)ds = 1$. В качестве $J(s)$ можно взять функцию $J(s) = I(0 \le s \le 1)$. Основная задача состоит в оценивании условной ф.р. $F_x(t)$ по выборке наблюдений над парой (Z, X): (Z_1, X_1), $(Z_2, X_2), \ldots, (Z_n, X_n)$, где подвыборка $\{Z_1, \ldots, Z_n\}$ удовлетворяет следующему условию α-перемешивания при $n \to \infty$:

$$\alpha(n) = \left\{\sup_{k\ge1} |P(A \cap B) - P(A)P(B)| : A \in F_1^k(Z),\, B \in F_{k+n}^{\infty}(Z)\right\} \to 0,$$

где $F_i^k(Z)$ — σ-алгебра событий, порождённая набором с.в. $\{Z_j,\, i \le j \le k\}$. Среди большинства классов зависимых с.в. α-перемешивающиеся с.в. наиболее часто встречаются на практике (см. [2]). Следует отметить, что индикаторы $\{I(Z_i \le t)\}$ также образуют последовательность α-перемешивающихся случайных величин.

В качестве оценки для $F_x(t)$ рассмотрим следующую статистику Надарая—Уотсона (см. [3, 4]):

$$F_{xh}\left(t\right) = \sum_{i=1}^{n} \Psi_{ni}\left(x; h_n\right) I(Z_i \le t), \quad (t; x) \in R^+ \times \mathbf{D}_X\,, \qquad (2)$$

где весы

$$\Psi_{ni}\left(x; h_n\right) = \left(\sum_{j=1}^{n} k\left(\frac{x - X_j}{h_n}\right)\right)^{-1} k\left(\frac{x - X_i}{h_n}\right),\;\; i = 1, \ldots, n,$$

заданы последовательностью $h_n \downarrow 0$ при $n \to \infty$ и ядром $k\left(\cdot\right)$. Прямые вычисления показывают, что условные математическое ожидание и дисперсия оценки (2) при заданных $X_1, \ldots, X_n$ равны

$$\begin{aligned} \mathbf{M}^* F_{xh}(t) &= \mathbf{M}\left[F_{xh}(t)\,|X_1, ..., X_n\right] = \sum_{i=1}^{n} \Psi_{ni}(x; h_n)\, F_{X_i}(t)\,, \\ \mathbf{D}^* F_{xh}(t) &= \mathbf{D}\left[F_{xh}(t)\,|X_1, ..., X_n\right] = \sum_{i=1}^{n} \Psi_{ni}^2(x; h_n)\, F_{X_i}(t)\,(1 - F_{X_i}(t))\,. \end{aligned} \tag{3}$$

Поскольку при достаточно больших n и $x \in (h_n, 1-h_n)$, $F_{X_i}(t) \approx F_x(t)$ и $\sum\limits_{i=1}^{n} \Psi_{ni}(x; h_n) = 1$, то естественно ожидать, что правые стороны формул (3) являются асимптотически несмещенными оценками соответствующих математических ожиданий, т.е. при $n \to \infty$ и $(t, x) \in R^+ \times D_X$:

$$\mathbf{M} F_{xh}(t) = F_x(t) + o(1), \quad \mathbf{D} F_{xh}(t) = \frac{1}{nh_n} F_x(t)\,(1 - F_x(t)) + o(1). \tag{4}$$

В данной работе будет оценён квадратический риск оценки (2). Обозначим $\dot{F}_x(t) = \frac{\partial F_x(t)}{\partial x}$, $\ddot{F}_x(t) = \frac{\partial^2 F_x(t)}{\partial x^2}$. Для последовательности $\{h_n, n \geq 1\}$, ядра $k(\cdot)$ и условной ф.р. F_x введем условия:

(У1) $h_n \downarrow 0$ и $nh_n \to \infty$ при $n \to \infty$;

(У2) ядро $k(\cdot)$ является непрерывной, ограниченной и симметричной плотностью с компактным носителем $[-M, M]$ при $M > 0$;

(У3) при фиксированном $t \in R^+$ существует и ограничена производная $\ddot{F}_x(t)$ в окрестности U_x точки x.

Далее понадобится также и следующее утверждение.

Пусть ξ и η две с.в., измеримые относительно $F_1^k(Z)$ и $F_{k+n}^{\infty}(Z)$ соответственно, а также

$$\|\xi\|_\infty = ess - \sup|\xi| = \inf\{t \in R^+ \cup \{+\infty\} : P(|\xi| > t) = 0\}.$$

Лемма [2]. *Справедливо неравенство*

$$|Cov(\xi, \eta)| \leq 4\alpha(n)\|\xi\|_\infty \cdot \|\eta\|_\infty.$$

Например, для индикатора $I(Z_i \leq t)$: $ess - \sup|I(Z_i \leq t)| = 1$.

Следующее утверждение даёт оценку для среднеквадратического отклонения $F_{xh}(t)$ от $F_x(t)$.

Теорема 1. *Пусть выполнены условия (У1)–(У3) и последовательность $\{Z_i, i \geq 1\}$ удовлетворяет условию α-перемешивания так, что*

$$\alpha(n) \to 0 \quad \text{и} \quad \frac{1}{nh_n^2} \sum_{i=1}^{n-1} \alpha(i) \to 0$$

при $n \to \infty$. *Тогда при* $n \to \infty$

$$\mathbf{M}[F_{xh}(t) - F_x(t)]^2 = \left[\frac{h_n^2}{2}\ddot{F}_x(t)\int u^2 k(u)du\right]^2 +$$

$$+\frac{1}{nh_n}\int k^2(u)du \cdot F_x(t)(1 - F_x(t)) + o\left(h_n^4 + \frac{1}{nh_n}\right).$$

Доказательство. Так как

$$\mathbf{M}[F_{xh}(t) - F_x(t)]^2 = \mathbf{D}F_{xh}(t) + [\mathbf{M}F_{xh}(t) - F_x(t)]^2, \tag{5}$$

то оценим слагаемые в правой части последнего равенства по отдельности. Обозначим $\xi_i = (x - X_i)/h_n$. Используя разложение Тейлора, имеем равенства

$$F_{X_i}(t) = F_{x-h_n\xi_i}(t) = F_x(t) - h_n\xi_i\dot{F}_x(t) + \frac{h_n^2}{2}\xi_i^2\ddot{F}_x(t) + o(h_n^2), \tag{6}$$

$$F_{X_i}^2(t) = F_{x-h_n\xi_i}^2(t) = F_x^2(t) - 2h_n\xi_i F_x(t)\dot{F}_x(t) +$$
$$+ \frac{h_n^2}{2}\xi_i^2\left(\dot{F}_x(t)\right)^2 + \frac{h_n^2}{2}\xi_i^2 F_x(t)\ddot{F}_x(t) + o(h_n^2). \tag{7}$$

С учетом (6) и симметричности ядра $k(\cdot)$ для математического ожидания имеем

$$\mathbf{M}F_{xh}(t) = F_x(t) + \frac{h_n^2}{2}\ddot{F}_x(t)\frac{\sum\limits_{i=1}^n \xi_i^2 k(\xi_i)}{\sum\limits_{i=1}^n k(\xi_i)} + o(h_n^2) =$$

$$= F_x(t) + \frac{h_n^2}{2}\ddot{F}_x(t)\int u^2 k(u)du + o(1), \tag{8}$$

где использованы равенства

$$R_n^{(0)} = \frac{1}{nh_n}\sum_{i=1}^n k(\xi_i) = \int k(u)du + o(1) = 1 + o(1), \ \ n \to \infty,$$

$$R_n^{(1)} = \frac{1}{nh_n}\sum_{i=1}^n \xi_i k(\xi_i) = \int uk(u)du + o(1) = o(1), \ \ n \to \infty, \tag{9}$$

$$R_n^{(2)} = \frac{1}{nh_n}\sum_{i=1}^n \xi_i^2 k(\xi_i) = \int u^2 k(u)du + o(1), \ \ n \to \infty.$$

Из (8), в частности, при $n \to \infty$ следует асимптотическая несмещенность оценки $\mathbf{M}F_{xh}(t) = = F_x(t) + O(h_n^2)$.

Теперь вычислим условную дисперсию оценки:

$$\mathbf{D}F_{xh}(t) = \frac{\sum\limits_{i=1}^{n} k^2(\xi_i)\left[F_{X_i}(t) - F_{X_i}^2(t)\right]}{\left[\sum\limits_{i=1}^{n} k(\xi_i)\right]^2} +$$

$$+\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} \frac{k(\xi_i)k(\xi_j)Cov\left(I(Z_i \le t), I(Z_j \le t)\right)}{\left[\sum\limits_{i=1}^{n} k(\xi_i)\right]^2} = D_{1n} + D_{2n}. \qquad (10)$$

Исследуем слагаемые в правой части (10). С этой целью вычислим

$$Q_n^{(0)} = \frac{1}{nh_n}\sum_{i=1}^{n} k^2(\xi_i) = \int k^2(u)du + o(1), \;\; n \to \infty, \qquad (11)$$

а также с учетом симметричности $k(\cdot)$

$$Q_n^{(1)} = \frac{1}{nh_n}\sum_{i=1}^{n} \xi_i k^2(\xi_i) = \int uk^2(u)du + o(1) = o(1), \;\; n \to \infty. \qquad (12)$$

Используя соотношения (6), (7), (9), (11) и (12), имеем

$$D_{1n} = \left[\sum_{i=1}^{n} k(\xi_i)\right]^{-2} \left\{\left[F_x(t) - F_x^2(t)\right]\sum_{i=1}^{n} k^2(\xi_i) + \right.$$

$$\left. + \frac{h_n^2}{2}\left[\ddot{F}_x(t) - (\dot{F}_x(t))^2 - F_x(t)\ddot{F}_x(t)\right]\sum_{i=1}^{n} \xi_i^2 k^2(\xi_i) + o(1)\right\} =$$

$$= \frac{1}{nh_n}\left[F_x(t) - F_x^2(t)\right]\int k^2(u)du + o(1). \qquad (13)$$

Теперь оценим D_{2n}. Пусть $c(j-i) = Cov\left(I(Z_i \le t), I(Z_j \le t)\right)$. Тогда с учетом первого равенства в (9), леммы и условий теоремы получим

$$|D_{2n}| \le \left[R_n^{(0)}\right]^{-2}\frac{1}{n^2h_n^2}\left|\sum_{\substack{i=1\\i\neq j}}^{n}\sum_{j=1}^{n} k(\xi_i)k(\xi_j)c(j-i)\right| \le$$

$$\le \frac{2}{n^2h_n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} k(\xi_i)k(\xi_j)\left|c(j-i)\right| \le \frac{2k_0^2}{n^2h_n^2}\sum_{i=1}^{n}(n-i)|c(i)| \le$$

$$\le \frac{2k_0^2}{nh_n^2}\sum_{i=1}^{n-1}|c(i)| \le \frac{8k_0^2}{nh_n^2}\sum_{i=1}^{n-1}\alpha(i) = o(1),\ n\to\infty, \qquad (14)$$

где $k_0 = \sup\limits_{u\in R} k(u) \in (0,\infty)$. Теперь утверждение теоремы следует из (5), (8), (10), (13) и (14). Теорема 1 доказана.

Следствие. *В условиях теоремы 1 при достаточно больших n имеем следующее асимптотическое представление для дисперсии оценки* $F_{xh}(t)$*:*

$$\mathbf{D}F_{xh}(t) = \frac{1}{nh_n}F_x(t)(1-F_x(t))\int k^2(u)du + o(1),\ x\in(h_n, 1-h_n). \qquad (15)$$

При заданных ковариатах $X_1,\ldots,X_n$ из теоремы 1 также следует среднеквадратическая состоятельность оценки $F_{xh}(t)$. В условиях теоремы 1 можно записать следующее асимптотическое соотношение для квадратического риска оценки $F_{xh}(t)$ с заданной весовой функцией $w(\cdot)$ при фиксированном $t\in R^+$:

$$\int_a^b \mathbf{M}[F_{xh}(t)-F_x(t)]^2 w(x)dx \approx I(h_n),$$

где $I(h_n) = RAh_n^4/4 + QB/(nh_n)$, $Q = \int\limits_{-M}^{M} k^2(u)du$, $R = \int\limits_{-M}^{M} u^2k(u)du$, $A = \int\limits_a^b \ddot F_x(t)w(x)dx$ и $B = \int\limits_a^b F_x(t)(1-F_x(t))w(x)dx$. Для отыскания оптимальной последовательности $\{h_n, n\ge 1\}$, доставляющей наименьшее значение риску $I(h_n)$, решим уравнение:

$$\frac{\partial I(h_n)}{\partial h} = RAh_n^3 - \frac{QB}{nh_n^2} = 0,$$

откуда находим $h_{n,opt} = Cn^{-1/5}$, где $C = (QB/RA)^{1/5}$. Например, если $k(\cdot)$ – плотность равномерного распределения на $[-1,1]$, то $Q = 1/2$ и $R = 1/3$ и значение C зависит только от степени гладкости функции $F_x(t)$ по $x\in D_X$.

Библиографический список

1. *Akritas M. G., Van Keilegom I.* Nonparametric estimation of the residual estimation // Scand. J. Statist. – 2001. – V.28, №3. – P. 549–567.

2. *Doukhan P.* Mixing. Properties and Examples. – Springer-Verlag, New York, 1994. – 144 p.

3. *Надарая Э. А.* Непараметрическое оценивание плотности вероятности и кривой регрессии. – Тбилиси: ТГУ, 1983. – 194 с.

4. *Watson J. S.* Smooth regression analysis // Sankhya Ser. A – 1964. – V. 26, Part 4 – P. 359–372.

Estimation of conditional distribution function in heteroscedastical regression model with weak dependent observations.

A.A. Abdushukurov[1], F.A. Abdikalikov[2]

[1]*Branch of Moscow State University in Tashkent named after M. V. Lomonosov, Faculty of Applied Mathematics and Informatics, Department of Applied Mathematics and Informatics, Uzbekistan, Tashkent; a_ abdushukurov@rambler.ru*

[2]*Karakalpak State University named after Berdakh, Faculty of Physics and Mathematics, Department of Functional Analysis, Algebra and Geometry, Uzbekistan, Republic of Karakalpakstan; abdikalikovf@mail.ru*

Abstract. *Paper is devoted to estimation of conditional distribution function in heteroscedastical regression model, in which random variables of interest are α-mixing. In this model estimator of conditional distribution function is investigated. It is found the expression for mean square deviation of estimator and optimal window width sequence.*

Key words: *model of heteroscedastical regression, α-mixing, kernel estimate.*

УДК 519.2

О неравенствах типа Каца–Петрова–Розовского для некоторых случайных сумм[1]

А.В. Дорофеева

Факультет вычислительной математики и кибернетики МГУ имени М.В. Ломоносова; 119991, Москва, ГСП-1, Ленинские горы, МГУ, 2-й уч. корп., ф-т ВМК; alex.dorofeyeva@gmail.com

Аннотация. *Доказан аналог неравенства Каца–Петрова–Розовского, устанавливающий неравномерную оценку скорости сходимости в центральной предельной теореме для сумм случайного числа независимых одинаково распределенных случайных величин для случаев, когда индекс суммирования (число слагаемых в сумме) имеет биномиальное или пуассоновское распределение и стохастически независим от слагаемых. Рассмотрена ситуация, когда на слагаемые накладываются ослабленные моментные условия, в частности, слагаемые могут не иметь моменты степенных порядков, больших второго. На биномиальные и пуассоновские суммы перенесен результат Л. В. Розовского, из которого вытекает, что если слагаемые имеют момент порядка* $2+\delta$ *с* $\delta \in (0, 1)$*, то остаточный член в соответствующей версии центральной предельной теоремы имеет более высокий порядок малости (по числу слагаемых и аргументу), нежели степенные порядки, устанавливаемые неравенствами типа Нагаева–Бикялиса.*

Ключевые слова: *центральная предельная теорема; нормальная аппроксимация; случайная сумма; биномиальное распределение; распределение Пуассона; скорость сходимости.*

1. Введение

В данной заметке приводится результаты, уточняющие некоторые утверждения статьи [2].

Оценки точности нормальной аппроксимации для распределений сумм случайных величин традиционно являются объектом пристального внимания специалистов в области теории вероятностей, поскольку они играют важную роль во многих прикладных задачах. Такие оценки помогают осознанно принимать решения об адекватности или неадекватности нормальной модели для наблюдаемых статистических закономерностей. При этом особый интерес представляет

[1]Работа выполнена при поддержке РФФИ (проект 18-07-01405)

ситуация, в которой распределения слагаемых могут иметь довольно тяжелые хвосты. Эта ситуация в статье формализована в виде ослабленных моментных условий, налагаемых на слагаемые. В частности, слагаемые могут не иметь моменты степенных порядков, больших второго. Пусть $X_1, X_2, \ldots$ – независимые случайные величины с $\mathsf{E}X_i = 0$ и $0 < \mathsf{E}X_i^2 \equiv \sigma_i^2 < \infty$, $i = 1, 2, \ldots$ Для $n \in \mathbb{N}$ обозначим $S_n = X_1 + \ldots + X_n$, $B_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$. Стандартную нормальную функцию распределения обозначим $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2} dz, \qquad x \in \mathbb{R}.$$

Обозначим

$$\Delta_n(x) = |\mathsf{P}(S_n < xB_n) - \Phi(x)|, \quad \Delta_n = \sup_x \Delta_n(x).$$

Всюду далее символ $\mathbb{I}(A)$ будет обозначать индикаторную функцию события A.

Пусть $\mathcal{G}$ – класс вещественных функций $g(x)$ аргумента $x \in \mathbb{R}$ таких, что

- функция $g(x)$ четна;
- функция $g(x)$ неотрицательна при всех x и $g(x) > 0$ при $x > 0$;
- функция $g(x)$ не убывает при $x > 0$;
- функция $x/g(x)$ не убывает при $x > 0$.

Классу $\mathcal{G}$ принадлежат, например, функции $g(x) \equiv \mathrm{const} > 0$, $g(x) = |x|^\delta$ при $0 \le \delta \le 1$, $g(x) = \min\{|x|, \mathrm{const}\}$.

В 1963 г. М. Кац [12] доказал, что, какой бы ни была функция $g \in \mathcal{G}$, если случайные величины $X_1, X_2, ...$ одинаково распределены, причем $\mathsf{E}X_1^2 g(X_1) < \infty$, то существует конечная положительная абсолютная постоянная C_1 такая, что

$$\Delta_n \le C_1 \cdot \frac{\mathsf{E}X_1^2 g(X_1)}{\sigma_1^2 g\big(\sigma_1\sqrt{n}\big)}. \tag{1}$$

В 1965 г. этот результат был обобщен В. В. Петровым [6] на случай неодинаково распределенных случайных величин (также см. [7]): какой бы ни была функция $g \in \mathcal{G}$, если $\mathsf{E}X_k^2 g(X_k) < \infty$, $k = 1, ..., n$, то существует конечная положительная абсолютная постоянная C_2 такая, что

$$\Delta_n \le \frac{C_2}{B_n^2 g(B_n)} \sum_{k=1}^{n} \mathsf{E}X_k^2 g(X_k). \tag{2}$$

В 1979 г. В. В. Петров [8] доказал неравномерный аналог неравенства (2) (также см. [6]): какой бы ни была функция $g \in \mathcal{G}$, если $\mathsf{E}X_k^2 g(X_k) < \infty$, $k = 1, ..., n$, то существует конечная положительная абсолютная постоянная C_3 такая, что для любого $x \in \mathbb{R}$

$$\Delta_n(x) \leq \frac{C_3}{B_n^2(1+|x|)^2 g\big(B_n(1+|x|)\big)} \sum_{k=1}^{n} \mathsf{E}X_k^2 g(X_k). \tag{3}$$

В частности, если случайные величины одинаково распределены и $\mathsf{E}X_1^2 g(X_1) < \infty$ для некоторой функции $g \in \mathcal{G}$, то существует абсолютная постоянная C такая, что для любого $x \in \mathbb{R}$

$$\Delta_n(x) \leq \frac{C \cdot \mathsf{E}X_1^2 g(X_1)}{\sigma_1^2(1+|x|)^2 g\big(\sigma_1\sqrt{n}(1+|x|)\big)}. \tag{4}$$

Неравенство (3) получено в [8] как следствие оценки

$$\Delta_n(x) \leq C \sum_{k=1}^{n} \left[\frac{\mathsf{E}X_k^2 \mathbb{I}\big(|X_k| \geq (1+|x|)B_n\big)}{B_n^2(1+|x|)^2} + \frac{\mathsf{E}|X_k|^3 \mathbb{I}\big(|X_k| < (1+|x|)B_n\big)}{B_n^3(1+|x|)^3} \right], \tag{5}$$

которая для случая одинаково распределенных слагаемых принимает вид

$$\Delta_n(x) \leq \frac{C}{\sigma^2(1+|x|)^2} \cdot \mathsf{E}X_1^2 \min\left\{1,\ \frac{|X_1|}{\sigma\sqrt{n}(1+|x|)}\right\}. \tag{6}$$

(Интересно отметить, что в свою очередь, неравенства (5) и (6) являются частными случаями неравенств (3) и (4) соответственно со специальной функцией $g(y) = \min\{|y|,\ 1\}$).

В статье Л. В. Осипова и В. В. Петрова [5] было получено уточнение результатов С. В. Нагаева [4] и А. Бикялиса [1]. А именно, было доказано, что если $\mathsf{E}|X_1|^{2+\delta} < \infty$ при некотором $\delta \in (0, 1)$, то

$$\Delta_n(x) \leq \frac{Q\big(\sqrt{n}(1+|x|)\big)}{n^{\delta/2}(1+|x|)^{2+\delta}},$$

где $Q(y)$ – ограниченная функция, такая что $Q(y) \to 0$ при $y \to \infty$. Другими словами, если $\mathsf{E}|X_1|^{2+\delta} < \infty$ при $0 < \delta < 1$, то $\Delta_n(x) = o\big(n^{-\delta/2}(1+|x|)^{-(2+\delta)}\big)$. Условие $\delta < 1$ существенно, при $\delta = 1$ этот результат не имеет места.

В работе [9] упомянутая оценка была обобщена. А именно, пусть $\mathcal{G}_0$ – класс, содержащий функции $g(x)$ такие, что

- $g(x) \in \mathcal{G}$;

- $\lim\limits_{x\to\infty} \dfrac{g(x^2)}{xg(x)} = 0.$

Классу $\mathcal{G}_0$ принадлежит, например, функция $g_\delta(x) \equiv |x|^\delta$ при $0 < \delta < 1$. В то же время функция $g_1(x) \equiv |x|$, принадлежащая классу $\mathcal{G}$, не входит в класс $\mathcal{G}_0$. Еще одним примером функции из $\mathcal{G}_0$ является $g(x) = \exp\left\{\left(\log(e+|x|)\right)^p\right\}$ при $0 < p < 1$ (см. [9]).

В работе [9] было доказано, что если $\mathsf{E}X_1^2 g(X_1) < \infty$ для некоторой функции $g(x) \in \mathcal{G}_0$, то

$$\Delta_n(x) \le \frac{Q\big(\sigma\sqrt{n}(1+|x|)\big)}{(1+|x|)^2 g\big(\sigma\sqrt{n}(1+|x|)\big)}$$

для всех $n \in \mathbb{N}$ и $x \in \mathbb{R}$, где $Q(y)$ – ограниченная функция, такая, что $Q(y) \to 0$ при $y \to \infty$.

В работе [2] упомянутые выше результаты В. В. Петрова были перенесены на биномиальные и пуассоновские случайные суммы и явно указан вид функции $Q(y)$.

Неравенство (8) было уточнено Л. В. Розовским [10], который показал, что (8) справедливо для функций из более широкого класса функций, нежели $\mathcal{G}_0$. А именно это неравенство остается справедливым для $g \in \mathcal{G}_1$, где $\mathcal{G}_1$ – класс, содержащий функции $g(x)$ такие, что

- $g(x) \in \mathcal{G}$;
- $\lim\limits_{x\to\infty} \dfrac{x}{g(x)} = \infty.$

Заметим, что из результатов работ [5], [9] и [10] вытекает, что если слагаемые имеют момент порядка $2+\delta$ с $\delta \in (0, 1)$, то остаточный член в соответствующей версии центральной предельной теоремы имеет более высокий порядок малости (по числу слагаемых и аргументу), нежели степенные порядки, устанавливаемые классическими неравенствами типа Нагаева–Бикялиса.

Цель данной работы – перенести на биномиальные и пуассоновские случайные суммы неравенство Л. В. Розовского [10], распространяющее упомянутый выше результат В. В. Петрова [9] на более широкий класс функций.

Вспомогательные результаты

Всюду далее рассматриваются независимые одинаково распределенные случайные величины $X_1, X_2, \ldots$ с $\mathsf{E}X_i = 0$ и $0 < \mathsf{E}X_i^2 \equiv \sigma^2 < \infty$. Пусть

$p \in (0,1]$ – произвольно. Рассмотрим случайную величину $N_{n,p}$, имеющую биномиальное распределение с параметрами n и p:

$$\mathsf{P}(N_{n,p} = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, \ldots, n.$$

Предположим, что при каждом $n \in \mathbb{N}$ случайные величины $N_{n,p}, X_1, X_2, \ldots$ взаимно независимы. Рассмотрим *биномиальную случайную сумму*

$$S_{N_{n,p}} = X_1 + \ldots + X_{N_{n,p}}.$$

При этом если $N_{n,p} = 0$, то $S_{N_{n,p}} = 0$.

Обозначим

$$\Delta_{n,p}(x) = \Big|\mathsf{P}\big(S_{N_{n,p}} < x\sigma\sqrt{np}\big) - \Phi(x)\Big|.$$

В работе [3] доказано следующее утверждение.

Лемма 1 [3]. *Для любых $n \in \mathbb{N}$ и $p \in (0,1]$ справедливо неравенство*

$$\Delta_{n,p}(x) \leq$$

$$\leq 37\left[\frac{\mathsf{E}X_1^2\mathbb{I}\big(|X_1| \geq (1+|x|)\sigma\sqrt{np}\big)}{\sigma^2(1+|x|)^2} + \frac{\mathsf{E}|X_1|^3\mathbb{I}\big(|X_1| < (1+|x|)\sigma\sqrt{np}\big)}{\sigma^3\sqrt{np}(1+|x|)^3}\right] =$$

$$= \frac{37}{\sigma^2(1+|x|)^2} \cdot \mathsf{E}X_1^2 \min\left\{1,\ \frac{|X_1|}{\sigma\sqrt{np}(1+|x|)}\right\}.$$

В работе [2] с помощью леммы 1 доказана следующая теорема.

Теорема 1 [2]. *Предположим, что $\mathsf{E}X_1^2 g(X_1) < \infty$ для некоторой функции $g \in \mathcal{G}$. Тогда для любых $n \in \mathbb{N}$, $p \in (0,1]$ и $x \in \mathbb{R}$ справедливо неравенство*

$$\Delta_{n,p}(x) \leq \frac{37\,\mathsf{E}X_1^2 g(X_1)}{\sigma^2(1+|x|)^2 g((1+|x|)\sigma\sqrt{np})}.$$

Пусть N_λ – случайная величина, имеющая распределение Пуассона с параметром $\lambda > 0$:

$$\mathsf{P}(N_\lambda = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N} \cup \{0\}.$$

Предположим, что при каждом λ случайные величины $N_\lambda, X_1, X_2, \ldots$ независимы. Рассмотрим *пуассоновскую случайную сумму*

$$S_{N_\lambda} = X_1 + \ldots + X_{N_\lambda}.$$

Если $N_\lambda = 0$, то полагаем $S_{N_\lambda} = 0$. Несложно убедиться, что $\mathsf{E}S_{N_\lambda} = 0$ и $\mathsf{D}S_{N_\lambda} = \lambda\sigma^2$. Точность нормальной аппроксимации для распределений пуассоновских случайных сумм изучалась многими авторами (см. исторические обзоры в работах [11, 13]). Равномерные оценки точности нормальной аппроксимации для распределений пуассоновских случайных сумм при ослабленных моментных условиях получены в статье [14]. Насколько известно автору, неравномерные оценки величины

$$\Delta_\lambda(x) = \left|\mathsf{P}\big(S_{N_\lambda} < x\sigma\sqrt{\lambda}\big) - \Phi(x)\right|$$

для такой ситуации впервые были получены в работе [3], где доказан следующий результат.

Лемма 2 [3]. *Для любых $\lambda > 0$ и $x \in \mathbb{R}$ справедлива оценка*

$$\Delta_\lambda(x) \leq$$

$$\leq 37\left[\frac{\mathsf{E}X_1^2\mathbb{I}\big(|X_1| \geq (1+|x|)\sigma\sqrt{\lambda}\big)}{\sigma^2(1+|x|)^2} + \frac{\mathsf{E}|X_1|^3\mathbb{I}\big(|X_1| < (1+|x|)\sigma\sqrt{\lambda}\big)}{\sigma^3\sqrt{\lambda}(1+|x|)^3}\right] =$$

$$= \frac{37}{\sigma^2(1+|x|)^2}\cdot\mathsf{E}X_1^2\min\left\{1,\, \frac{|X_1|}{\sigma\sqrt{\lambda}(1+|x|)}\right\}.$$

В работе [2] с помощью леммы 2 доказана следующая теорема.

Теорема 2. *Предположим, что $\mathsf{E}X_1^2 g(X_1) < \infty$ для некоторой функции $g \in \mathcal{G}$. Тогда для любых $\lambda > 0$ и $x \in \mathbb{R}$ справедливо неравенство*

$$\Delta_\lambda(x) \leq \frac{37\,\mathsf{E}X_1^2 g(X_1)}{\sigma^2(1+|x|)^2 g((1+|x|)\sigma\sqrt{\lambda})}.$$

Основные результаты

В работе [2] теоремы 1 и 2 были уточнены для функций из класса $\mathcal{G}_0$. Здесь будет получен аналогичный результат для функций из класса $\mathcal{G}_1$, рассмотренного Л. В. Розовским [10].

Леммы 1 и 2 позволяют, дословно повторяя рассуждения из статьи [10], заменив при этом n на np в случае биномиальных сумм и на λ в случае пуассоновских сумм, получить следующие результаты.

Теорема 3. *Предположим, что $\mathsf{E}X_1^2 g(X_1) < \infty$ для некоторой функции $g \in \mathcal{G}_0$. Тогда для любых $n \in \mathbb{N}$, $p \in (0,1]$ и $x \in \mathbb{R}$ справедливо неравенство*

$$\Delta_{n,p}(x) \leq \frac{Q^*\big((1+|x|)\sigma\sqrt{np}\big)}{\sigma^2(1+|x|)^2 g\big((1+|x|)\sigma\sqrt{np}\big)},$$

где

$$Q^*(y) = 37\,g(y)\mathsf{E}\left[X_1^2\min\left\{1,\, \frac{|X_1|}{y}\right\}\right]. \tag{7}$$

При этом $\lim_{y\to\infty} Q^*(y) = 0$.

Теорема 4. *Предположим, что* $\mathsf{E}X_1^2 g(X_1) < \infty$ *для некоторой функции* $g \in \mathcal{G}_0$. *Тогда для любых* $\lambda > 0$ *и* $x \in \mathbb{R}$ *справедливо неравенство*

$$\Delta_\lambda(x) \leq \frac{Q^*\big((1+|x|)\sigma\sqrt{\lambda}\big)}{\sigma^2(1+|x|)^2 g\big((1+|x|)\sigma\sqrt{\lambda}\big)},$$

где функция $Q^*(y)$ *определена в* (7). *При этом* $\lim_{y\to\infty} Q^*(y) = 0$.

Доказательство. Чтобы убедиться в том, что функция $Q^*(y)$ обладает требуемыми свойствами, воспользуемся леммами 1 и 2, следуя работе [10], заметим, что для любого положительного y

$$\mathsf{E}X_1^2 \min\Big\{1,\ \frac{|X_1|}{y}\Big\} = \frac{1}{g(y)} \cdot \mathsf{E}X_1^2 g(X_1) A(X_1, y),$$

где

$$A(x,y) = \frac{g(y)}{g(x)} \cdot \min\Big\{1,\ \frac{x}{y}\Big\}.$$

Так как $g \in \mathcal{G}_1$, то $A(x,y) \leq 1$ и $A(x,y) \to 0$ при $y \to \infty$ при любом x. Поэтому по теореме о мажорируемой сходимости $Q^*(y) = \mathsf{E}\big[A(X_1,y)X_1^2 g(X_1)\big] \to 0$ при $y \to \infty$.

Библиографический список

1. *Бикялис А.* Оценки остаточного члена в центральной предельной теореме // Литовский математический сборник, 1966. – Т. 6. – Вып. 3. – С. 323–346.

2. *Дорофеева А.В.* Неравенства типа Каца–Петрова для некоторых случайных сумм // Статистические методы оценивания и проверки гипотез, 2018. – Вып. 28. – С. 66–75.

3. *Королев В. Ю., Дорофеева А.В.* О неравномерных оценках точности нормальной аппроксимации для распределений некоторых случайных сумм при ослабленных моментных условиях // Информатика и ее применения, 2018. – Т. 12. – Вып. 4. – С. 91–96.

4. *Нагаев С. В.* Некоторые предельные теоремы для больших уклонений // Теория вероятностей и ее применения, 1965. – Т. 10. – Вып. 2. – С. 231–254.

5. *Осипов Л.В., Петров В.В.* Об оценке остаточного члена в центральной предельной теореме // Теория вероятностей и ее применения, 1967. – Т. 12. – Вып.2. – С. 322–329.

6. *Петров В.В.* Одна оценка отклонения распределения суммы независимых случайных величин от нормального закона // Доклады АН СССР, 1965. – Т. 160. – Вып. 5. – С. 1013–1015.

7. *Петров. В. В.* Суммы независимых случайных величин.– М.: Наука, 1972. – 416 с.

8. *Петров В.В.* Одна предельная теорема для сумм независимых неодинаково распределенных случайных величин // Записки научных семинаров ЛОМИ, 1979. – Т.85. – С. 188–192.

9. *Петров В.В.* Об оценке остаточного члена в центральной предельной теореме // Записки научных семинаров ПОМИ, 2007. – Т. 341. – С. 142–146.

10. *Розовский Л.В.* Одна неравномерная оценка остаточного члена в центральной предельной теореме // Записки научных семинаров ПОМИ, 2007. – Т. 351. – С. 238–241.

11. *Шевцова И.Г.* О точности нормальной аппроксимации для обобщенных пуассоновских распределений // Теория вероятностей и ее применения, 2014. – Т.58. – Вып.1. – С. 138–158.

12. *Katz M.* Note on the Berry–Esseen theorem // Annals of Math. Statist., 1963. – Vol. 39. – №4. – P. 1348–1349.

13. *Korolev V.Yu., Shevtsova I.G.* An improvement of the Berry–Esseen inequality with applications to Poisson and mixed Poisson random sums // Scandinavian Actuarial Journal, 2012. – №2. – P. 81–105.

14. *Korolev V., Dorofeeva A.* Bounds of the accuracy of the normal approximation to the distributions of random sums under relaxed moment conditions // Lithuanian Mathematical Journal, 2017. – Vol. 57. – №1. – P. 38–58.

On the Katz–Petrov–Rozovskii-type inequalities for some random sums

A.V. Dorofeeva

Faculty of Computational Mathematics and Cybernetics of Moscow State University; 119991, Moscow, GSP-1, Leninskie Gory, Moscow State University, Faculty of Computational Mathematics and Cybernetics; alex.dorofeyeva@gmail.com

Abstract. *Analogues of the Katz–Petrov–Rozovskii-type inequalities are proved that establish non-uniform convergence rate estimates in the central limit theorem for sums of a random number of independent identically distributed random variables for cases where summation index (the number of terms in sum) has the binomial or Poisson distribution and is stochastically independent of the terms. The case is considered where relaxed moment conditions are imposed on the terms, in particular, the terms may not have moments higher than the second. A result of L. V. Rozovskii sharpening the Katz-Petrov inequalities is extended to binomial and Poisson random sums. From this result it follows that in the case where the terms have order of moments equal to* $2 + \delta$ *with* $\delta \in (0, 1)$*, the remainder term in the corresponding version of the central limit theorem has a higher order of smallness (in terms of the number of terms and argument) than power orders established by the Nagaev–Bikelis-type inequalities.*

Key words: *central limit theorem; normal approximation; random sum; binomial distribution; Poisson distribution; convergence rate.*

УДК 519.24

Асимптотический анализ распределения суммы линейного процесса

Т. М. Зупаров

Национальный Университет Узбекистана имени Мирзо Улугбека, математический факультет, Узбекистан, Ташкент. rasulova_nargiza@mail.ru;(+99871)246-90-80; Узбекистан, 100174, г. Ташкент, ВУЗ городок, Национальный Университет Узбекистана, кафедра "Теория вероятностей и математическая статистика"

Аннотация. *Данная работа посвящена асимптотическому анализу распределения суммы первых n членов линейного процесса, порожденного инновационной последовательностью* $\{\xi_{kn}\}$. *С помощью Бевереджа-Нельсона разложения (БН-разложения) доказан критерий слабой сходимости нормированной суммы линейного процесса к предельному распределению и как следствие этого результата получен аналог теоремы Линдеберга-Феллера для линейных процессов, порожденных* $\varphi-$ *перемешивающейся инновационной последовательностью.*

Ключевые слова: *линейный процесс, БН-разложения, почти наверное сходимость, теорема Линдберга-Феллера.*

Пусть $\{\xi_{kn}, k \in \mathbb{Z}, k \leq n, n \geq 1\}-$ последовательность серий случайных величин удовлетворяет следующему условию:
(A) для любого $\varepsilon > 0$ справедливо соотношение

$$P\left\{\sup_{k\in\mathbb{Z},k\leq n} |\xi_{kn}| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0.$$

Пусть $\{a_{in}, i \in Z, n \geq 1\}$ – такая последовательность чисел, что выполнены условия:

(B) $\sum\limits_{j=0}^{\infty} |\tilde{a}_{jn}| < \infty$, $\sum\limits_{j=0}^{\infty} a_{jn} = a(n) \neq 0$ и $\sup\limits_{0\leq j<\infty} \left|\frac{\tilde{a}_{jn}}{a(n)}\right| \leq C$, где C – число, не зависящее от n, и $\tilde{a}_{jn} = \sum\limits_{h=j+1}^{\infty} a_{hn}$.

Определение 1. Если ряд $X_{kn} = \sum\limits_{i=0}^{\infty} a_{in}\xi_{k-i,n}$ сходится с вероятностью 1, тогда последовательность случайных величин $\{X_{kn}, k \in \mathbb{Z}\}$ называется линейным процессом с коэффициентами $\{a_{kn}, k \in Z, n \geq 1\}$, порожденным инновационным процессом $\{\xi_{kn}, k \in Z\}$.

Замечание. Если выполнены условия (A) и (B), то ряд $\sum\limits_{i=0}^{\infty} a_{in}\xi_{k-i,n}$ сходится с вероятностью 1, и следовательно линейный процесс в этом случае определен вполне корректно.

Работа посвящена асимптотическому анализу распределения суммы первых n членов линейного процесса, порожденного инновационной последовательностью $\{\xi_{kn}\}$. С помощью Бевереджа–Нельсона разложения (БН-разложения) доказан критерий слабой сходимости нормированной суммы линейного процесса к предельному распределению и как следствие этого результата получен аналог теоремы Линдеберга–Феллера для линейных процессов, порожденных φ-перемешивающейся инновационной последовательностью. Кроме того, доказан усиленный закон больших чисел для линейного процесса, порожденного сильно зависимой стационарной последовательностью. Многие результаты по асимптотике распределения суммы линейного процесса можно найти в [2,6–10].

Пусть L – оператор сдвига, который отображает случайный процесс $\{\xi_{kn}, k \in \mathbb{Z}, k \leq n\}$ в себя так, что $L\xi_{kn} = \xi_{k-1,n}$ и $A(L) = \sum\limits_{j=0}^{\infty} a_{jn}L^j$– многочлен от оператора сдвига, с помощью которого мы можем записать линейный процесс X_{kn} в виде $X_{kn} = A(L)\xi_{kn} = \sum\limits_{j=0}^{\infty} a_{kn}\xi_{k-j,n}$.

При установлении асимптотического поведения распределения суммы линейного процесса важную роль играет следующее БН-разложение:

$$A(L) = A(1) - (1 - L)\sum_{j=0}^{\infty} \tilde{a}_{jn}. \tag{1}$$

БН-разложение (1) формально было применено в работе Beveridje и Nelson [1] при изучении флуктуации циклов в коммерческой деятельности. Phillips и Solo [2] дали общую трактовку БН-разложения и применили его для доказательства закона больших чисел (ЗБЧ) и принципа инвариантности для линейных процессов, порожденных независимыми одинаково распределенными инновациями.

Согласно БН-разложению получаем соответствующие разложения для линейного процесса и суммы первых n ее членов:

$$X_{kn} = A(L)\xi_{kn} = A(1)\xi_{kn} - (1-L)\sum_{j=0}^{\infty} \tilde{a}_{jn} L^j = a(n)\xi_{kn} - \left(\tilde{\xi}_{kn} - \tilde{\xi}_{k-1,n}\right),$$

$$S_n = \sum_{k=1}^{n} X_{kn} = a(n)\sum_{k=1}^{n} \xi_{kn} + \left(\tilde{\xi}_{0n} - \tilde{\xi}_{nn}\right), \tag{2}$$

где $\tilde{\xi}_{tn} = \sum\limits_{j=0}^{\infty} \tilde{a}_{jn}\xi_{t-j,n}$.

С помощью разложения (2) получим следующую теорему переноса.

Теорема 1. *Пусть $\{X_{kn}\}$ – линейный процесс, с коэффициентами $\{a_{kn}\}$, порожденный инновационным процессом $\{\xi_{kn}\}$, который удовлетворяет условию (A). Если, кроме того, выполнены условия (B) и $F_n(x) \Rightarrow F(x)$ (условие (C)), где $F_n(x) = P\left(\sum\limits_{k=1}^{n} \xi_{kn} \le x\right)$, а $F(x)$ – произвольная функция распределения, то*

$$P\left(\frac{1}{a(n)}\sum_{k=1}^{n} X_{kn} \le x\right) \Rightarrow F(x).$$

Доказательство. Согласно (2) имеем

$$\frac{1}{a(n)}\sum_{k=1}^{n} X_{kn} = \sum_{k=1}^{n} \xi_{kn} + \frac{\tilde{\xi}_{0n} - \xi_{nn}}{a(n)}. \tag{3}$$

Поскольку по условию (C) функция распределения $F_n(x)$ суммы $\sum\limits_{k=1}^{n} \xi_{kn}$ слабо сходится к функции распределения $F(x)$, то для доказательства теоремы нам достаточно показать сходимость по вероятности к нулю второго слагаемого правой части равенства (3). Это утверждение следует из следующей леммы.

Лемма 1. *При выполнении условий (A) и (B) имеет место соотношение:*

$$\frac{\tilde{\xi}_{tn}}{a(n)} \xrightarrow{\text{п.н.}} 0. \tag{4}$$

Доказательство. Соотношение (4) эквивалентно (см. [3], стр.269) утверждению: для любого $\varepsilon > 0$, при $n \to \infty$

$$P\left(\sup_{t \ge n} \left|\frac{\tilde{\xi}_{tn}}{a(n)}\right| > \varepsilon\right) \to 0. \tag{5}$$

Имеем

$$P_n(\varepsilon) := P\left(\sup_{k\geq n}\left|\frac{\tilde{\xi}_{kn}}{a(n)}\right| > \varepsilon\right) =$$

$$= P\left(\sup_{k\geq n}\left|\frac{1}{a(n)}\sum_{j=0}^{\infty}\tilde{a}_{jn}\xi_{k-j,n}\right| > \varepsilon\right) \leq P\left(\sum_{j=0}^{\infty}\frac{|\tilde{a}_{jn}|}{|a(n)|}\sup_k|\xi_{kn}| > \varepsilon\right).$$

в силу условия (B) ряд $\sum\limits_{j=1}^{\infty}[\tilde{a}_{jn}/a(n)]$ абсолютно сходится. Отсюда следует, что его остаток стремится к нулю. Другими словами, для любого $\varepsilon > 0$ существует целое положительное число $K = K(\varepsilon)$, зависящее только от ε такое, что выполнено неравенство

$$\sum_{j=k+1}^{\infty}\left|\frac{\tilde{a}_{jn}}{a(n)}\right| < \frac{\varepsilon}{2}. \tag{6}$$

Имеет место неравенство:

$$P_n(\varepsilon) \leq P\left(\sum_{j=1}^{K}\left|\frac{\tilde{a}_{jn}}{a(n)}\right|\sup_{k\in\mathbb{Z}}\left|\tilde{\xi}_{kn}\right| > \frac{\varepsilon}{2}\right) +$$

$$+P\left(\sum_{j=K+1}^{\infty}\left|\frac{\tilde{a}_{jn}}{a(n)}\right|\sup_{k\in\mathbb{Z}}\left|\tilde{\xi}_{kn}\right| > \frac{\varepsilon}{2}\right) =: P_{n1}(\varepsilon) + P_{n2}(\varepsilon). \tag{7}$$

Теперь для доказательства леммы 1 нам достаточно показать стремление к нулю обоих слагаемых в правой части неравенства (7).

Применяя условия (B) и (A), имеем

$$P_{n1}(\varepsilon) = P\left(\sum_{j=1}^{K}\left|\frac{\tilde{a}_{jn}}{a(n)}\right|\sup_{k\in\mathbb{Z}}\left|\tilde{\xi}_{kn}\right| > \frac{\varepsilon}{2}\right) \leq P\left(C\sum_{j=1}^{K}\sup_{k\in\mathbb{Z}}|\xi_{kn}| > \frac{\varepsilon}{2}\right) =$$

$$= P\left(\sup_{k\in\mathbb{Z}}|\xi_{kn}| > \frac{\varepsilon}{2KC}\right) \to 0 \text{ при } n \to \infty.$$

Теперь докажем стремление к нулю второго слагаемого в правой части неравенства (7). Пусть $A_n = \left\{\omega : \sup\limits_{k\in\mathbb{Z}}|\xi_{kn}| \leq 1\right\}$, тогда в силу условия (A)

$$P(\bar{A}_n) \to 0, \text{ при } n \to \infty. \tag{8}$$

Рассмотрим выражение

$$P_{n2}(\varepsilon) = P\left(\sum_{j=K+1}^{\infty}\left|\frac{\tilde{a}_{jn}}{a(n)}\right|\sup_{k\in\mathbb{Z}}|\xi_{kn}| > \varepsilon\right) =$$

$$P\left(\sum_{j=K+1}^{\infty}\left|\frac{\tilde{a}_{jn}}{a(n)}\right|\sup_{k\in\mathbb{Z}}|\xi_{kn}|>\varepsilon, A\right)+P\left(\sum_{j=K+1}^{\infty}\left|\frac{\tilde{a}_{jn}}{a(n)}\right|\sup_{k\in\mathbb{Z}}|\xi_{kn}|>\varepsilon, \bar{A}\right).$$

Здесь первое слагаемое равно нулю согласно (6), а второе, в силу соотношения (8), стремится к нулю при $n\to\infty$, так как

$$P\left(\sum_{j=K+1}^{\infty}\left|\frac{\tilde{a}_{jn}}{a(n)}\right|\sup_{k\in\mathbb{Z}}|\xi_{kn}|>\varepsilon, \bar{A}\right)\le P(\bar{A})\to 0.$$

Лемма доказана.

Замечание 3. С помощью леммы 1 и БН-разложения можно перенести многие результаты, связанные с почти наверное сходимостью для сумм случайных величин, на суммы для линейных процессов. Однако, в некоторых случаях условие (A) либо не выполняется, либо его доказательство является трудно выполнимой задачей. Для иллюстрации этого утверждения мы здесь приведем доказательство усиленного закона больших чисел для линейного процесса, порожденного стационарным, в узком смысле инновационным процессом, удовлетворяющим условию сильной зависимости.

В работе [4] доказана следующая теорема.

Теорема 2 [4]. *Пусть* $\{\xi_i\}$ – *неотрицательные случайные величины и* $\eta_n=\sum_{i=1}^{n}\xi_i$. *Если* $\sup\mathbf{M}\xi_i<\infty$, $\mathbf{M}\eta_n\to\infty$ *и существует* $\gamma>1$ *такое, что*

$$\mathbf{D}\eta_n=O\left(\frac{(\mathbf{M}\eta_n)^2}{(\ln\mathbf{M}\eta_n)(\ln\ln\mathbf{M}S_n)^{\gamma}}\right),$$

то $\dfrac{\eta_n}{\mathbf{M}\eta_n}\xrightarrow{\text{п.н.}}1$ *при* $n\to\infty$.

Следствие 1 (см. [4]). *Пусть* $\{\xi_i\}$ – *одинаково распределенные случайные величины,* $\mathbf{M}\xi_i=\mu$, $\mathbf{M}\xi_i^2<\infty$ *и* $\eta_n=\sum_{i=1}^{n}\xi_i$. *Если* $\mathbf{M}\xi_i>-M$ *для некоторого* $M>0$, *и существует* $\gamma>1$ *такое, что*

$$\sum_{1\le i<j\le n}\left(\mathbf{M}\xi_i\xi_j-\mu^2\right)=O\left(\frac{n^2}{(\ln n)(\ln\ln n)^{\gamma}}\right), \tag{9}$$

то $\eta_n/n\to\mu$ *п.н.*

С помощью следствия 1 можно получить следующий закон больших чисел для сильно зависимых стационарных последовательностей.

Теорема 3. *Пусть* $\{\xi_k, k \geq 1\}$ *– стационарная в широком смысле последовательность случайных величин,* $\mathbf{M}\xi_i = \mu$, $\mathbf{M}\xi_i^2 = \sigma^2 < \infty$ *и* $\eta_n = \sum\limits_{i=1}^{n} \xi_i$. *Если* $\mathbf{M}\xi_i \geq -M$ *для некоторого* $M > 0$, *и существует* $\gamma > 1$ *такое, что*

$$Cov\,(\xi_1, \xi_k) \leq K \frac{\sigma^2}{(\ln k)|\ln\ln k|^{\gamma}}, \;\; k = 2, ..., \tag{10}$$

где K *– некоторая константа, то* $\eta_n/n \to \mu$ *п.н.*

Доказательство. Воспользуемся следствием 1. Для доказательства теоремы 2 нам достаточно показать выполнения условия (9). Имеем

$$\sum_{1\leq i<j\leq n} \left(\mathbf{M}\xi_i\xi_j - \mu^2\right) = \sum_{1\leq i<j\leq n} (\mathbf{M}\xi_i\xi_j - \xi_i\xi_j) =$$

$$= \sum_{1\leq i<j\leq n} Cov\,(\xi_i, \xi_j) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov\,(\xi_i, \xi_j) = \sum_{i=1}^{n-1}\sum_{v=1}^{n-i} Cov\,(\xi_i, \xi_{i+v}).$$

Отсюда, используя неравенство Коши-Шварца, условие (10) и стационарность последовательности $\{\xi_k, k \geq 1\}$, находим

$$\sum_{1\leq i<j\leq n} \left(\mathbf{M}\xi_i\xi_j - \mu^2\right) = \sum_{i=1}^{n-1}\sum_{v=1}^{n-i} Cov\,(\xi_i, \xi_{i+v}) \leq (n-1)\sum_{v=1}^{n} |Cov\,(\xi_1, \xi_{v+1})| \leq$$

$$\leq K(n-1)\left(1 + \sum_{k=2}^{n} \frac{1}{(\ln k)|\ln\ln k|^{\gamma}}\right)\sigma^2. \tag{11}$$

Сумма внутри скобки имеет оценку

$$\sum_{k=2}^{n} \frac{1}{(\ln k)|\ln\ln k|^{\gamma}} = O\left(\frac{n}{(\ln n)(\ln\ln n)^{\gamma}}\right). \tag{12}$$

Действительно, пусть $Q = Q(n)$– некоторое целое положительное число, подбираемое ниже, такое, что $Q \to \infty$, $n \to \infty$. Тогда

$$\sum_{k=2}^{n} \frac{1}{(\ln k)|\ln\ln k|^{\gamma}} = \sum_{k=2}^{Q} \frac{1}{(\ln k)|\ln\ln k|^{\gamma}} + \sum_{k=Q+1}^{n} \frac{1}{(\ln k)(\ln\ln k)^{\gamma}} = \Sigma_1 + \Sigma_2.$$

Первое слагаемое полученной суммы Σ_1 имеет следующую очевидную оценку:

$$\Sigma_1 = \sum_{k=2}^{Q} \frac{1}{(\ln k)|\ln\ln k|^{\gamma}} \leq Q, \tag{13}$$

где $C = C_\gamma > 0$ – некоторая положительная постоянная. Так как функция $\ln x(\ln\ln x)^\gamma$ возрастает при $\gamma \geq 0$, то

$$\Sigma_2 = \sum_{k=Q+1}^{n} \frac{1}{(\ln k)|\ln\ln k|^\gamma} \leq \frac{n-k}{\ln Q(\ln\ln Q)^\gamma}. \tag{14}$$

Выберем теперь $Q = \dfrac{n}{\ln n(\ln\ln n)^\gamma}$. Тогда, учитывая неравенства (13) и (14), придем к оценке (12), так как при $n \to \infty$

$$\frac{\ln n(\ln\ln n)^\gamma}{\ln\frac{n}{\ln n}(\ln\ln n)^\gamma\left[\ln\ln n\frac{n}{\ln n}(\ln\ln n)^\gamma\right]^\gamma} = 1 + o(1).$$

Из неравенства (11) и оценки (12) следует, что условие (9) следствия 1 выполнено. Теорема 3 доказана.

С помощью следствия 1 и представления (3) можно получить следующий результат для линейных процессов.

Теорема 4. *Пусть $\{X_k,\ k = 1, 2, ...\}$ – линейный процесс, имеющий коэффициенты $\{a_{in}, i \in Z, n \geq 1\}$, удовлетворяющие условиям*

$$\sum_{k=0}^{\infty} (k+1)^2 \left(\frac{a_{k+1} + ... + a_{k+n}}{a(n)}\right)^2 \leq C_1 \frac{n}{\ln n(\ln\ln n)^\alpha}; \tag{15}$$

$$\frac{1}{a^2(n)} \sum_{j=0}^{n-1} a_{jn}^2 \leq C_2 \frac{1}{\ln n(\ln\ln n)^\beta}, \tag{16}$$

где C_1 и C_2 – постоянные, не зависящие от n, а $\alpha > 1, \beta > 1$ – произвольные вещественные числа. Для порождающей его инновационной последовательности $\{\xi_k, k \in Z, k \leq n\}$ с одинаково распределенными членами выполнены условия следствия 1. Тогда $\dfrac{1}{na(n)}\sum\limits_{k=1}^{n} X_k \xrightarrow{п.н.} \mu$.

Доказательство. БН-разложение для суммы линейного процесса $\{X_k\}$ имеет вид:

$$S_n = \sum_{k=1}^{n} X_k = a(n) \sum_{k=1}^{n} \xi_k + \left(\tilde{\xi}_0 - \tilde{\xi}_n\right),$$

где $\tilde{\xi}_t = \sum\limits_{j=0}^{\infty} \tilde{a}_{jn}\xi_{t-j}$. Отсюда $\mathbf{M}S_n = \mu n a(n)$. Положим $\xi_{kn} = \xi_k/n$, $k \in \mathbb{Z}, k \leq n$. Тогда

$$\frac{S_n}{na(n)} = \frac{1}{n}\sum_{k=1}^{n} \xi_k + \frac{\tilde{\xi}_0 - \tilde{\xi}_n}{na(n)}. \tag{17}$$

Первый член разложения (17) стремится к μ согласно следствия 1. Следовательно, для доказательства теоремы 5 нам достаточно доказать, что при $n \to \infty$

$$\frac{\tilde{\xi}_0 - \tilde{\xi}_n}{na(n)} \xrightarrow{\text{п.н.}} 0.$$

В действительности для доказательства этого утверждения достаточно показать сходимость ряда (см. [3], стр. 270)

$$P(\varepsilon) = \sum_{n=1}^{\infty} P\left(\left|\frac{\tilde{\xi}_0 - \tilde{\xi}_n}{na(n)}\right| > \varepsilon\right) < \infty$$

для каждого $\varepsilon > 0$.

По неравенству Чебышева

$$P(\varepsilon) \leq \sum_{n=1}^{\infty} \frac{\mathbf{D}\left(\tilde{\xi}_0 - \tilde{\xi}_n\right)}{\varepsilon^2 n^2 a^2(n)}. \tag{18}$$

Таким образом, для доказательства теоремы нужно доказать сходимость последнего ряда для каждого $\varepsilon > 0$.

Справедливо равенство

$$\tilde{\xi}_0 - \tilde{\xi}_n = \sum_{k=0}^{\infty} (a_{k+1,n} + ... + a_{k+n,n})\xi_{-k} - \sum_{j=1}^{n} a_{n-j,n}\xi_j.$$

Отсюда

$$\mathbf{D}(\tilde{\xi}_0 - \tilde{\xi}_n) \leq 2\mathbf{D}\left(\sum_{k=0}^{\infty} (a_{k+1,n} + ... + a_{k+n,n})\xi_{-k}\right) + 2\mathbf{D}\left(\sum_{j=1}^{n} a_{n-j,n}\xi_j\right) =$$

$$= 2\mathbf{M}\left[\sum_{k=0}^{\infty} (a_{k+1,n} + ... + a_{k+n,n})\left(\xi_{-k} - \mathbf{M}\xi_{-k}\right)\right]^2 + 2\mathbf{M}\left[\sum_{j=1}^{n} a_{n-j,n}(\xi_j - \mathbf{M}\xi_j)\right]^2.$$

Теперь, применяя неравенство Коши-Шварца, получим

$$\mathbf{D}\left(\tilde{\xi}_0 - \tilde{\xi}_n\right) \leq 2\sum_{k=0}^{\infty} k^2 (a_{k+1,n} + ... + a_{k+n,n})^2 \sum_{k=1}^{\infty} \frac{\mathbf{D}\xi_{-k}}{k^2} +$$

$$+2\left(\sum_{j=1}^{n} a_{n-j,n}^2\right) \sum_{j=1}^{n} \mathbf{D}\xi_j.$$

Используя неравенство (18), условия (15) и (16), оценим

$$\frac{\varepsilon^2}{2}P(\varepsilon) \le C_1 \sum_{n=1}^{\infty} \frac{1}{n \ln n (\ln\ln n)^{\alpha}} \left(\sum_{k=0}^{\infty} \frac{\mathbf{D}\xi_{-k}}{(k+1)^2} \right) +$$

$$+C_2 \sum_{n=1}^{\infty} \frac{1}{n \ln n (\ln\ln n)^{\beta}} \sum_{j=1}^{n} \frac{\mathbf{D}\xi_j}{n}. \quad (19)$$

Отсюда в силу одинаковой распределенности членов последовательности $\{\xi_k\}$ и сходимости рядов находящихся в правой части неравенства (19), следует сходимость ряда (18) при любом $\varepsilon > 0$. Теорема доказана.

Из теорем 3 и 4 непосредственно следует усиленный ЗБЧ для линейных процессов, порожденных сильно зависимыми инновационными последовательностями.

Следствие 2. *Пусть $\{X_k,\ k = 1, 2, ...\}$ – линейный процесс, имеющий коэффициенты $\{a_{in}, i \in Z, n \ge 1\}$, которые удовлетворяют условиям (15) и (16), порожденный стационарной инновационной последовательностью с одинаково распределенными членами $\{\xi_k, k \in Z, k \le n\}$, для которой выполнены условия теоремы 3. Тогда $\frac{1}{na(n)} \sum_{k=1}^{n} X_k \xrightarrow{п.н.} \mu$.*

Для изложения следующего результата, связанного с основной теоремой 1, нам потребуется понятие равномерного сильного перемешивания.

Определение 2. Пусть $\{\xi_{kn},\ k \in \mathbb{Z}\}$ – инновационная последовательность случайных величин в схеме серий. Через $\mathcal{M}_a^b$ обозначим σ-алгебру событий, порожденную случайными величинами $\xi_{an}, \xi_{a+1,n}, ..., \xi_{bn}$. Скажем, что последовательность $\{\xi_{kn},\ k \in \mathbb{Z}\}$ удовлетворяет условию равномерного сильного перемешивания, если

$$\varphi(k) = \sup \left| \frac{P(AB) - P(A)P(B)}{P(A)} \right| \to 0 \text{ при } \ k \to \infty,$$

где верхняя грань берется по всем $A \in \mathcal{M}_{-\infty}^{l},\ B \in \mathcal{M}_{l+k}^{\infty}$.

Пусть $\{\xi_{kn}, k \in \mathbb{Z}\}$ – последовательность серий случайных величин, удовлетворяющая условию равномерного сильного перемешивания. Положим $\eta_n = \xi_{1n} + ... + \xi_{nn},\ m_{kn} = \mathbf{M}\xi_{kn},\ \sigma_{kn}^2 = \mathbf{D}\xi_{kn} < \infty$ и предположим, что $\mathbf{D}\eta_n = 1$. Мы скажем, что последовательность $\{\xi_{kn},\ k \in \mathbb{Z}\}$ удовлетворяет условию Линдеберга, если для любого $\varepsilon > 0$

$$\text{(Л)} \quad \sum_{k=1}^{n} \int\limits_{|x - m_{kn}| > \varepsilon} (x - m_{kn})^2 dF_k(x) \xrightarrow[n \to \infty]{} 0,$$

где $F_k(x)$ – функция распределения случайной величины ξ_{kn}. Не ограничивая общности рассуждений, можно считать, что $m_{kn} = 0, k \in \mathbb{Z}$.

Теорема 5. *Пусть* $\{X_{kn},\ k = 1, 2, ...\}$ – *линейный процесс, имеющий коэффициенты* $\{a_{in}, i \in Z, n \geq 1\}$, *удовлетворяющие условию (B), порожденный инновационной последовательностью* $\{\xi_{in}, i \in Z\}$, *удовлетворяющей условию равномерного сильного перемешивания, для которой выполнены условия (A) и*

$$\sum_{k=1}^{\infty} \varphi^{1/2}(k) < \infty. \tag{20}$$

Тогда, для того чтобы линейный процесс удовлетворял центральной предельной теореме, т.е. имело место соотношение

$$P\left(\frac{1}{a(n)}\sum_{k=1}^{n} X_{kn} \leq x\right) \xrightarrow[n\to\infty]{} \Phi\left(x\right), \tag{21}$$

где $\Phi(x)$ – *функция распределения стандартной нормальной случайной величины, необходимо и достаточно выполнения условия Линдеберга (Л).*

Следствие 2. *Пусть* $\{X_{kn},\ k = 1, 2, ...\}$ – *линейный процесс, имеющий коэффициенты* $\{a_{in}, i \in Z, n \geq 1\}$, *которые удовлетворяют условию (B), порожденный последовательностью* $\{\xi_{in}, i \in Z\}$, *независимых в каждой серии случайных величин, удовлетворяющих условию (A). Тогда, для того чтобы линейный процесс удовлетворял центральной предельной теореме, необходимо и достаточно выполнения условия Линдеберга (Л).*

Доказательство теоремы 5. Достаточность. В работе [5] доказано, что при выполнении условия Линдеберга и (20), любая последовательность $\{\xi_{kn}\}$ с равномерным сильным перемешиванием подчиняется принципу инвариантности, отсюда как следствие получаем центральную предельную теорему. Из этого результата, согласно теореме 1, следует справедливость соотношения (21). Необходимость условий теоремы 6 непосредственно следует из известной теоремы Линдеберга–Феллера.

Библиографический список

1. *Beveridje S., Nelson C. R.* A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle"// Journal of monetary Economics. – 1981. – V. 7. – P. 151–174.

2. *Phillips P.C.B., Solo V.* Asymptotes for linear processes// Annals of statistics. – 1992. – V. 20. – P. 971–1001.

3. *Ширяев А.Н.* Вероятность. – М.: Наука, 1980. – 576 с.

4. *Nuno Luzia.* A Borel-Cantelli lemma and its applications// arXiv. 1201, 5866. – 15 p.

5. *Зупаров Т.М., Мухамедов А.К.* Принцип инвариантности для процессов с равномерно сильным перемешиванием// В сб.: Функционалы от случайных процессов и статистические выводы. – Ташкент: Фан, 1989. – С. 27–36.

6. *Зупаров Т.М.* Предельные теоремы для линейных процессов// Материалы респуб. Научно-практической конференции "Статистика и ее применения". – Ташкент, 2012. – С. 112–123.

7. *Peligrad M. and Utev S.* Central limit theorem for linear processes// Ann Probab. – 1997. – V. 25, №1. – P. 443–456.

8. *Peligrad M. and Utev S.* Central limit theorem for stationary linear processes// Ann Probab. – 2006. – V. 34, №4. – P. 1608–1626.

9. *Wu W.B.* Central limit theorem for functional of linear processes and their applications// Statist, Sinica. – 2002. – V. 12. – P.635–649.

10. *Wu W.B. and Min W.* On linear processes with dependent innovations// Stohastic Process. Appl. – 2005. – V. 115, №6. – P. 939–958.

Asymptotic analysis of the distribution of the sum of the linear process

Zuparov T.M.

National University of Uzbekistan named after Mirzo Ulugbek, faculty of mechanics and mathematics, department of Probability Theory and Mathematical Statistics, Uzbekistan, Tashkent; tm_zuparov@rambler.ru

Abstract. This paper is devoted to the asymptotic analysis of the distribution of the sum of the first n members of a linear process generated by the innovation sequence $\{\xi_{kn}\}$. Using the Beverage-Nelson decomposition (BN-decomposition) proved criterion of weak convergence of the normalized sum of a linear process to the limit distribution, as a consequence of this result, obtained an analogue of the Lindeberg-Feller theorem for linear processes generated by φ-mixing innovation sequence.

Key words: *linear process, BN-decomposition, almost surely convergence, Lindberg–Feller theorem.*

УДК 519.24

Об эмпирическом процессе независимости с оцениваемым параметром и его применении к модели пропорциональных интенсивностей

Л. Р. Какаджанова

Национальный университет Узбекистана имени Мирзо Улугбека, математический факультет, Узбекистан, Ташкент; leyla_tvms@rambler.ru; Узбекистан, 100174, г.Ташкент, ВУЗ городок, улица Университетская, 4, Национальный университет Узбекистана имени Мирзо Улугбека, кафедра "Теория вероятностей и математическая статистика"

Аннотация. *Приведены асимптотические результаты для эмпирического процесса независимости случайных элементов и событий, индексированных классом измеримых функций. Показано применение этих результатов к модели пропорциональных интенсивностей.*

Ключевые слова: *эмпирический процесс, независимость, метрическая энтропия, теоремы Гливенко-Кантелли и Донскера, пропорциональная интенсивность.*

Введение

Следуя [1–2], рассмотрим последовательность экспериментов, в которых наблюдаемые данные состоят из пар $\{(X_k, A_k), k \geq 1\}$, где $\{X_k\}$ – независимые и одинаково распределенные случайные элементы в измеримом пространстве $(\mathfrak{X}, \mathfrak{B})$ с вероятностным законом $\mathbb{P}$, а $\{A_k\}$ – события с общей вероятностью появления $p = \mathbb{P}(A_k) \in (0,1)$. Пусть $\delta_k = I(A_k)$ – индикатор события A_k. На n-м шаге эксперимента наблюдаются данные $\mathbb{S}^{(n)} = \{(X_k, \delta_k), 1 \leq k \leq n\}$. Каждая пара (X_k, δ_k) образует статистическую модель с выборочным пространством $\mathfrak{X} \otimes \{0,1\}$, σ-алгеброй $\mathcal{G}$ множеств $B \times D$ и распределением $\mathbb{Q}^*(\cdot)$ на $(\mathfrak{X} \otimes \{0,1\}, \mathcal{G})$:

$$\mathbb{Q}^*(B \times D) = \mathbb{P}(X_k \in B, \delta_k \in D), B \in \mathfrak{B}, D \subset \{0,1\}.$$

Наш интерес сфокусирован на субмерах $\mathbb{Q}_m(B) = \mathbb{Q}^*(B \times \{m\})$, $m = 0,1$ и $\mathbb{Q}(B) = \mathbb{Q}_0(B) + \mathbb{Q}_1(B) = \mathbb{Q}^*(B \times \{0,1\})$, $B \in \mathfrak{B}$. С практической точки зрения интерес представляет гипотеза $\mathcal{H}$ о независимости случайных элементов X_k и событий A_k для каждого $k \geq 1$. Для проверки этой гипотезы используем знакопеременную меру $\Lambda(B) = \mathbb{Q}_1(B) - p\mathbb{Q}(B)$, $B \in \mathfrak{B}$, где $p = \mathbb{Q}_1(\mathfrak{X})$. Тогда справедливость гипотезы $\mathcal{H}$ эквивалентна равенству $\Lambda(B) = 0$ для каждого $B \in \mathfrak{B}$. Чтобы построить статистику для проверки гипотезы $\mathcal{H}$, введем $\Lambda_n(B)$ – эмпирический аналог меры $\Lambda(B)$, полагая $\Lambda_n(B) = \mathbb{Q}_{1n}(B) - p_n\mathbb{Q}_n(B)$ для каждого $B \in \mathfrak{B}$. Здесь

$$p_n = \mathbb{Q}_{1n}(\mathfrak{X}), \mathbb{Q}_n(B) = \mathbb{Q}_{0n}(B) + \mathbb{Q}_{1n}(B),$$

$$\mathbb{Q}_{0n}(B) = \frac{1}{n}\sum_{k=1}^{n}(1-\delta_k)I(X_k \in B), \mathbb{Q}_{1n}(B) = \frac{1}{n}\sum_{k=1}^{n}\delta_k I(X_k \in B) \qquad (1)$$

– эмпирические аналоги p, $\mathbb{Q}(B)$ и $\mathbb{Q}_m(B), m = 0,1$, соответственно. Согласно усиленному закону больших чисел (УЗБЧ), для каждого $B \in \mathfrak{B}$ при $n \to \infty$, $\Lambda_n(B) \overset{\text{п.н.}}{\to} \Lambda(B)$ и, следовательно, при справедливости гипотезы $\mathcal{H}$, $\Lambda_n(B) \overset{\text{п.н.}}{\to} 0$. Тогда естественным становится исследование предельного поведения нормированного эмпирического процесса $\{\chi_n = a_n(\Lambda_n(B) - \Lambda(B)), B \in \mathcal{J}\}$, который индексирован некоторым классом $\mathcal{J}$ множеств $\mathfrak{B}$, где $\{a_n, n \geq 1\}$ (возможно случайная) последовательность положительных чисел. В предыдущих работах [1,2] мы исследовали специальный нормированный эмпирический процесс независимости, индексированный классом $\mathcal{F}$-измеримых функций f на $\mathfrak{X}$, который совпадает с χ_n, где $f = I(\cdot)$ является индикаторной функцией.

Эмпирический процесс независимости индексированный классом измеримых функций

Для знакопеременной меры G и класса $\mathcal{F}$ борелевских измеримых функций $f : \mathfrak{X} \to \mathbb{R}$ определим интеграл

$$Gf = \int\limits_{\mathfrak{X}} f dG, f \in \mathcal{F}.$$

Введем следующее $\mathcal{F}$-индексированное представление эмпирических мер (1) для $f \in \mathcal{F}$:

$$\mathbb{Q}_{0n} f = \frac{1}{n} \sum_{k=1}^{n} (1 - \delta_k) f(X_k), \ \ \mathbb{Q}_{1n} f = \frac{1}{n} \sum_{k=1}^{n} \delta_k f(X_k),$$

$$\mathbb{Q}_n f = \mathbb{Q}_{0n} f + \mathbb{Q}_{1n} f = \frac{1}{n} \sum_{k=1}^{n} f(X_k), \tag{2}$$

а также $\Lambda_n f = \mathbb{Q}_{1n} f - p_n \mathbb{Q}_n f$, где $p_n = \mathbb{Q}_{1n} 1 = \mathbb{Q}_{1n}(\mathfrak{X}) = \frac{1}{n} \sum\limits_{k=1}^{n} \delta_k$. Видно, что (1) является специальным классом (2), когда $\mathcal{F} = \{I(B), B \in \mathcal{J}\}$.

Введем $\mathcal{F}$-индексированный эмпирический процесс $G_n : \mathcal{F} \to \mathbb{R}$

$$f \mapsto G_n f = \sqrt{n}(\mathbb{Q}_n - \mathbb{Q}) f = n^{-1/2} \sum_{k=1}^{n} (f(X_k) - \mathbb{Q} f), f \in \mathcal{F}. \tag{3}$$

Здесь $G_n f = G_{0n} f + G_{1n} f$ с субэмпирическими процессами

$$G_{jn} f = \sqrt{n}(\mathbb{Q}_{jn} - \mathbb{Q}_j) f, \ \ j = 0, 1, f \in \mathcal{F}. \tag{4}$$

Для данной функции f по УЗБЧ и центральной предельной теореме (ЦПТ)

$$(a) \ \ \mathbb{Q}_n f \xrightarrow[n \to \infty]{\text{п.н.}} \mathbb{Q} f \ \text{ при } \ \mathbb{Q}|f| < \infty; \tag{5}$$

$$(b) \ \ G_n f \Rightarrow G f \stackrel{d}{=} N\left(0, \sigma_{\mathbb{Q}}^2(f)\right), \ \ n \to \infty \ \text{ при } \ \mathbb{Q} f^2 < \infty, \tag{6}$$

где $\sigma_{\mathbb{Q}}^2(f) = \mathbb{Q}(f - \mathbb{Q} f)^2$.

Равномерные варианты утверждений (5) и (6) для класса $\mathcal{F}$-измеримых функций, имеют очень солидную теорию (см., например, [3-10]). Имеются разные варианты классических теорем Гливенко–Кантелли и Донскера для $\mathcal{F}$-индексированных эмпирических процессов (3) при определенных условиях на множество $\mathcal{F}$ - измеримых функций. Эти условия обеспечивают сходимость $n^{-1/2} \|G_n f\|_{\mathcal{F}} = \sup\left\{n^{-1/2}|G_n f|, f \in \mathcal{F}\right\}$ к нулю, либо по вероятности, либо почти наверное. Такие классы $\mathcal{F}$ называются слабыми или сильными классами Гливенко–Кантелли, соответственно. Теоремы типа Донскера обеспечивают обобщенные условия на класс $\mathcal{F}$, при которых имеет место сходимость

$$G_n f \Rightarrow G f \quad \text{в} \quad l^{\infty}(\mathcal{F}), \tag{7}$$

где $l^{\infty}(\mathcal{F})$ – пространство всех ограниченных функций $f : \mathfrak{X} \to \mathbb{R}$ с супремум нормой $\|.\|_{\mathcal{F}}$ (см., [8], с.81). Класс $\mathcal{F}$, для которого имеет место сходимость (7), называется классом Донскера. Предельное поле $\{Gf, \ f \in \mathcal{F}\}$

в (7) называется $\mathbb{Q}$-броуновским мостом. Оно является плотным борелевским элементом $l^{\infty}(\mathcal{F})$ и является гауссовским полем с нулевым средним и ковариационной функцией

$$\operatorname{cov}(Gf, Gg) = \mathbb{Q}fg - \mathbb{Q}f\mathbb{Q}g, \quad f, g \in \mathcal{F}. \tag{8}$$

Таким образом, $\mathbb{Q}$-броуновский мост $\{Gf, f \in \mathcal{F}\}$ может быть представлен по распределению через $\mathbb{Q}$-броуновский лист $\{W(f), f \in \mathcal{F}\}$ с нулевым средним и ковариацией

$$\operatorname{cov}(W(f), W(g)) = \mathbb{Q}fg, \quad f, g \in \mathcal{F}, \tag{9}$$

с помощью равенства

$$Gf \overset{d}{=} W(f) - W(1)\mathbb{Q}f, \ \ f \in \mathcal{F}. \tag{10}$$

По УЗБЧ при условиях $\mathbb{Q}_j |f| < \infty, \ j = 0, 1$ для данной функции f:

$$\Lambda_n f \underset{n\to\infty}{\overset{\text{п.н.}}{\to}} \Lambda f \overset{\text{при } \mathcal{H}}{=} 0. \tag{11}$$

Более того, для заданного f, величина $\sqrt{n}(\Lambda_n - \Lambda) f$ будучи линейным функционалом субэмпирического процесса (4) при условиях $\mathbb{Q}_j f^2 < \infty, \ j = 0, 1$, в пределе имеет нормальное распределение $N\left(0, \sigma^2_{\mathbb{Q}}(f)\right)$. В [1] авторы доказали равномерные УЗБЧ и ЦПТ для специальных нормированных эмпирических $\mathcal{F}$-индексированных процессов

$$\left\{\Delta_n f = \left(\frac{n}{p_n(1-p_n)}\right)^{1/2} (\Lambda_n - \Lambda) f, \ \ f \in \mathcal{F}\right\}, \tag{12}$$

и показали, что предельное распределение совпадает с $\mathbb{Q}$-броуновским мостом $\{Gf, \ \ f \in \mathcal{F}\}$ с ковариацией (8).

Асимптотические результаты

Пусть $\mathcal{L}_q(\mathbb{Q})$ – пространство функций $f : \mathfrak{X} \to \mathbb{R}$ с нормой

$$\|f\|_{\mathbb{Q},q} = (\mathbb{Q}|f|^q)^{1/q} = \left\{\int\limits_{\mathfrak{X}} |f|^q d\mathbb{Q}\right\}^{1/q}.$$

Для формулировки $\mathcal{F}$-равномерных вариантов теорем Гливенко–Кантелли и Донскера определим сложность или энтропию класса $\mathcal{F}$. Для этого необходимо ввести понятие ε-скобки. ε-скобка в $\mathcal{L}_q(\mathbb{Q})$ – это пара функций

$\varphi, \psi \in \mathcal{L}_q(\mathbb{Q})$ таких, что $\mathbb{Q}\left(\varphi(X) \le \psi(X)\right) = 1$ и $\|\psi - \varphi\|_{\mathbb{Q},q} \le \varepsilon$, т.е. $\mathbb{Q}(\psi - \varphi)^q \le \varepsilon^q$. Функция $f \in \mathcal{F}$ в скобке (или покрывается скобкой) $[\varphi, \psi]$, если $\mathbb{Q}\left(\varphi(X) \le f(X) \le \psi(X)\right) = 1$. Таким образом, φ и ψ могут и не принадлежать классу $\mathcal{F}$, но они должны иметь конечные нормы. Числом скобок (или покрытий) $N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right)$ называется минимальное число ε-скобок в $\mathcal{L}_q(\mathbb{Q})$ необходимых для покрытия $\mathcal{F}$ (см. [8,9]):

$$N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right) = \min \begin{cases} k: \text{ для некоторых } f_1, ..., f_k \in \mathcal{L}_q\left(\mathbb{Q}\right), \\ \quad \mathcal{F} \subset \bigcup\limits_{i,j} \left[f_i, f_j\right] : \|f_j - f_i\|_{\mathbb{Q},q} \le \varepsilon. \end{cases}$$

Число $H_q\left(\varepsilon\right) = \log N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}\right)\right)$ называется метрической энтропией со скобкой класса $\mathcal{F}$ в $\mathcal{L}_q(\mathbb{Q})$. Числа $H_{jq}\left(\varepsilon\right) = \log N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_q\left(\mathbb{Q}_j\right)\right), j = 0, 1$ означают метрические энтропии класса $\mathcal{F}$ в $\mathcal{L}_q(\mathbb{Q}_j)$, $j = 0, 1$, соответственно. Для слабой сходимости $\mathcal{F}$-индексированного эмпирического процесса (12) используется интеграл от метрической энтропии со скобкой

$$J_{j[]}^{(q)}\left(\delta\right) = J_{j[]}\left(\delta; \mathcal{F}; \mathcal{L}_q\left(\mathbb{Q}_j\right)\right) = \int\limits_0^\delta \left(H_{jq}\left(\varepsilon\right)\right)^{1/2} d\varepsilon, j = 0, 1, \text{ for } 0 < \delta < 1.$$

Напомним, что важными свойствами чисел $N_{[]}\left(\cdot\right)$ является то, что они стремятся к $+\infty$ при $\varepsilon \downarrow 0$. Однако, для теоремы Донскера необходимо, чтобы они стремились не очень быстро к $+\infty$. Эта скорость измеряется через интеграл-скобки $J_{j[]}^{(q)}\left(\delta\right)$ (см. [8,9]).

В следующей теореме показана справедливость теоремы типа Гливенко–Кантелли для процесса $\{\Delta_n f, \ f \in \mathcal{F}\}$. Здесь $*$ – означает сходимость п.н. по внешней вероятности.

Теорема 1 [1]. *Пусть класс $\mathcal{F}$ таков, что*

$$N_{[]}\left(\varepsilon, \mathcal{F}, \mathcal{L}_1\left(\mathbb{Q}_j\right)\right) < \infty, j = 0, 1. \tag{13}$$

Тогда при справедливости гипотезы $\mathcal{H}$ при $n \to \infty$

$$\left\| n^{-1/2} \Delta_n f \right\|_{\mathcal{F}}^{*} \overset{п.н.}{\to} 0. \tag{14}$$

Для установления слабой сходимости процесса (12) к гауссовскому процессу сначала исследуются предельные свойства двумерного эмпирического поля $\{(\mathbb{A}_n f, \mathbb{A}_{1n} g)\ , \ f, g \in \mathcal{F}\}$, где $\mathbb{A}_n f = n^{1/2}\left(\mathbb{Q}_n - \mathbb{Q}\right) f$ и $\mathbb{A}_{1n} g = n^{1/2}\left(\mathbb{Q}_{1n} - \mathbb{Q}_1\right) g$.

Теорема 2 [1]. *Пусть класс* $\mathcal{F}$ *таков, что*

$$\mathcal{F} \subset \mathcal{L}_2(\mathbb{Q}_j) \text{ и } J^{(2)}_{j[\,]}(1) < \infty, \ j = 0, 1. \tag{15}$$

Тогда при $n \to \infty$ *последовательность* $\{(\mathbb{A}_n f, \mathbb{A}_{1n} g)\ ,\ f, g \in \mathcal{F}\}$ *отображений* $\mathcal{F} \to \mathbb{R}^2$ *слабо сходится в* $l^\infty(\mathcal{F}) \times l^\infty(\mathcal{F})$ *к двумерному гауссовскому полю* $\{(\mathbb{A} f, \mathbb{A}_1 g)\ ,\ f, g \in \mathcal{F}\}$ *с нулевым средним и ковариационной структурой для* $f, g \in \mathcal{F}$*:*

$$\begin{aligned} \mathbb{E}\,(\mathbb{A} f \cdot \mathbb{A} g) &= \mathbb{Q} f g - \mathbb{Q} f \mathbb{Q} g, \\ \mathbb{E}\,(\mathbb{A}_1 f \cdot \mathbb{A}_1 g) &= \mathbb{Q}_1 f g - \mathbb{Q}_1 f \mathbb{Q}_1 g, \\ \mathbb{E}\,(\mathbb{A} f \cdot \mathbb{A}_1 g) &= \mathbb{Q}_1 f g - \mathbb{Q} f \mathbb{Q}_1 g. \end{aligned} \tag{16}$$

Замечание. *В формуле (16) при* $g \equiv 1$ *имеем* $\mathbb{Q}_1 1 = p$ *и*

$$\mathbb{E}\,(\mathbb{A} f \cdot \mathbb{A}_1 1) = \mathbb{Q}_1 f - p\,\mathbb{Q} f,\ f \in \mathcal{F}. \tag{17}$$

Отсюда при справедливости гипотезы $\mathcal{H}$ ковариация (17) равна нулю для всех $f \in \mathcal{F}$. Таким образом, при справедливости гипотезы $\mathcal{H}$ броуновский мост $\{\mathbb{A}\, f,\ f \in \mathcal{F}\}$ и с.в. $\mu_0 = \mathbb{A}_1 1$ с нормальным распределением $\mathcal{N}\,(0, p\,(1-p))$ независимы.

При исследовании процесса (12) основное свойство слабой сходимости к $\mathbb{Q}$-броуновскому мосту обеспечивается следующим утверждением.

Теорема 3 [1]. *В условиях теоремы 2 для* $n \to \infty$

$$\Delta_n f \Rightarrow \Delta f \text{ в } l^\infty(\mathcal{F}), \tag{18}$$

где $\{\Delta f,\ f \in \mathcal{F}\}$ *– гауссовское поле с нулевым средним, которое при справедливости гипотезы* $\mathcal{H}$ *совпадает по распределению с* $\mathbb{Q}$*-броуновским мостом.*

Теперь рассмотрим сходимость эмпирического процесса независимости (12), когда класс $\mathcal{F}$-измеримых функций $f_{\theta,\eta} : \mathfrak{X} \to \mathbb{R}$ индексируется множествами Θ и $\mathcal{K}$: $\mathcal{F} = \{f_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{K}\}$, где η является оцениваемым параметром. Пусть η_n – оценка для η, являющаяся случайным элементом со значениями в $\mathcal{K}$, определенными на том же вероятностном пространстве, что и $X_1, ..., X_n$, а $\eta_0 \in \mathcal{K}$ – фиксированный элемент, который является пределом по вероятности последовательности η_n. Для некоторых приложений интерес представляет следующий результат: при $n \to \infty$

$$\sup_{\theta \in \Theta} |\Delta_n (f_{\theta,\eta_n} - f_{\theta,\eta_0})| \xrightarrow{p} 0. \tag{19}$$

Этот результат может быть применен к оцениванию функционалов $\theta \mapsto \Delta f_{\theta,\eta}$. При неизвестном η мы сможем решить эту задачу, используя оценку $\Delta_n f_{\theta,\eta_n}$.

Будем говорить, что η_n является состоятельной оценкой для η_0, если

$$\sup_{\theta\in\Theta}\left|\Delta_n(f_{\theta,\eta_n}-f_{\theta,\eta_0})^2\right| \xrightarrow[n\to\infty]{p} 0. \qquad (20)$$

Теорема 4. *Если условия (15) и (20) выполнены, то (19) имеет место.*

Утверждение теоремы можно доказать, используя разложение

$$\Delta_n f_{\theta,\eta_n} - \Delta f_{\theta,\eta_0} = \qquad (21)$$
$$= \Delta_n(f_{\theta,\eta_n}-f_{\theta,\eta_0}) + \Delta_n f_{\theta,\eta_0} + \sqrt{\frac{n}{p_n(1-p_n)}}\Delta_n(f_{\theta,\eta_n}-f_{\theta,\eta_0}).$$

Первое слагаемое разложения (21) стремится к нулю ввиду (19), второе слагаемое стремится к гауссовскому процессу по теореме 3, а третье слагаемое стремится к нулю согласно (20).

Применение к модели пропорциональных интенсивностей

Рассмотрим модель случайного цензурирования справа, где $X_i = \min\{T_i, C_i\}$ и $A_i = \{T_i \le C_i\}$. Здесь с.в. T_i и C_i – продолжительность жизни и время цензурирования i-го элемента; $\{T_i\}$ и $\{G_i\}$ – независимые с.в. с общими непрерывными функциями распределения F и G соответственно, $F(0) = G(0) = 0$. Имеются данные $\mathbb{S}^{(n)} = \{(X_i, \delta_i), 1 \le i \le n\}$ с $\delta_i = I(A_i)$. Интересующие нас с.в. T_i наблюдаются, когда A_i происходят, т.е. при $\delta_i = 1$. Заметим, что $\{X_i\}$ имеют общую функцию распределения $H = 1 - (1-F)(1-G)$ и субраспределения определены формулами

$$\mathbb{Q}_0(B) = \mathbb{P}(X_k \in B, \delta_k = 0) = \mathbb{P}(C_k \in B \cap [0, T_k)) \int_B (1 - F(t))G(dt),$$

$$\mathbb{Q}_1(B) = \mathbb{P}(X_k \in B, \delta_k = 1) = \mathbb{P}(T_k \in B \cap [0, C_k]) \int_B (1 - G(t))F(dt). \qquad (22)$$

Теперь рассмотрим модель пропорциональных интенсивностей (МПИ) или модель Козиола–Грина, которая считается в статистической литературе полезной и интересной с практической точки зрения (см., например, [1]). В МПИ предполагается наличие параметрической связи вида

$$1 - G = (1 - F)^{\beta} \text{ для некоторого } \beta > 0. \tag{23}$$

Применяя (23), получим равенство $1 - F = (1 - H)^p$, где $p = \frac{1}{1+\beta} = \mathbb{P}(A_k)$. Одним из базовых свойств МПИ является то, что (23) выполняется, лишь в случае, когда X_k и δ_k независимы. Такое свойство МПИ играет базовую роль в построении и исследовании оценок многих функционалов распределения F. Следующая достаточная оценка максимального правдоподобия для F была впервые введена и изучена Абдушукуровым (1984) в МПИ:

$$F_n(t) = 1 - (1 - H_n(t))^{p_n}, \tag{24}$$

где $H_n(t) = \frac{1}{n}\sum\limits_{k=1}^{n} I(X_k \leq t)$ и $p_n = \frac{1}{n}\sum\limits_{k=1}^{n} \delta_k$ – независимые эмпирические оценки $H(t)$ и p соответственно.

Достаточно много работ посвящено статистическому анализу F_n. В них подчеркивается превосходство методов оценивания и проверки гипотез в МПИ, основанных на F_n, по сравнению с множительной оценкой Каплана-Мейера. Следовательно, возникает вопрос о проверке справедливости МПИ, т.е. сложной гипотезы, описываемой соотношением (23). Согласно характеризации МПИ это соотношение эквивалентно гипотезе $\mathcal{H}$ о независимости подвыборок $(X_1, ..., X_n)$ и $(\delta_1, ..., \delta_n)$. Для проверки этой гипотезы рассмотрим следующий вариант процесса (12), когда $f = I(\cdot)$:

$$\Delta_n(t) = \left(\frac{n}{p_n\,(1 - p_n)}\right)^{1/2} (H_{1n}(t) - p_n H_n(t)), \; -\infty < t < \infty, \tag{25}$$

где $H_{1n}(t) = \frac{1}{n}\sum\limits_{k=1}^{n} I(X_k \leq t, \delta_k = 1)$. Тогда справедливо следующее следствие из теоремы 3: если $\mathcal{H}$ выполнено, то при $n \to \infty$

$$\Delta_n(\cdot) \Rightarrow \mathbb{B}(H(\cdot)), \tag{26}$$

где $\{\mathbb{B}(y), 0 \leq y \leq 1\}$ – броуновский мост. Некоторые статистики для проверки гипотезы $\mathcal{H}$ могут быть построены на основе сходимости (26).

Библиографический список

1. *Abdushukurov A. A., Kakadjanova L. R.* A class of special empirical process of independence //Journal of Siberian Federal University. Mathematics & Physics. – 2015. – V. 8, №2. – P. 1–9.

2. *Abdushukurov A. A., Kakadjanova L. R.* Sequential empirical process of independence //Journal of Siberian Federal University. Mathematics & Physics. – 2018. – V. 11, №5. – P. 634–643.

3. *Alexander K. S.* Probability inequalities for empirical processes and a law of the iterated logarithm // Ann. Probab. – 1984. – V. 12, №4. – P. 1041–1067.

4. *Dudley R. M.* Central limit theorems for empirical measures // Ann. Probab. – 1978. – V. 6. – P. 899–929.

5. *Gaensler P., Stute W.* Empirical processes: a survey of results for independent and identically distributed random variables // Ann. Probab. – 1979. – V. 7, №2. – P. 193–243.

6. *Gine E., Zinn J.* Some limit theorems for empirical processes// Ann. Probab. – 1984. – V. 12, №4. – P. 929–989.

7. *Shorack G. R., Wellner J. A.* Empirical processes with applications to statistics. – John Wiley&Sons, 1986.

8. *Van der Vaart A. W., Wellner J. A.* Weak convergence and empirical processes. – Springer, 1986.

9. *Van der Vaart A. W.* Asymptotic Statistics. – Cambridge University Press: Cambridge, 1998.

10. *Van der Vaart A. W., Wellner J. A.* IMS Lecture Notes-Monograph Series. Asymptotics: Particless, Processes and Inverse Problems. – 2007. – V. 55 – P. 234–252.

On empirical process of independence with estimated parameter and its application to testing of proportional hazards model.

L. R. Kakadjanova

National University of Uzbekistan named after Mirzo Ulugbek, Faculty of Mathematics, Department of Probability Theory and Mathematical Statistics, Uzbekistan, Tashkent; leyla_tvms@rambler.ru

Abstract. *The asymptotical results of empirical processes of independence of random elements and events indexed by class of measurable functions are considered. Showed their application to proportional hazards model.*

Key words: *empirical process, independence, metric entropy, Glivenko-Cantelly and Donsker theorems, proportional hazards.*

УДК 519.2

О «равновесном распределении» как интегральной функции концентрации[1]

В.Ю. Королев

Факультет вычислительной математики и кибернетики МГУ имени М. В. Ломоносова; 119991, Москва, ГСП-1, Ленинские горы, МГУ, 2-й уч. корп., ф-т ВМК; Институт проблем информатики ФИЦ «Информатика и управление» РАН; vkorolev@cs.msu.ru

Аннотация. *Замечено, что хорошо известное «равновесное распределение» может служить в качестве интегральной функции концентрации. Описаны простейшие свойства интегральной функции концентрации, в том числе ее применение в уточненном варианте неравенства Маркова. Рассмотрена связь интегральной и «обычной» функций концентрации. Приведен обзор свойств итераций интегральной функции концентрации, в частности, указаны соотношения для их моментов и преобразований Лапласа–Стильтьеса. Доказана предельная теорема о сходимости таких итераций к показательному распределению.*

Ключевые слова: *функция концентрации; интегральная функция концентрации; равновесное распределение; преобразование Лапласа–Стильтьеса; показательное распределение; неподвижная точка.*

1. Определение интегральной функции концентрации

Пусть ξ – случайная величина с функцией распределения $F(x) = \mathsf{P}(\xi < x)$. Функцией концентрации случайной величины ξ принято называть функцию $Q_\xi(x)$ неотрицательного аргумента x, определяемую соотношением

$$Q_\xi(x) = \sup_z \mathsf{P}(z \le \xi \le z + x), \tag{1}$$

см., например, [4, 6]. Если слегка изменить определение функции концентрации и вместо (1) положить

$$Q^*_\xi(x) = \sup_z \mathsf{P}(z \le \xi < z + x), \quad x \ge 0,$$

то несложно заметить, что

[1]Работа выполнена при поддержке РФФИ (проект 18-07-01405).

$$Q_\xi^*(x) = \sup_z \big[F(z+x) - F(z)\big] = \rho(F_x, F), \tag{2}$$

где $\rho(F, G)$ – равномерное расстояние (расстояние Колмогорова) между функциями распределения F и G, а $F_x(z) = F(z+x)$, $z \in \mathbb{R}$, $x \geq 0$. Можно убедиться, что функция $Q_\xi^*(x)$ обладает свойствами функции распределения (некоторой неотрицательной случайной величины).

Заменим в (2) метрику ρ на какую-нибудь другую вероятностную метрику. Во многих задачах удобно иметь дело с расстоянием $\zeta_1 = L_1$. Однако, если попытаться определить аналог функции концентрации как

$$Q_\xi^\star(x) = L_1(F_x, F) = \int_{-\infty}^{\infty} \big[F(z+x) - F(z)\big] dz =$$

$$= \int_{-\infty}^{\infty} \mathsf{P}(z \leq \xi < z + x) dz, \quad x \geq 0, \tag{3}$$

то несложно убедиться, что $Q_\xi^\star(x) = L_1(F_x, F) \equiv x$ независимо от вида функции распределения F.

Тем не менее, если ограничиться рассмотрением только неотрицательных случайных величин ξ с конечным математическим ожиданием, отличным от нуля, то функции $Q_\xi^\star(x)$ можно придать свойства функции распределения, заменив в (3) расстояние L_1 его нормированным аналогом и получив в итоге функцию

$$Q_\xi^{(1)}(x) = \frac{1}{\mathsf{E}\xi} \int_0^{\infty} \big[F(z+x) - F(z)\big] dz, \quad x \geq 0.$$

При этом несложно видеть, что, так как

$$\mathsf{E}\xi = \int_0^{\infty} \big[1 - F(z)\big] dz,$$

то

$$Q_\xi^{(1)}(x) = \frac{1}{\mathsf{E}\xi} \int_0^{\infty} \big[F(z+x) - 1 + 1 - F(z)\big] dz =$$

$$= 1 - \frac{1}{\mathsf{E}\xi} \int_0^{\infty} \big[1 - F(z+x)\big] dz = 1 - \frac{1}{\mathsf{E}\xi} \int_x^{\infty} \big[1 - F(z)\big] dz =$$

$$= 1 - \frac{1}{\mathsf{E}\xi} \int_0^{\infty} \big[1 - F(z)\big] dz + \frac{1}{\mathsf{E}\xi} \int_0^{x} \big[1 - F(z)\big] dz = \frac{1}{\mathsf{E}\xi} \int_0^{x} \big[1 - F(z)\big] dz. \tag{4}$$

Функцию, стоящую в правой части (4), назовем *интегральной функцией концентрации*. В теории процессов восстановления функция $Q_\xi^{(1)}(x)$, $x \geq 0$, известна под названием *стационарного распределения запаздывания* [2]. В теории надежности эту функцию называют *равновесным распределением* (equilibrium distribution) [7]. Она также фигурирует в хорошо известной в теории массового обслуживания и актуарной математике формуле Поллачека–Хинчина–Беекмана [3]. В актуарной математике функция $Q_\xi^{(1)}(x)$ также называется *распределением интегрированного хвоста* (integrated tail distribution) [8].

Интерпретация функции $Q_\xi^{(1)}(x)$, $x \geq 0$, как интегральной функции концентрации возможна и для случайных величин χ, принимающих значения обоих знаков. С этой целью в качестве неотрицательной случайной величины ξ можно рассматривать, например, интегральную концентрацию неотрицательной части случайной величины χ, то есть $\xi = \chi^+ \equiv \max\{0, \chi\}$, интегральную концентрацию неположительной части случайной величины χ, то есть $\xi = \chi^- \equiv -\min\{0, \chi\}$, или же интегральную абсолютную концентрацию случайной величины χ, то есть $\xi \equiv |\chi| = \chi^+ + \chi^-$. В последнем случае имеем

$$Q_\xi^{(1)}(x) = \int_0^\infty \mathsf{P}(z \leq |\chi| < z + x)dz, \quad x \geq 0.$$

2. Элементарные свойства интегральной функции концентрации и ее интерпретация

2.1. Интегральная функция концентрации функции распределения сама является функцией распределения, более того, независимо от типа функции распределения F интегральная функция концентрации является абсолютно непрерывной функцией распределения. Как легко видеть, $Q_\xi^{(1)}(0) = 0$, и функции распределения $Q_\xi^{(1)}(x)$ соответствует плотность

$$q^{(1)}(x) = \frac{1 - F(x)}{\mathsf{E}\xi}, \quad x \geq 0.$$

2.2. Интегральная функция концентрации задает оператор A, действующий из множества функций распределения, все точки роста которых сосредоточены на неотрицательной полупрямой, в множество функций распределения, все точки роста которых сосредоточены на неотрицательной полупрямой. Обратный оператор A^{-1} определен на множестве абсолютно непрерывных функций распределения с монотонной плотностью, конечной в нуле.

2.3. Пусть $a > 0$. Имеем

$$Q_{a\xi}^{(1)}(x) = \frac{1}{a\mathsf{E}\xi}\int_0^x \Big[1 - F\Big(\frac{z}{a}\Big)\Big]dz = \frac{1}{\mathsf{E}\xi}\int_0^{\frac{x}{a}} \big[1 - F(z)\big]dz = Q_\xi^{(1)}\Big(\frac{x}{a}\Big).$$

Пусть $b > 0$. Имеем

$$Q^{(1)}_{\xi+b}(x) = \frac{1}{\mathsf{E}\xi + b}\int_0^x \big[1 - F(z-b)\big]dz = \frac{1}{\mathsf{E}\xi + b}\int_0^{\max\{0,x-b\}} \big[1 - F(z)\big]dz =$$

$$= \frac{\mathsf{E}\xi}{\mathsf{E}\xi + b}\cdot Q^{(1)}_{\xi}\big(\max\{0, x-b\}\big).$$

2.4. Пусть $F, F_1, F_2, \ldots$ – функции распределения неотрицательных случайных величин $\xi, \xi_1, \xi_2, \ldots$ соответственно, причем $0 < \mathsf{E}\xi = \mathsf{E}\xi_1 = \mathsf{E}\xi_2 = \ldots < \infty$. Предположим, что

$$L_1(F_n,\, F) \longrightarrow 0 \tag{5}$$

при $n \to \infty$. Тогда при любом $x \geq 0$ мы имеем

$$\big|Q^{(1)}_{\xi_n}(x) - Q^{(1)}_{\xi}(x)\big| = \left|\frac{1}{\mathsf{E}\xi}\int_0^x \big[F(x) - F_n(x)\big]dx\right| \leq$$

$$\leq \frac{1}{\mathsf{E}\xi}\int_0^x \big|F(x) - F_n(x)\big|dx \leq \frac{1}{\mathsf{E}\xi}\int_0^\infty \big|F(x) - F_n(x)\big|dx \longrightarrow 0.$$

При этом из абсолютной непрерывности функции распределения $Q^{(1)}_{\xi}(x)$ вытекает, что условие (5) влечет (см., например, Лемму 2.10.1 на с. 101-102 в [1])

$$\lim_{n\to\infty}\sup_{x\geq 0}\big|Q^{(1)}_{\xi_n}(x) - Q^{(1)}_{\xi}(x)\big| = 0,$$

то есть сходимость функций распределения $F_n(x)$ к функции распределения $F(x)$ в метрике $\zeta_1 = L_1$ влечет равномерную сходимость соответствующих интегральных функций концентрации.

2.5. Предположим, что $\mathsf{E}\xi^2 < \infty$. Если η_1 – случайная величина с функцией распределения $Q^{(1)}_{\xi}(x)$, то

$$\mathsf{E}\eta_1 = \int_0^\infty x q^{(1)}(x)dx = \frac{1}{\mathsf{E}\xi}\int_0^\infty x\big[1 - F(x)\big]dx = \frac{1}{2\mathsf{E}\xi}\int_0^\infty \big[1 - F(x)\big]dx^2 =$$

$$= \frac{x^2\big[1 - F(x)\big]}{2\mathsf{E}\xi}\bigg|_{x=0}^{\infty} - \frac{1}{2\mathsf{E}\xi}\int_0^\infty x^2 d\big[1 - F(x)\big] = \frac{1}{2\mathsf{E}\xi}\int_0^\infty x^2 dF(x) = \frac{\mathsf{E}\xi^2}{2\mathsf{E}\xi}.$$

Другими словами, средняя интегральная концентрация случайной величины ξ равна половине отношения ее второго и первого моментов. При этом по неравенству Иенсена $\mathsf{E}\xi \leq \sqrt{\mathsf{E}\xi^2}$, так что

$$\mathsf{E}\eta_1 \geq \tfrac{1}{2}\sqrt{\mathsf{E}\xi^2} \geq \tfrac{1}{2}\mathsf{E}\xi.$$

Таким образом, среднее значение интегральной концентрации определяется моментом *второго* порядка исходной случайной величины. Более того,

$$\mathsf{E}\eta_1 = \frac{1}{2}\Big(\frac{\mathsf{D}\xi}{\mathsf{E}\xi} + \mathsf{E}\xi\Big).$$

В частности, для нормированной случайной величины ξ с $\mathsf{E}\xi = 1$

$$\mathsf{E}\eta_1 = \tfrac{1}{2}\big(\mathsf{D}\xi + 1\big).$$

Таким образом, возможна следующая интерпретация интегральной функции концентрации $Q_\xi^{(1)}(x)$: если дисперсия является *числовой* характеристикой разброса исходной случайной величины ξ, то интегральная функция концентрации $Q_\xi^{(1)}(x)$ является *функциональной* характеристикой разброса.

3. Элементарные неравенства для интегральной функции концентрации. Уточнение неравенства Маркова

3.1. Рассмотрим следующее уточнение неравенства Маркова

$$\mathsf{P}(\xi \geq x) \leq \frac{\mathsf{E}\xi}{x},$$

справедливого при любом $x > 0$ для неотрицательной случайной величины ξ с конечным математическим ожиданием.

Для любого $x \in \mathbb{R}$ имеем

$$Q_\xi^{(1)}(x) = \frac{1}{\mathsf{E}\xi}\int_0^x \big[1 - F(z)\big]dz \geq \frac{1}{\mathsf{E}\xi}\int_0^x \big[1 - F(x)\big]dz =$$

$$= \frac{\big[1 - F(x)\big]}{\mathsf{E}\xi}\int_0^x dx = \frac{x\big[1 - F(x)\big]}{\mathsf{E}\xi} = \mathsf{P}(\xi \geq x)\frac{x}{\mathsf{E}\xi}.$$

Отсюда

$$\mathsf{P}(\xi \geq x) \leq Q_\xi^{(1)}(x) \cdot \frac{\mathsf{E}\xi}{x}. \tag{6}$$

Так как $0 \leq Q_\xi^{(1)}(x) \leq 1$, то правая часть (6) меньше, чем правая часть неравенства Маркова. При этом, очевидно, правая часть (6) равна $\frac{1}{x}\int_0^x \big[1 - F(z)\big]dz$.

Пусть $0 \leq h \leq x$. Тогда

$$Q_\xi^{(1)}(x) = \frac{1}{\mathsf{E}\xi}\int_0^x \big[1 - F(z)\big]dz =$$

$$= \frac{1}{\mathsf{E}\xi}\int_0^h \big[1-F(z)\big]dz + \frac{1}{\mathsf{E}\xi}\int_h^x \big[1-F(z)\big]dz \le$$

$$\le \frac{h+(x-h)\big(1-F(h)\big)}{\mathsf{E}\xi} = \frac{x-xF(h)+hF(h)}{\mathsf{E}\xi} = \frac{x\big[1-(1-\frac{h}{x})F(h)\big]}{\mathsf{E}\xi}.$$

Таким образом, для любого $x>0$ и любого $\delta\in(0,1)$ имеет место следующая двусторонняя оценка функции $Q_\xi^{(1)}(x)$:

$$\frac{x\big(1-F(x)\big)}{\mathsf{E}\xi} \le Q_\xi^{(1)}(x) \le \frac{x\big[1-(1-\delta)F(\delta x)\big]}{\mathsf{E}\xi}.$$

Из этой оценки, очевидно, получаем двойное неравенство

$$1-\frac{\mathsf{E}\xi}{x}\cdot Q_\xi^{(1)}(x) \le F(x) \le \frac{1}{1-\delta}\Big[1-\frac{\delta\mathsf{E}\xi}{x}\cdot Q_\xi^{(1)}\Big(\frac{x}{\delta}\Big)\Big].$$

3.2. Найдем связь интегральной функции концентрации $Q_\xi^{(1)}(h)$ с (подправленной) "обычной" функцией концентрации $Q_\xi(h)$ при условии $F(0)=0$. Для $z>0$ справедливо соотношение

$$Q_\xi(z) = \sup_{x\ge 0}\big[F(x+z)-F(x)\big] \ge F(z)-F(0)=F(z),$$

откуда $F(z)\le Q_\xi(z)$. Поэтому

$$Q_\xi^{(1)}(h) = \frac{1}{\mathsf{E}\xi}\int_0^h [1-F(z)]dz \ge \frac{1}{\mathsf{E}\xi}\int_0^h [1-Q_\xi(z)]dz \ge \frac{h[1-Q_\xi(h)]}{\mathsf{E}\xi}.$$

Пусть ϑ – случайная величина с функцией распределения $Q_\xi(x)$. Тогда, полагая $h\to\infty$, из приведенного выше неравенства с учетом результатов п. 2.5 получим

$$1 \ge \frac{\mathsf{E}\vartheta}{\mathsf{E}\xi} = \frac{\mathsf{E}\vartheta\cdot\mathsf{E}\eta_1}{2\mathsf{E}\xi^2},$$

то есть величины $\mathsf{E}\vartheta$ и $\mathsf{E}\eta_1$ в некотором смысле обратно пропорциональны.

4. Моменты итераций интегральной функции концентрации

В свою очередь, можно рассмотреть интегральную функцию концентрации $Q_{\eta_1}^{(1)}(x)$ случайной величины η_1, функцию распределения которой обозначим $F_{\eta_1}(x)$. Очевидно, что

$$F_{\eta_1}(x) = Q_\xi^{(1)}(x), \quad x\ge 0.$$

Имеем

$$Q^{(1)}_{\eta_1}(x)=\frac{1}{\mathsf{E}\eta_1}\int_0^x\big[1-F_{\eta_1}(z)\big]dz=\frac{1}{\mathsf{E}\eta_1}\int_0^x\big[1-Q^{(1)}_\xi(z)\big]dz=$$
$$=\frac{2\mathsf{E}\xi}{\mathsf{E}\xi^2}\int_0^x\left(1-\frac{1}{\mathsf{E}\xi}\int_0^z\big[1-F(y)\big]dy\right)dz\equiv Q^{(2)}_\xi(x).$$

Определив функции $Q^{(1)}_\xi(x)$ и $Q^{(2)}_\xi(x)$, по аналогии для $n\in\mathbb{N}$ в предположении, что $0<\mathsf{E}\xi^n<\infty$, положим

$$Q^{(n)}_\xi(x)=\frac{1}{\mathsf{E}\eta_{n-1}}\int_0^x\big[1-Q^{(n-1)}_\xi(z)\big]dz,\quad x\ge 0,$$

где η_{n-1} – случайная величина с функцией распределения $Q^{(n-1)}_\xi(x)$, $\eta_0=\xi$, $Q^{(0)}_\xi(x)\equiv F(x)$.

Утверждение 1. *Предположим, что* $\mathsf{E}\xi^{k+1}<\infty$ *при некотором* $k\in\mathbb{N}$. *Тогда для любого* $n\le k$

$$\mathsf{E}\eta_n^k=\frac{\mathsf{E}\eta_{n-1}^{k+1}}{(k+1)\mathsf{E}\eta_{n-1}}.$$

Доказательство. Интегрируя по частям, получим

$$\mathsf{E}\eta_n^k=\int_0^\infty x^k dQ^{(n)}_\xi(x)=\int_0^\infty x^k d\left(\frac{1}{\mathsf{E}\eta_{n-1}}\int_0^x\big[1-Q^{(n-1)}_\xi(z)\big]dz\right)=$$
$$=\frac{1}{\mathsf{E}\eta_{n-1}}\int_0^\infty x^k\big[1-Q^{(n-1)}_\xi(x)\big]dx=\frac{1}{(k+1)\mathsf{E}\eta_{n-1}}\int_0^\infty\big[1-Q^{(n-1)}_\xi(x)\big]dx^{k+1}=$$
$$=\frac{1}{(k+1)\mathsf{E}\eta_{n-1}}\left\{\big[1-Q^{(n-1)}_\xi(x)\big]x^{k+1}\Big|_{x=0}^{\infty}-\int_0^\infty x^{k+1}d\big[1-Q^{(n-1)}_\xi(x)\big]\right\}=$$
$$=\frac{\mathsf{E}\eta_{n-1}^{k+1}}{(k+1)\mathsf{E}\eta_{n-1}},$$

что и требовалось доказать.

Утверждение 2. *Предположим, что* $\mathsf{E}\xi^{n+1}<\infty$ *при некотором* $n\in\mathbb{N}$. *Тогда*

$$\mathsf{E}\eta_n=\frac{\mathsf{E}\xi^{n+1}}{(n+1)\mathsf{E}\xi^n}.$$

Доказательство. Последовательно применим Утверждение 1 с $k=1,2,\ldots,n$ и получим

$$\mathsf{E}\eta_n=\frac{\mathsf{E}\eta_{n-1}^2}{2\cdot\mathsf{E}\eta_{n-1}}=\frac{\mathsf{E}\eta_{n-2}^3}{2\cdot 3\cdot\mathsf{E}\eta_{n-1}\cdot\mathsf{E}\eta_{n-2}}=\ldots=$$

$$= \frac{\mathsf{E}\eta_0^{n+1}}{(n+1)!\mathsf{E}\eta_{n-1} \cdot \mathsf{E}\eta_{n-2} \cdot \ldots \cdot \mathsf{E}\eta_0}. \tag{7}$$

Аналогично

$$\mathsf{E}\eta_{n-1} = \frac{\mathsf{E}\eta_0^{n}}{n!\mathsf{E}\eta_{n-2} \cdot \ldots \cdot \mathsf{E}\eta_0}. \tag{8}$$

Теперь подставим (8) в (7) и получим

$$\mathsf{E}\eta_n = \frac{\mathsf{E}\eta_0^{n+1} n! \mathsf{E}\eta_{n-2} \cdot \ldots \cdot \mathsf{E}\eta_0}{(n+1)!\mathsf{E}\eta_0^n \cdot \mathsf{E}\eta_{n-2} \cdot \ldots \cdot \mathsf{E}\eta_0} = \frac{\mathsf{E}\eta_0^{n+1}}{(n+1)\mathsf{E}\eta_0^n} = \frac{\mathsf{E}\xi^{n+1}}{(n+1)\mathsf{E}\xi^n}, \tag{9}$$

что и требовалось доказать.

Утверждение 3. *Предположим, что* $\mathsf{E}\xi^{n+k} < \infty$ *при некоторых* $n \in \mathbb{N}$ *и* $k \in \mathbb{N}$. *Тогда у случайной величины* η_n *существует момент порядка* k, *причем*

$$\mathsf{E}\eta_n^k = \frac{\mathsf{E}\xi^{n+k}}{C_{n+k}^k \mathsf{E}\xi^n}. \tag{10}$$

Доказательство. Будем поочередно применять Утверждения 1 и 2. Получим

$$\mathsf{E}\eta_n^k = \frac{\mathsf{E}\eta_{n-1}^{k+1}}{(k+1)\mathsf{E}\eta_{n-1}} = \frac{n\mathsf{E}\xi^{n-1}}{(k+1)\mathsf{E}\xi^n} \cdot \mathsf{E}\eta_{n-1}^{k+1} = \frac{n\mathsf{E}\xi^{n-1}}{(k+1)\mathsf{E}\xi^n} \cdot \frac{\mathsf{E}\eta_{n-2}^{k+2}}{(k+2)\mathsf{E}\eta_{n-2}} =$$

$$= \frac{(n-1)n\mathsf{E}\xi^{n-2}}{(k+1)(k+2)\mathsf{E}\xi^n} \cdot \mathsf{E}\eta_{n-2}^{k+2} = \ldots = \frac{n!k!}{(k+n)!} \cdot \frac{\mathsf{E}\xi^0 \mathsf{E}\xi^{n+k}}{\mathsf{E}\xi^n} = \frac{\mathsf{E}\xi^{n+k}}{C_{n+k}^k \mathsf{E}\xi^n},$$

что и требовалось доказать.

Следствие 1. *Пусть существует* $\mathsf{E}\xi^3$. *Тогда*

$$\mathsf{D}\eta_1 = \mathsf{E}\eta_1^2 - \big(\mathsf{E}\eta_1\big)^2 = \frac{\mathsf{E}\xi^3}{3\mathsf{E}\xi} - \left(\frac{\mathsf{E}\xi^2}{2\mathsf{E}\xi}\right)^2.$$

5. Свойства преобразований Лапласа–Стильтьеса итераций интегральной функции концентрации

Преобразование Лапласа–Стильтьеса функции распределения $Q_\xi^{(n)}(x) = Q_{\eta_n}^{(1)}(x)$ (случайной величины η_n) обозначим $\psi_n(s)$. Будем считать, что $\psi_0(s) \equiv \phi(s)$, где $\phi(s)$ – преобразование Лапласа–Стильтьеса исходной случайной величины ξ (функции распределения $F(x)$)

Утверждение 4. *Предположим, что* $\mathsf{E}\xi^{n+m} < \infty$ *при некоторых* $n \in \mathbb{N}$ *и* $m \in \mathbb{N}$. *Тогда для преобразования Лапласа–Стильтьеса* $\psi_n(s)$ *случайной величины* η_n *при* $s \to 0$ *справедливо представление*

$$\psi_n(s) = 1 + n! \sum\nolimits_{k=1}^{m} \frac{(-1)^k s^k}{(n+k)!} \frac{\mathsf{E}\xi^{n+k}}{\mathsf{E}\xi^n} + o(s^m). \tag{11}$$

Доказательство. Хорошо известно (см., например, главу XIII в книге [5]), что, если ζ – неотрицательная случайная величина и существует $\mathsf{E}\zeta^m$, то при $s \to 0$

$$\mathsf{E}e^{-s\zeta} = \sum\nolimits_{k=0}^{m} (-1)^k \frac{s^k}{k!} \mathsf{E}\zeta^k + o(s^m).$$

Поэтому, полагая $\zeta = \eta_n$, с помощью (10) имеем

$$\psi_n(s) = 1 + \sum\nolimits_{k=1}^{m} (-1)^k \frac{s^k}{k!} \mathsf{E}\eta_n^k + o(s^m) = 1 + \sum\nolimits_{k=1}^{m} (-1)^k \frac{s^k}{k!} \frac{\mathsf{E}\xi^{n+k}}{C_{n+k}^k \mathsf{E}\xi^n} + o(s^m) =$$

$$= 1 + n! \sum\nolimits_{k=1}^{m} \frac{(-1)^k s^k}{(n+k)!} \frac{\mathsf{E}\xi^{n+k}}{\mathsf{E}\xi^n} + o(s^m),$$

что и требовалось доказать.

Утверждение 5. *Пусть* $\mathsf{E}\xi_{m+1} < \infty$ *для некоторого* $m \in \mathbb{N}$. *Тогда для всех* $n = 1, 2, \ldots, m$ *справедливо рекуррентное соотношение*

$$\psi_n(s) = \frac{1 - \psi_{n-1}(s)}{s\mathsf{E}\eta_{n-1}}, \quad s \geq 0,$$

где $\psi_0(s) \equiv \phi(s)$.

Доказательство вытекает из следующей цепочки соотношений:

$$\psi_n(n) = \int_0^\infty e^{-sx} dQ_\xi^{(n)}(x) = \int_0^\infty e^{-sx} d_x \left(\frac{1}{\mathsf{E}\eta_{n-1}} \int_0^x \left[1 - Q_\xi^{(n-1)}(z)\right] dz \right) =$$

$$= \frac{1}{\mathsf{E}\eta_{n-1}} \int_0^\infty e^{-sx} \left[1 - Q_\xi^{(n-1)}(x)\right] dx = \frac{1}{s\mathsf{E}\eta_{n-1}} \left(1 + \int_0^\infty Q_\xi^{(n-1)}(x) de^{-sx} \right) =$$

$$= \frac{1}{s\mathsf{E}\eta_{n-1}} \left(1 + Q_\xi^{(n-1)}(x) e^{-sx} \Big|_{x=0}^\infty - \int_0^\infty e^{-sx} dQ_\xi^{(n-1)}(x) \right) = \frac{1 - \psi_{n-1}(s)}{s\mathsf{E}\eta_{n-1}},$$

что и требовалось доказать.

В качестве примера рассмотрим случай, когда $\psi_{n-1}(s) = (1+s)^{-1}$, то есть когда случайная величина η_{n-1} имеет стандартное показательное распределение. В таком случае согласно Утверждению 5

$$\psi_n(s) = \frac{1 - \psi_{n-1}(s)}{s\mathsf{E}\eta_{n-1}} = \frac{1}{s}\left(1 - \frac{1}{1+s}\right) = \frac{1}{s} \cdot \frac{s}{1+s} = \frac{1}{1+s}.$$

Таким образом, показательное распределение является неподвижной точкой преобразования, задаваемого интегральной функцией концентрации.

Вопрос о единственности неподвижной точки этого преобразования в классе распределений с одним и тем же (например, единичным) математическим ожиданием сводится к вопросу о единственности решения дифференциального уравнения

$$F + F' - 1 = 0$$

в классе функций распределения, все точки роста которых сосредоточены на неотрицательной полуоси. Несложно убедиться, что указанное уравнение допускает единственное решение $F(x) = 1 - e^{-x}$. Таким образом, показательное распределение оказывается единственной неподвижной точкой преобразования, задаваемого интегральной функцией концентрации.

Теорема 1. *Пусть* $0 < \mathsf{E}\xi^n < \infty$ *для некоторого* $n \in \mathbb{N}$. *Тогда*

$$\psi_n(s) = \frac{\phi(s) - \sum_{k=0}^{n-1} (-1)^k \dfrac{s^k}{k!} \mathsf{E}\xi^k}{(-1)^n \dfrac{s^n}{n!} \mathsf{E}\xi^n}, \quad s \geq 0.$$

Доказательство проведем по индукции.

1. Пусть $k = 1$. Случайной величине η_1 с функцией распределения $Q_\xi^{(1)}(x)$ соответствует преобразование Лапласа–Стильтьеса

$$\psi^{(1)}(s) = \int_0^\infty e^{-sx} q^{(1)}(x) dx = \frac{1}{\mathsf{E}\xi} \int_0^\infty e^{-sx} \big[1 - F(x)\big] dx =$$

$$= \frac{1}{s\mathsf{E}\xi} \int_0^\infty \big[F(x) - 1\big] de^{-sx} = \frac{e^{-sx}\big[F(x) - 1\big]}{s\mathsf{E}\xi} \Bigg|_{x=0}^{x=\infty} + \frac{1}{s\mathsf{E}\xi} \int_0^\infty e^{-sx} d\big[1 - F(x)\big] =$$

$$= \frac{1 - \phi(s)}{s\mathsf{E}\xi} = \frac{\phi(s) - 1}{(-1)^1 \dfrac{s^1}{1!} \mathsf{E}\xi^1}, \quad s \geq 0.$$

2. Пусть $k = n - 1$. Предположим, что

$$\psi_{n-1}(s) = \frac{\phi(s) - \sum_{k=0}^{n-2} (-1)^k \dfrac{s^k}{k!} \mathsf{E}\xi^k}{(-1)^{n-1} \dfrac{s^{n-1}}{(n-1)!} \mathsf{E}\xi^{n-1}}, \quad s \geq 0. \tag{12}$$

3. Докажем утверждение для $k = n$. С учетом утверждения 5 и соотношений (12) и (9) для произвольного $s \geq 0$ имеем

$$\psi_n(s) = \frac{1 - \psi_{n-1}(s)}{s\mathsf{E}\eta_{n-1}} = \frac{n\mathsf{E}\xi^{n-1}}{s\mathsf{E}\xi^n} \left(1 - \frac{\phi(s) - \sum_{k=0}^{n-2} (-1)^k \dfrac{s^k}{k!} \mathsf{E}\xi^k}{(-1)^{n-1} \dfrac{s^{n-1}}{(n-1)!} \mathsf{E}\xi^{n-1}} \right) =$$

$$= \frac{n\mathsf{E}\xi^{n-1}}{s\mathsf{E}\xi^n} - \frac{\phi(s) - \sum_{k=0}^{n-2} (-1)^k \dfrac{s^k}{k!} \mathsf{E}\xi^k}{(-1)^{n-1} \dfrac{s^n}{n!} \mathsf{E}\xi^n} =$$

$$= \frac{\phi(s) - \sum_{k=0}^{n-2}(-1)^k \frac{s^k}{k!}\mathsf{E}\xi^k}{(-1)^n \frac{s^n}{n!}\mathsf{E}\xi^n} - \frac{n\mathsf{E}\xi^{n-1}}{s\mathsf{E}\xi^n} =$$

$$= \frac{\phi(s) - \Big[\sum_{k=0}^{n-2}(-1)^k \frac{s^k}{k!}\mathsf{E}\xi^k + (-1)^{n-1}\mathsf{E}\xi^{n-1}\frac{s^{n-1}}{(n-1)!}\Big]}{(-1)^n \frac{s^n}{n!}\mathsf{E}\xi^n} =$$

$$= \frac{\phi(s) - \sum_{k=0}^{n-1}(-1)^k \frac{s^k}{k!}\mathsf{E}\xi^k}{(-1)^n \frac{s^n}{n!}\mathsf{E}\xi^n},$$

что и требовалось доказать.

Теорема 1 устанавливает, что преобразование Лапласа–Стильтьеса n-й итерации оператора, определяемого интегральной функцией концентрации, на функции распределения F с конечным n-м моментом, равно отношению разности преобразования Лапласа–Стильтьеса исходной функции распределения F и суммы первых $n-1$ членов его разложения в ряд Тейлора, деленной на n-й член этого разложения.

6. Предельная теорема о бесконечных итерациях интегральной функции концентрации

Предположим, что случайная величина ξ имеет моменты всех порядков. В данном разделе мы будем рассматривать асимптотическое поведение случайных величин η_n, нормированных своими математическими ожиданиями $\mathsf{E}\eta_n \equiv a_n$, и найдем условия, при которых существует случайная величина η такая, что

$$\frac{\eta_n}{\mathsf{E}\eta_n} \Longrightarrow \eta \tag{13}$$

при $n \to \infty$. Преобразование Лапласа–Стильтьеса случайной величины η обозначим $\psi(s)$, $s \geq 0$. Хорошо известно (см., например, главу XIII в книге [5]), что (13) имеет место тогда и только тогда, когда при каждом $s \geq 0$

$$\lim_{n\to\infty} \psi_n\Big(\frac{s}{a_n}\Big) = \psi(s). \tag{14}$$

Зафиксируем произвольное $s > 0$. Из Утверждения 5 вытекает, что

$$\lim_{n\to\infty} \psi_n\Big(\frac{s}{a_n}\Big) = \lim_{n\to\infty} \frac{1-\psi_{n-1}(\frac{s}{a_n})}{s\frac{a_{n-1}}{a_n}} = \lim_{n\to\infty} \frac{1-\psi_{n-1}(\frac{s}{a_{n-1}}\cdot\frac{a_{n-1}}{a_n})}{s\frac{a_{n-1}}{a_n}} =$$

$$= \frac{1 - \lim_{n\to\infty} \psi_{n-1}\left(\frac{s}{a_{n-1}} \cdot \frac{a_{n-1}}{a_n}\right)}{s \lim_{n\to\infty} \frac{a_{n-1}}{a_n}}.$$

Предположим, что

$$\lim_{n\to\infty} \frac{a_{n-1}}{a_n} = 1.$$

С учетом Утверждения 2 это условие принимает вид

$$\lim_{n\to\infty} \frac{\left(\mathsf{E}\xi^n\right)^2}{\mathsf{E}\xi^{n-1} \cdot \mathsf{E}\xi^{n+1}} = 1. \tag{15}$$

Покажем, что условия (14) и (15) влекут соотношение

$$\lim_{n\to\infty} \psi_{n-1}\Big(\frac{s}{a_{n-1}} \cdot \frac{a_{n-1}}{a_n}\Big) = \psi(s). \tag{16}$$

Имеем

$$\left|\psi_{n-1}\Big(\frac{s}{a_{n-1}} \cdot \frac{a_{n-1}}{a_n}\Big) - \psi(s)\right| \leq$$

$$\leq \left|\psi_{n-1}\Big(\frac{s}{a_{n-1}} \cdot \frac{a_{n-1}}{a_n}\Big) - \psi\Big(s \cdot \frac{a_{n-1}}{a_n}\Big)\right| + \left|\psi\Big(s \cdot \frac{a_{n-1}}{a_n}\Big) - \psi(s)\right| \equiv M_1^{(n)} + M_2^{(n)}.$$

Возьмем произвольно малое $\epsilon > 0$. Так как функции $\psi_{n-1}\big(\frac{s}{a_{n-1}}\big)$ и $\psi(s)$ монотонны, ограничены и непрерывны по $s \in [0, \infty)$, то сходимость (14) равномерна (см., например, Лемму 2.10.1 на с. 101-102 в [1]). Поэтому каким бы ни было $\epsilon > 0$, найдется n_1 такое, что $M_1^{(n)} < \epsilon$ для всех $n \geq n_1$. Поскольку функция $\psi(s)$ непрерывна, каким бы ни было $\epsilon > 0$, найдется n_2 такое, что $M_2^{(n)} < \epsilon$ для всех $n \geq n_2$. Таким образом, для всех $n \geq \max\{n_1, n_2\}$

$$\left|\psi_{n-1}\Big(\frac{s}{a_{n-1}} \cdot \frac{a_{n-1}}{a_n}\Big) - \psi(s)\right| < 2\epsilon,$$

то есть имеет место (16).

Пусть имеет место сходимость (13). Тогда в силу (14), (15) и (16) имеем

$$s\,\psi(s) = 1 - \psi(s). \tag{17}$$

Рассмотрим (17) как уравнение относительно $\psi(s)$. Тогда, очевидно, его решение будет иметь вид

$$\psi(s) = \frac{1}{1+s}.$$

Полученное представление для $\psi(s)$ соответствует преобразованию Лапласа–Стильтьеса стандартного показательного распределения с параметром $\lambda = 1$, определяемого функцией распределения $E(x) = 1 - e^{-x}$, $x \geq 0$.

Теорема 2. *Предположим, что случайная величина* ξ *имеет моменты всех порядков, удовлетворяющие условию* (15). *Тогда*

$$\lim_{n\to\infty}\sup_{x\geq 0}\left|Q_{\xi}^{(n)}\left(\frac{x\mathsf{E}\xi^{n+1}}{(n+1)\mathsf{E}\xi^{n}}\right)-E(x)\right|=0. \tag{18}$$

Доказательство. Докажем теорему от противного. Предположим, что условие (15) выполнено, но сходимость (18) не имеет места. В таком случае существует $s_0>0$ такое, что последовательность

$$\left\{\psi_n\left(\frac{s_0}{a_n}\right)\right\}_{n\geq 1} \tag{19}$$

не имеет предела. В силу ограниченности последовательности (19) это может быть, только если существуют две подпоследовательности $\mathcal{N}_1$ и $\mathcal{N}_2$ такие, что существуют

$$\lim_{n\to\infty, n\in\mathcal{N}_1}\psi_n\left(\frac{s_0}{a_n}\right)\neq\lim_{n\to\infty, n\in\mathcal{N}_2}\psi_n\left(\frac{s_0}{a_n}\right).$$

Но повторяя рассуждения, приведенные выше, мы убеждаемся, что последнее невозможно, и пределы всех сходящихся подпоследовательностей последовательности (19) в точке s_0 совпадают с преобразованием Лапласа–Стильтьеса стандартного показательного распределения в этой точке. Таким образом, установлена поточечная сходимость преобразований Лапласа–Стильтьеса нормированных итераций интегральных функций концентрации к преобразованию Лапласа–Стильтьеса стандартного показательного распределения, эквивалентная поточечной сходимости соответствующих функций распределения. Так как предельная показательная функция распределения непрерывна, то поточечная сходимость (13) функций распределения $Q_{\xi}^{(n)}(x)$ случайных величин η_n равномерна. Теорема доказана.

Библиографический список

1. *Галамбош Я.* Асимптотическая теория экстремальных порядковых статистик. – М.: Наука, 1984.

2. *Каштанов В. А., Беляев Ю. К., Соловьев А. Д., Коваленко И. Н., Барзилович Е. Ю., Ушаков И. А.* Вопросы математической теории надежности. – М.: Радио и связь, 1983.

3. *Королев В. Ю., Бенинг В. Е., Шоргин С. Я.*. Математические основы теории риска. 2-е изд. – М.: ФИЗМАТЛИТ, 2011.

4. *Петров В. В.* Суммы независимых случайных величин.– М.: Наука, 1972.

5. *Феллер В.* Введение в теорию вероятностей и ее применения. Т. 2. – М.: Мир, 1967.

6. *Хенгартнер В., Теодореску Р.* Функции концентрации. – М.: Наука, 1980.

7. *Chatterjee A., Mukherjee S. P.* Equilibrium distribution – its role in reliability theory / *Balakrishnan N., Rao C. R.* (Eds). Handbook of Statistics. Volume 20. – Amsterdam: Elsevier Science, 2001. – Chapter 4. P. 105–137.

8. *Sheldon Lin X.* Integrated Tail Distribution / Encyclopedia of Actuarial Science, Volume II. – New York: John Wiley and Sons, 2004. – P. 909–911.

On the «equilibrium distribution» as a concentration function

V.Yu. Korolev

Faculty of Computational Mathematics and Cybernetics of Moscow State University; 119991, Moscow, GSP-1, Leninskie Gory, Moscow State University, Faculty of Computational Mathematics and Cybernetics; Institute of Informatics Problems, Federal Research Center "Commputer Science and Control", Russian Academy of Sciences; vkorolev@sc.msu.ru

Abstract. *It is noticed that the well-known «equilibrium distribution» can be regarded as an integral concentration function. Elementary properties of the integral concentration function are described including its application to the improvement of the Markov inequality. The relationship between the integral and «usual» concentration functions is considered. A survey of the properties of iterations of the integral concentration function is presented. In particular, relations for the moments and Laplace–Stieltjes transforms are given. The limit theorem on convergence of iterations of the integral concentration function to the exponential distribution is proved.*

Key words: *concentration function; integral concentration function; equilibrium distribution; Laplace–Stieltjes transform; exponential distribution; fixed point.*

УДК 519.24

Асимптотические свойства оценки байесовского типа для неизвестного параметра в модели конкурирующих рисков

Н. С. Нурмухамедова

Национальный Университет Узбекистана имени Мирзо Улугбека, математический факультет, Узбекистан, Ташкент; rasulova_nargiza@mail.ru; (+99871)246-90-80; Узбекистан, 100174, г. Ташкент, ВУЗ городок, Национальный Университет Узбекистана, кафедра "Эконометрика и экономическое моделирование"

Аннотация. *В данной работе с использованием свойства локальной асимптотической нормальности для статистики отношения правдоподобия доказана асимптотическая нормальность оценок байесовского типа в модели конкурирующих рисков при случайном цензурировании интервалом ненаблюдения.*

Ключевые слова: *конкурирующие риски, случайное цензурирование, статистика отношения правдоподобия, локальная асимптотическая нормальность, асимптотическая эффективность.*

1. Введение

Статистика отношения правдоподобия (СОП) играет фундаментальную роль в теории принятия решений, в особенности в теории проверки статистических гипотез. Среди разнообразных критериев следует особо выделить критерии, основанные на СОП. Так, например, при проверке простой гипотезы H_0 против сложной альтернативы H_1 о неизвестном законе распределения критерии, построенные на основе СОП, согласно фундаментальной лемме Неймана–Пирсона являются равномерно наиболее мощными при любом объеме наблюдений n [8]. Интерес представляет случай, когда гипотеза H_1 становится близкой к основной гипотезе H_0 с ростом n, т.е. $H_1 = H_{1n} \to H_0$ при $n \to \infty$. Предположим, что основная гипотеза H_0 является простой: $H_0 : \theta = \theta_0$, а альтернатива сложной, но близкой к H_0, т.е. $H_{1n}: \theta_n = \theta_0 + \Gamma_n(\theta_0)^{-1} \cdot h_n$, где $h_n \to h \in \Theta$ при $n \to \infty$. Зададимся СОП

$$l_n(h_n) = \frac{dP^{(n)}_{\theta_0 + \Gamma_n^{-1} \cdot h_n}}{dP^{(n)}_{\theta_0}}. \tag{1.1}$$

На базе асимптотических свойств статистики (1.1) создана целая теория, играющая важную роль в математической статистике. Самое важное свойство СОП (1.1) это свойство локальной асимптотической нормальности (ЛАН), которое дает возможность развития асимптотической теории оценок максимального правдоподобия и байесовских оценок, а также контигуальности вероятностных мер [7–11]. В работах авторов [1–5] свойства ЛАН для СОП установлены в некоторых моделях неполных наблюдений, получаемых случайным цензурированием наблюдений в модели конкурирующих рисков (МКР). В данной работе, используя свойство ЛАН в общей статистической модели, исследуются асимптотические свойства оценки байесовского типа для неизвестного параметра, частности, доказана ее асимптотическая эффективность.

2. Модель конкурирующих рисков при случайном цензурировании интервалом не наблюдения

В МКР интерес представляют случайная величина (с.в.) X со значениями в измеримом пространстве $(\mathcal{X}, \mathcal{B})$ и события $(A^{(1)}, ..., A^{(k)})$, образующие полную группу, где k – фиксировано. На практике МКР отвечает с.в. X, означающая время безотказной работы некоторого объекта (индивидуума, физического устройства), подвергаемого k конкурирующим рискам и выходящего из строя по причине осуществления одного из событий $\{A^{(i)},\ i = 1, ..., k\}$. В таких ситуациях парами $\{(X, A^{(i)}),\ i = 1, ..., k\}$ определяются время и причина выхода из строя объекта (о МКР подробно см. [1]). Повторив эксперимент, в котором наблюдаются совокупности $(X, A^{(1)}, ..., A^{(k)})$ в однородных условиях независимым образом, получаем последовательность копий $\{(X_j, A_j^{(1)}, ..., A_j^{(k)}),\ \ j \geq 1\}$. Пусть $\delta_j^{(i)} = I(A_j^{(i)})$ – индикатор события $A_j^{(i)}$. Каждый случайный вектор $\zeta_j = (X_j, \delta_j^{(1)}, ..., \delta_j^{(k)})$ индуцирует статистическую модель с выборочным пространством

$$\mathcal{Y} = \mathcal{X} \times \{0,1\}^{(k)} = \mathcal{X} \times \{0,1\} \times ... \times \{0,1\}$$

с σ-алгеброй C множеств вида $B \times D_1 \times ... \times D_k$, где $B \in \mathcal{B}$ и $D_i \subset \{0,1\}$, $i = 1, ..., k$. Далее предложим, что распределение вектора ζ_j на $(\mathcal{Y}, C)$ задано с точностью до некоторого неизвестного параметра $\theta = (\theta_1, ..., \theta_s) \in \Theta$:

$$Q^*_\theta(B \times D_1 \times ... \times D_k) = P_\theta(X_1 \in B, \delta_1^{(1)} \in D_1, ..., \delta_1^{(k)} \in D_k), \qquad (2.1)$$

где Θ – открытое множество в R^s. Пусть распределение (2.1) абсолютно непрерывно относительно σ-конечной меры $\nu(x) = \mu(x) \times \varepsilon_1 \times ... \times \varepsilon_k$, где μ-лебегова мера на R и ε_i – считающие меры, сосредоточенные на точках $y^{(i)} \in \{0, 1\}$, $i = \overline{1, k}$. В дальнейшем рассмотрим такую статистическую схему, согласно которой совокупность $(X_j, A_j^{(1)}, ..., A_j^{(k)})$ не наблюдаема, если с.в. X_j окажется в интервале $[Y_{1j}, Y_{2j}]$, где $\{(Y_{1j}, Y_{2j}),\ \ j \geq 1\}$ – последовательность независимых одинаково распределенных (н.о.р.) случайных векторов с (возможно неявно зависящим от θ) неизвестным распределением $G(u, v)$, $(u, v) \in R^2$. При этом совокупности $(X_j, A_j^{(1)}, ..., A_j^{(k)})$ и пары (Y_{1j}, Y_{2j}) предполагаются независимыми и $P_\theta(Y_{1j} \leq Y_{2j}) = 1$ при каждом $j \geq 1$. Реально это соответствует, например, таким испытаниям на долговечность, в которых наблюдения за j-м объектом с временем безотказной работы X_j могут в случайный момент Y_{1j} прерваны и затем в случайный момент Y_{2j} возобновлены. Такую статистическую модель назовем МКР с случайным цензурированием интервалом не наблюдения. В такой ситуации, на самом деле, вместо событий $(A_j^{(1)}, ..., A_j^{(k)})$ наблюдаемыми будут события $(D_j^{(0)}, D_j^{(1)}, ..., D_j^{(k)})$, где

$$\begin{aligned} &D_j^{(0)} = \{\omega : Y_{1j}(\omega) \leq X_j(\omega) \leq Y_{2j}(\omega)\}, \\ &D_j^{(i)} = A_j^{(i)} \cap (\{\omega : X_j(\omega) < Y_{1j}(\omega)\} \cup \{\omega : X_j(\omega) > Y_{2j}(\omega)\}), i = 1, ..., k. \end{aligned}$$

Пусть $\Delta_j^{(i)} = I(D_j^{(i)})$, $i = 0, 1, ..., k$ и $w_j = \varepsilon_{1j} + \varepsilon_{2j}$, где $\varepsilon_{1j} = I(X_j < Y_{1j})$ и $\varepsilon_{2j} = I(X_j > Y_{2j})$. Тогда очевидно $\Delta_j^{(0)} = 1 - w_j$ и $\Delta_j^{(i)} = w_j \delta_j^{(i)}$. Поскольку в МКР интерес представляют совместные свойства пар $\{(X_j, A_j^{(i)}),\ i = \overline{1, k}\}$, то в связи с этим определим субраспределения

$$Q_{i\theta}(B) = Q^*_\theta(B \times \{0\} \times ... \times \{0\} \times \{1\} \times \{0\} \times ... \times \{0\}),\ i = 1, ..., k,\ (2.2)$$

получаемые из (2.1) при $D_i = \{1\}$ и $D_j = \{0\}$, $i \neq j$, $j = 1, ..., k$. Пусть $Q_\theta(B) = \sum\limits_{i=1}^{k} Q_{i\theta}(B)$. Через $h^{(i)}$ и h обозначим плотности субраспределений $Q_{i\theta}$ и Q_θ:

$$Q_{i\theta}(B) = \int\limits_B h^{(i)}(x; \theta) \mu(dx),\ \ i = 1, ..., k,$$

$$Q_\theta(B) = \int\limits_B h(x;\theta)\mu(dx), \tag{2.3}$$

где $h = h^{(1)} + ... + h^{(k)}$. При $B = (-\infty; x]$ обозначим $Q_{i\theta}((-\infty; x]) = H^{(i)}(x;\theta)$, $i = \overline{1,k}$ и $Q_\theta((-\infty; x]) = H(x;\theta)$. Далее определим интегральные функции интенсивности (и.ф.и.), соответствующие парам $(X, A^{(i)})$:

$$\Lambda^{(i)}(x;\theta) = \int\limits_{(-\infty;x]} \lim_{\Delta\downarrow 0} P_\theta(t < X \leq t+\Delta,\ A^{(i)}|X > t)\mu(dt) =$$

$$= \int\limits_{(-\infty;x]} \frac{dH^{(i)}(t;\theta)}{1 - H(t;\theta)},\ \ i = 1,...,k,\ \ x \in R^1. \tag{2.4}$$

Тогда и.ф.и., соответствующая с.в. X есть $\Lambda(x;\theta) = \sum\limits_{i=1}^{k} \Lambda^{(i)}(x;\theta)$. В МКР интерес представляют экспоненциальные функционалы $F^{(i)}(x;\theta) = 1 - \exp\{-\Lambda^{(i)}(x;\theta)\}$, $i = \overline{1,k}$, которые характеризуют распределение i-риска. Поскольку $\Lambda(x;\theta) = -\log(1 - H(x;\theta))$, то

$$1 - H(x;\theta) = P_\theta(X > x) = \prod_{i=1}^{k}(1 - F^{(i)}(x;\theta)). \tag{2.5}$$

Определим плотности $f^{(i)}(x;\theta) = \frac{\partial}{\partial x}F^{(i)}(x;\theta)$, $i = \overline{1,k}$. Тогда плотность интенсивности отказов по i-риску есть $f^{(i)}/(1 - F^{(i)})$. С другой стороны, в силу формул (2.3)–(2.5) для любого $(x;\theta) \in R^1 \times \Theta$ и $i = 1,...,k$

$$\frac{f^{(i)}(x;\theta)}{1 - F^{(i)}(x;\theta)} = \frac{h^{(i)}(x;\theta)}{1 - H(x;\theta)},$$

что эквивалентно

$$h^{(i)}(x;\theta) = f^{(i)}(x;\theta) \prod_{\substack{j=1 \\ j\neq i}}^{k} (1 - F^{(j)}(x;\theta)). \tag{2.6}$$

Пусть на n-м шаге испытаний наблюдению доступна выборка $\mathbb{Z}^{(n)} = (Z_1, ..., Z_n)$, для $Z_j = w_j X_j + (1 - w_j)[Y_{1j}, Y_{2j}]$, т.е. каждое наблюдение Z_j есть либо с.в. X_j (при $w_j = 1$) либо интервал $[Y_{1j}, Y_{2j}]$ (при $w_j = 0$). Через $p(z;\theta)$ обозначим плотность распределения одного наблюдения, которая включает в себя и множители, зависящие от неизвестного распределения G. Отбросив эти множители и учитывая представление (2.6), получаем следующую "усеченную"(т.е. частичную) функцию правдоподобия выборки $\mathbb{Z}^{(n)}$:

$$p_n(\mathbb{Z}^{(n)};\theta) = \prod_{m=1}^{n} p(Z_m;\theta) = \prod_{m=1}^{n} \left[H(Y_{2m};\theta) - H(Y_{1m};\theta)\right]^{1-w_m} \times$$

$$\times \prod_{m=1}^{n}\prod_{i=1}^{k}\left[f^{(i)}(X_m;\theta)\prod_{\substack{j=1\\ j\neq i}}^{k}(1-F^{(j)}(X_m;\theta))\right]^{\delta_m^{(i)} w_m} =$$

$$= \prod_{m=1}^{n}\left\{\prod_{i=1}^{k}\left[h^{(i)}(X_m;\theta)\right]^{\delta_m^{(i)} w_m}\left[H(Y_{2m};\theta) - HY_{1m};\theta\right]^{1-w_m}\right\}. \qquad (2.7)$$

Пусть при каждом $u \in R^s$, $\theta + n^{-1/2}u = \Psi_n(u;\theta) \in \Theta$ и $\widetilde{Q}_\theta^{(n)}$-распределение, индуцированное выборкой $\mathbb{Z}^{(n)}$. Определим СОП модели

$$L_{n,\theta}(u) = d\widetilde{Q}^{(n)}_{\Psi_n(u;\theta)}(\mathbb{Z}^{(n)})/d\widetilde{Q}_\theta^{(n)}(\mathbb{Z}^{(n)}) = \frac{p_n(\mathbb{Z}^{(n)};\Psi_n(u;\theta))}{p_n(\mathbb{Z}^{(n)};\theta)} =$$

$$= \prod_{m=1}^{n}\prod_{i=1}^{k}\left[\frac{h^{(i)}(X_m;\Psi_n(u;\theta))}{h^{(i)}(X_m;\theta)}\right]^{\delta_m^{(i)} w_m} \times$$

$$\times \prod_{m=1}^{n}\left[\frac{H(Y_{2m};\Psi_n(u;\theta)) - H(Y_{1m};\Psi_n(u;\theta)}{H(Y_{2m};\theta) - H(Y_{1m};\theta)}\right]^{1-w_m}. \qquad (2.8)$$

Пусть $N^{(i)} = \{x : h^{(i)}(x;\theta) > 0\}$ и $N = \bigcap_{i=1}^{k} N^{(i)}$. Далее будут необходимы следующие условия регулярности:

(У1) множество $\{N^{(i)},\ i=\overline{1,k}\}$ не зависит от θ и $N \neq \emptyset$;

(У2) существуют производные $\dfrac{\partial^m h^{(i)}(x;\theta)}{\partial\theta_j^m}$, $m = 1,2$; $i = 1,...,k$; $j = 1,...,s$ для всех $\theta \in \Theta$;

(У3) $\int\limits_{-\infty}^{\infty}\left|\frac{\partial^m h^{(i)}(x;\theta)}{\partial\theta_j^m}\right|\mu(dx) < \infty$, $m = 1,2$; $i = 1,...,k$; $j = 1,...,s$ для всех $\theta \in \Theta$;

(У4) интегралы $I_{lj}^{(i)}(\theta) = M_\theta\left[\frac{\partial}{\partial\theta_l}\log h^{(i)}(X;\theta)\frac{\partial}{\partial\theta_j}\log h^{(i)}(X;\theta)\right]$ существуют и конечны для всех $l,j = 1,...,s$ и $\theta \in \Theta$;

(У5) матрица $I_X(\theta) = \left\|I_{lj}^X(\theta)\right\|_{l,j=\overline{1,s}} = \left\|\sum_{i=1}^{k} I_{lj}^{(i)}(\theta)\right\|_{l,j=\overline{1,s}} = \sum_{i=1}^{k} I^{(i)}(\theta)$ положительно определена для всех $\theta \in \Theta$.

Легко проверить, что $I^{(i)}(\theta)$ – матрица информации Фишера, отвечающая паре $(X,\delta^{(i)})$, а $I_X(\theta)$ – с.в. X.

Пусть

$$S_n(\mathbb{Z}^{(n)};\theta)=\frac{\partial\log p_n(\mathbb{Z}^{(n)};\theta)}{\partial\theta}=\sum_{j=1}^{n} l_\theta(X_j,Y_{1j},Y_{2j},w_j),$$

где

$$l_\theta(x,y_1,y_2,w)=w\sum_{i=1}^{k}\delta^{(i)}\frac{\partial\log h^{(i)}(x;\theta)}{\partial\theta}+(1-w)\frac{\partial\log(H(y_2;\theta)-H(y_1;\theta))}{\partial\theta}.$$

Введем матрицу $\mathbb{J}(\theta)=\mathbb{J}_1(\theta)+\mathbb{J}_2(\theta)$, где

$$\mathbb{J}_1(\theta)=\sum_{i=1}^{k}\int_{-\infty}^{\infty}\int_{y_1}^{\infty}\left[\int_{-\infty}^{y_1}\frac{\partial\log h^{(i)}(x;\theta)}{\partial\theta}\left(\frac{\partial\log h^{(i)}(x;\theta)}{\partial\theta}\right)^T dH^{(i)}(x;\theta)+\right.$$

$$\left.+\int_{y_2}^{\infty}\frac{\partial\log h^{(i)}(x;\theta)}{\partial\theta}\left(\frac{\partial\log h^{(i)}(x;\theta)}{\partial\theta}\right)^T dH^{(i)}(x;\theta)\right]dG(y_1,y_2),$$

$$\mathbb{J}_2(\theta)=\int_{-\infty}^{\infty}\int_{-\infty}^{y_1}\frac{\partial\log(H(y_2;\theta)-H(y_1;\theta))}{\partial\theta}\left(\frac{\partial\log(H(y_2;\theta)-H(y_1;\theta))}{\partial\theta}\right)^T\times$$

$$\times(H(y_2;\theta)-H(y_1;\theta))dG(y_1,y_2).$$

Пусть (u,v) – скалярное произведение для векторов $u,v\in R^s$. Следующая теорема содержит условие ЛАН для СОП.

Теорема 2.1 [5]. *Пусть выполнены условия регулярности (У1)–(У5) и* $\det\{\mathbb{J}(\theta)\}\neq 0$ *для* $\theta\in\Theta$. *Тогда для СОП* $L_{n,\theta}(u)$ *имеет место представление*

$$L_{n,\theta}(u)=\exp\left\{\sum_{j=1}^{n}\frac{(l_\theta(X_j,Y_{1j},Y_{2j},w_j),u)}{\sqrt{n}}-\frac{(\mathbb{J}(\theta)u,u)}{2}+R_n(u;\theta)\right\},\quad(2.9)$$

где при $n\to\infty$ *для каждого* $u\in R^s$

$$R_n(u;\theta)\overset{\widetilde{Q}_\theta^{(n)}}{\to}0\qquad(2.10)$$

и

$$\mathcal{L}\left(n^{-1/2}\sum_{j=1}^{n} l_\theta(X_j,Y_{1j},Y_{2j},w_j)|\widetilde{Q}_\theta^{(n)}\right)\to\mathbb{N}_s(0;\mathbb{J}(\theta)).\qquad(2.11)$$

3. Асимптотически нормальный класс оценок неизвестных параметров

Пусть $\{\pi(u), u \in \Theta\}$ – неотрицательная измеримая функция и $l(d;\theta) == (d-\theta)^2$ – функция потерь на множестве $D \times \Theta$, где D – множество возможных оценок для θ. Рассмотрим оценки $\theta_n \in D$, определяемые соотношением

$$\widehat{\theta}_n = \arg\min_{d \in D} \frac{\int\limits_{\Theta} l(d;\theta) p_n(\tilde{Z}^{(n)};\theta)\pi(\theta)d\theta}{\int\limits_{\Theta} p_n(\tilde{Z}^{(n)};\theta)\pi(\theta)d\theta}. \tag{3.1}$$

Заметим, что если θ – с.в. с априорной плотностью π, то $\widehat{\theta}_n$ является байесовской оценкой для θ. Докажем асимптотическую нормальность оценок $\widehat{\theta}_n$, предельные распределения которых не зависят от функций π.

Теорема 3.1. *Пусть выполнены условия регулярности (У1)-(У5), функция $\pi(\theta)$ непрерывна в окрестности точки θ_0 и $\pi(\theta_0) \neq 0$. Тогда при $n \to \infty$,*

$$\mathcal{L}\left(\sqrt{n}(\theta_n - \theta_0)/Q_{\theta_0}^{(n)}\right) \to N(0, \mathbb{J}^{-1}(\theta_0)).$$

Доказательство теоремы 3.1. В условиях (У1) – (У5) из теоремы 2.1 следует справедливость ЛАН (2.11) для СОП, где

$$\mathbb{J}(\theta) = \sum_{i=1}^{k} \int\limits_{-\infty}^{\infty} \int\limits_{y_1}^{\infty} \left[\int\limits_{-\infty}^{y_1} \left(\frac{\partial \log h^{(i)}(x;\theta)}{\partial \theta} \right)^2 dH^{(i)}(x;\theta) + \right.$$

$$\left. + \int\limits_{y_2}^{\infty} \left(\frac{\partial \log h^{(i)}(x;\theta)}{\partial \theta} \right)^2 dH^{(i)}(x;\theta) \right] dG(y_1, y_2) +$$

$$+ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{y_1} \left(\frac{\partial \log(H(y_2;\theta) - H(y_1;\theta))}{\partial \theta} \right)^2 (H(y_2;\theta) - H(y_1;\theta)) dG(y_1, y_2).$$

Согласно определению (3.1) оценка $\widehat{\theta}_n$ имеет представление

$$\widehat{\theta}_n = \frac{\int\limits_{\Theta} \theta p_n(\tilde{Z}^{(n)};\theta)\pi(\theta)d\theta}{\int\limits_{\Theta} p_n(\tilde{Z}^{(n)};\theta)\pi(\theta)d\theta}. \tag{3.2}$$

В интегралах (3.2) переменную θ заменим на близкую альтернативу $\theta_0 + \dfrac{u}{\sqrt{n}} = \theta_n \in \Theta$, $u \in \mathbb{R}^1$. Тогда, используя $L_{n,\theta_0}(u)$, будем иметь

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{\int\limits_{-\infty}^{+\infty} u \exp\left(L_{n,\theta_0}(u)\right) \pi\left(\theta_0 + \frac{u}{\sqrt{n}}\right) du}{\int\limits_{-\infty}^{+\infty} \exp\left(L_{n,\theta_0}(u)\right) \pi\left(\theta_0 + \frac{u}{\sqrt{n}}\right) du}. \tag{3.3}$$

Согласно (2.9) при каждом $u \in R^1$: $\exp\left(L_{n,\theta}(u)\right) \overset{D}{\Rightarrow} \exp\left(L(u)\right)$ при $n \to \infty$, где $L(u) = u\mathbb{J}^{1/2}(\theta_0)\zeta - \dfrac{u^2}{2}\mathbb{J}(\theta_0)$, здесь ζ – стандартная нормальная с.в. Отсюда следует, что конечномерные распределения процесса $L_{n,\theta}(u)$ сходятся к конечномерным распределениям процесса $L(u)$. В (3.3), с учетом условий (У4), формально переходим к пределу под знаком интеграла и тогда получаем равенство

$$\zeta\mathbb{J}^{1/2}(\theta_0) = \frac{\int\limits_{-\infty}^{+\infty} u \exp\left(L(u)\right) du}{\int\limits_{-\infty}^{+\infty} \exp\left(L(u)\right) du}. \tag{3.4}$$

Для обоснования такого перехода выберем фиксированное число $C > 0$ и для $u \in [-C, C]$ покажем сперва непрерывность процесса $L_{n,\theta}(u)$. Пусть числа u_1 и u_2 таковы, что $\theta_0 + u_m \in [-C, C]$, $m = 1, 2$. Покажем, что при достаточно больших n:

$$M_{\theta_0}(L_{n,\theta_0}(u_1) - L_{n,\theta_0}(u_2))^2 \le \alpha(u_1 - u_2)^2, \tag{3.5}$$

где число $\alpha > 0$. Мы имеем

$$M_{\theta_0}(L_{n,\theta_0}(u_1) - L_{n,\theta_0}(u_2))^2 =$$

$$= M_{\theta_0}\left\{\sum_{m=1}^{n}\left\{w_m \sum_{i=1}^{k}\left\{\log h^{(i)}\left(Z_m; \theta_0 + \frac{u_1}{\sqrt{n}}\right) - \log h^{(i)}\left(Z_m; \theta_0 + \frac{u_2}{\sqrt{n}}\right)\right\} +\right.\right.$$

$$+(1 - w_n)\left\{\left[\log H\left(Y_{2m}; \theta_0 + \frac{u_1}{\sqrt{n}}\right) - \log H\left(Y_{1m}; \theta_0 + \frac{u_1}{\sqrt{n}}\right)\right] -\right.$$

$$\left.\left.\left. - \left[\log H\left(Y_{2m}; \theta_0 + \frac{u_2}{\sqrt{n}}\right) - \log H\left(Y_{1m}; \theta_0 + \frac{u_2}{\sqrt{n}}\right)\right]\right\}\right\} \le$$

$$\le \sum_{i=1}^{k} \int\limits_{-\infty}^{\infty} \int\limits_{y_1}^{\infty} \left[\int\limits_{-\infty}^{y_1} \left[\log \frac{h^{(i)}(x; \theta_0 + u_1/\sqrt{n})}{h^{(i)}(x; \theta_0 + u_2/\sqrt{n})}\right]^2 dH^{(i)}(x; \theta) +\right.$$

$$+\int\limits_{y_2}^{+\infty}\left[\log\frac{h^{(i)}\left(x;\theta_0+u_1/\sqrt{n}\right)}{h^{(i)}\left(x;\theta_0+u_2/\sqrt{n}\right)}\right]^2 dH^{(i)}(x;\theta)\Bigg] dG(y_1,y_2)+$$

$$+\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{y_1}\left\{\log\frac{H\left(y_2;\theta_0+u_1/\sqrt{n}\right)}{H\left(y_1;\theta_0+u_1/\sqrt{n}\right)}-\log\frac{H\left(y_2;\theta_0+u_2/\sqrt{n}\right)}{H\left(y_1;\theta_0+u_2/\sqrt{n}\right)}\right\}^2\times$$

$$\times[(H(y_2;\theta_0)-H(y_1;\theta_0))]dG(y_1,y_2)=\mathbb{J}(\theta_0)(u_1-u_2)^2, \tag{3.5}$$

что доказывает (3.4). Таким образом, согласно (3.5) процесс $\{L_n(u),\ u\in[-C,C]\}$ является элементом пространства $\mathbb{C}[-C,C]$. С другой стороны, при любых t_1 и t_2 функционал

$$\Phi(\psi)=t_1\int\limits_{-C}^{C}u\,\psi(u)du+t_2\int\limits_{-C}^{C}u\,\psi(u)du$$

является непрерывным по ψ. Согласно теореме Крамера–Уолда ввиду непрерывности и условия (VI) из теоремы 2 в [6, с. 523] следует, что распределения случайных векторов

$$\left(\int\limits_{-C}^{C}u\exp\left(L_n(u)\right)\pi\left(\theta_0+\frac{u}{\sqrt{n}}\right)du\,,\ \int\limits_{-C}^{C}\exp\left(L_n(u)\right)\pi\left(\theta_0+\frac{u}{\sqrt{n}}\right)du\right)$$

сходятся к распределению вектора

$$\left(\pi\left(\theta_0\right)\int\limits_{-C}^{C}u\exp\left(L(u)\right)du\,,\ \pi\left(\theta_0\right)\int\limits_{-C}^{C}\exp\left(L(u)\right)du\right).$$

С другой стороны, для любого $\varepsilon>0$ существует $\delta>0$ такое, что

$$P\left(\pi\left(\theta_0\right)\int\limits_{|u|>C}u\exp\left(L(u)\right)du\ >\delta\right)<\varepsilon \tag{3.6}$$

и

$$P\left(\pi\left(\theta_0\right)\int\limits_{|u|>C}\exp\left(L(u)\right)du\ >\delta\right)<\varepsilon. \tag{3.7}$$

При достаточно больших n неравенства типа (3.6) и (3.7) имеют место для $L_n(u)$. Более того, при достаточно больших C и n верны неравенства

$$P\left(t_1 \int\limits_{|u|>C} u \exp\left(L_n(u)\right) \pi\left(\theta_0 + \frac{u}{\sqrt{n}}\right) du + \right.$$

$$\left. + t_2 \int\limits_{|u|>C} \exp\left(L_n(u)\right) \pi\left(\theta_0 + \frac{u}{\sqrt{n}}\right) du > \frac{1}{C^N}\right) \leq$$

$$\leq \sum_{|l|>C} P\left(\int\limits_{l}^{l+1} (|u|+1) \exp\left(L_n(u)\right) \pi\left(\theta_0 + \frac{u}{\sqrt{n}}\right) du > \frac{1}{l^N (t_1 \vee t_2)}\right) \leq$$

$$\leq \sum_{|l|>C} P\left(\max_{u \in [l, l+1]} \{\exp\left(L_n(u)\right)\} > \frac{l^{-(N+M+2)}}{(t_1 \wedge t_2)}\right) \leq \frac{\lambda_N}{C^N}. \tag{3.8}$$

Ввиду (3.8), при больших n

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{\int\limits_{-C}^{C} u \exp\left(L_n(u)\right) du}{\int\limits_{-C}^{C} \exp\left(L_n(u)\right) du} + r_n(C),$$

где $P\left(|r_n(C)| > \delta\right) < \varepsilon$. Таким образом, равенство (3.3) имеет место, и тем самым теорема доказана.

Замечание. Согласно теореме 3.1 и определению Фишера [7, с.127] оценку $\widehat{\theta}_n$ можно назвать асимптотически эффективной.

Библиографический список

1. *Абдушукуров А. А.* Статистика неполных наблюдений. – Ташкент: Изд-во Ташкентского университета, 2009. – 296 с.

2. *Абдушукуров А. А., Нурмухамедова Н. С.* Асимптотика для статистики отношения правдоподобия в модели конкурирующих рисков при многократном цензурировании справа // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2012. – Вып. 24. – С.4–15.

3. *Абдушукуров А. А., Нурмухамедова Н. С.* Локальная асимптотическая нормальность в модели конкурирующих рисков // Узбекский математический журнал. – 2012. – №2. – С.5–12.

4. *Abdushukurov A. A., Nurmuhamedova N. S.* Local approximate normality of likelihood ratio statistics in competing ricks model under random censorship from both sides// Far East Journal of Theoretical Statistics. – 2013. – V.42, №2. – P.107–122.

5. *Abdushukurov A. A., Nurmukhamedova N. S.* Locally asymptotically normality of the family of distributions by incomplete observations //Journal of Siberian Federal University. Mathematics & Physics. – 2014. – V.7, №2. – P.141–154.

6. *Гихман И. И., Скороход А.В.* Введение в теорию случайных процессов. – М.:Наука, 1977. – 568с.

7. *Ибрагимов И. А., Хасьминский Р. З.* Асимптотическая теория оценивания. – М.: Наука, 1979. – 527 с.

8. *Лэман Э.* Проверка статистических гипотез. – М.: Наука, 1964. – 408 с.

9. *Русас Дж.* Контигуальность вероятностных мер. – М.: Мир, 1975. – 254 с.

10. *Bilingsley P.* Convergence of Probability Measures. – M.: Nauka, 1977. – 351 p.

11. *Hajek J.* Local asymptotic minimax and admissibility in estimation, Proc. Sixth. Berkeley Symp. on Math. Statist. and Prob. – 1972. – V.1. – P.175–194.

Asymptotic properties of Bayesian type estimation for unknown parameter in competing risk model under random censoring

N.S. Nurmukhamedova

National University of Uzbekistan named after Mirzo Ulugbek, faculty of mechanics and mathematics, department of Econometrics and economic modeling, Uzbekistan, Tashkent; rasulova_ nargiza @ mail.ru

Abstract. *In this paper, using the property of local asymptotic normality for likelihood ratio statistics, we prove the asymptotic normality of Bayesian type estimates in in competing risks model under random censoring by non observing intervals.*

Key words: *competing risks; random censoring; likelihood ratio statistics; locally asymptotically normality, asymptotic efficiency.*

УДК 519.24

Многообразие критериев проверки однородности двух независимых выборок

А.И. Орлов

Московский государственный технический университет им. Н.Э. Баумана, Институт высоких статистических технологий и эконометрики; Москва, Россия; prof-orlov@mail.ru; http://orlovs.pp.ru; +7(916)8305117; 123104, Москва, Сытинский пер., д.7/14, кв.14.

Аннотация. *Дана сводка основных результатов, касающихся методов проверки однородности двух независимых выборок. Она позволяет системно анализировать многообразие таких методов с целью выбора наиболее адекватного для обработки конкретных данных. На основе базовой вероятностно-статистической модели сформулированы основные постановки задачи проверки однородности двух независимых выборок. Дан сравнительный анализ критериев Стьюдента и Крамера-Уэлча, предназначенных для проверки однородности математических ожиданий. Из непараметрическх методов проверки однородности рассмотрены критерии Вилкоксона, Смирнова, Лемана-Розенблатта. Для проверки абсолютной однородности рекомендовано использовать критерий Лемана-Розенблатта. Обсуждаются проблемы разработки и применения непараметрических критериев.*

Ключевые слова: *проверка статистических гипотез, независимые выборки, однородность характеристик, абсолютная однородность, критерий Крамера-Уэлча, критерий Вилкоксона, критерий Смирнова, критерий Лемана-Розенблатта*

1. Введение

Проверка однородности двух независимых выборок – классическая область математической статистики. Основополагающей является статья Стьюдента 1908 г. [1]. За более чем 111 лет получены многочисленные результаты различными авторами, в том числе нами. Однако чувствуется потребность в анализе всего многообразия постановок задач проверки статистических гипотез однородности двух независимых выборок, а также соответствующих критериев. Такому анализу посвящена настоящая статья.

2. Базовая вероятностно-статистическая модель

В математико-статистических терминах постановка задачи такова: имеются две выборки $x_1, x_2, ..., x_m$ и $y_1, y_2, ..., y_n$, требуется проверить их однородность. Выборка моделируется как совокупность независимых одинаково распределенных числовых случайных величин. Термин "однородность"уточняется ниже. Противоположным к "однородности"понятием является "различие"(или "наличие эффекта"). Можно переформулировать задачу: требуется проверить, есть ли различие между выборками. Если различия нет, то для дальнейшего изучения с целью практического применения выводов две рассматриваемые выборки часто объединяют в одну. Например, в маркетинге важно выделить сегменты потребительского рынка. Если установлена однородность двух выборок мнений потребителей, то возможно объединение сегментов, из которых эти выборки взяты, в один. В дальнейшем это позволит осуществлять по отношению к ним одинаковую маркетинговую политику (проводить одни и те же рекламные мероприятия и т.п.). Если же установлено различие, то поведение потребителей в двух сегментах различно, объединять эти сегменты нельзя, и могут понадобиться различные маркетинговые стратегии, своя для каждого из этих сегментов. Для обоснованного выбора и применения организационно-экономических (эконометрических, статистических) методов необходимо прежде всего построить и обосновать *вероятностную модель порождения данных.* При проверке однородности двух выборок общепринята модель, в которой $x_1, x_2, ..., x_m$ рассматриваются как результаты m независимых наблюдений некоторой случайной величины X с функцией распределения $F(x)$, неизвестной статистику, а $y_1, y_2, ..., y_n$ – как результаты n независимых наблюдений, вообще говоря, другой случайной величины Y с функцией распределения $G(x)$, также неизвестной статистику. Предполагается также, что наблюдения в одной выборке не зависят от наблюдений в другой, поэтому выборки и называют независимыми. Возможность применения модели в конкретной реальной ситуации требует обоснования. Независимость и одинаковая распределенность результатов наблюдений (измерений, испытаний, анализов, опытов, обследований) могут быть установлены или исходя из методики проведения конкретных наблюдений, или путем проверки статистических гипотез независимости и одинаковой распределенности с помощью соответствующих критериев проверки статистических гипотез [2].

3. Основные постановки задачи проверки однородности двух независимых выборок

Наивысшая степень однородности (абсолютная однородность) достигается, если обе выборки взяты из одной и той же генеральной совокупности, т. е. справедлива нулевая гипотеза

$$H_0 : F(x) = G(x) \text{ } \textit{при всех } x.$$

Отсутствие абсолютной однородности означает, что верна альтернативная гипотеза, согласно которой

$$H_1 : F(x_0) \neq G(x_0)$$

хотя бы при одном значении аргумента x_0. Если гипотеза H_0 принята, то выборки можно объединить в одну, если нет – то нельзя.

В некоторых случаях целесообразно проверять не совпадение функций распределения, а лишь совпадение некоторых характеристик случайных величин X и Y – математических ожиданий, медиан, дисперсий, коэффициентов вариации и др. (т.е. проверять однородность тех или иных характеристик). Например, однородность математических ожиданий означает, что справедлива гипотеза

$$H_0' : M(X) = M(Y),$$

где $M(X)$ и $M(Y)$ – математические ожидания случайных величин X и Y, результаты наблюдений над которыми составляют первую и вторую выборки соответственно. Доказательство различия между выборками в рассматриваемом случае – это доказательство справедливости альтернативной гипотезы

$$H_1' : M(X) \neq M(Y).$$

Если гипотеза H_0 верна, то и гипотеза H_0' верна, но из справедливости H_0', вообще говоря, не следует справедливость H_0. Математические ожидания могут совпадать для различающихся между собой функций распределения. В частности, если в результате обработки выборочных данных принята гипотеза H_0', то отсюда *не следует*, что две выборки можно объединить в одну.

Однако в ряде ситуаций целесообразна проверка именно гипотезы H_0'. Например, пусть функция спроса на определенный товар или услугу оценивается путем опроса потребителей (первая выборка) или с помощью данных о продажах (вторая выборка). Тогда маркетологу важно проверить гипотезу об отсутствии систематических расхождений результатов этих двух методов, т.е. гипотезу о равенстве математических ожиданий.

Другой пример – из производственного менеджмента. Пусть изучается эффективность управления бригадами рабочих на предприятии с помощью двух организационных схем, результаты наблюдения – объем производства продукции или услуг на одного члена бригады (производительность), а показатель эффективности организационной схемы – средний (по предприятию) объем производства на одного рабочего. Тогда для сравнения эффективности организационных схем достаточно проверить гипотезу H_0'. Если она принята, то нет оснований заявлять о том, что организационные схемы различаются по эффективности.

Иногда нужно проверить однородность дисперсий. Например, различаются ли два способа измерения по величине случайной ошибки – т.е. по дисперсии случайных погрешностей. Или, например, однородность коэффициентов вариации [3].

4. Проверка однородности характеристик

Наиболее часто рассматривают проверку однородности математических ожиданий.

Традиционный метод проверки однородности (двухвыборочный критерий Стьюдента). Для дальнейшего критического разбора опишем традиционный статистический метод проверки однородности [1]. Он широко использовался в течение всего XX в. Хотя к настоящему времени этот метод устарел (см. ниже), но по традиции продолжает встречаться в учебной литературе, и потому и применяться для анализа конкретных данных.

При использовании традиционного метода проверки однородности вычисляют выборочные средние арифметические в каждой выборке

$$\overline{x} = \frac{1}{m}\sum_{1\le i\le m} x_i, \quad \overline{y} = \frac{1}{n}\sum_{1\le i\le n} y_i,$$

затем выборочные дисперсии

$$s_x^2 = \frac{1}{m-1}\sum_{1\le i\le m}(x_i-\overline{x})^2, \quad s_y^2 = \frac{1}{n-1}\sum_{1\le i\le n}(y_i-\overline{y})^2$$

и статистику Стьюдента t, на основе которой принимают решение,

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{(m-1)s_x^2 + (n-1)s_y^2}} \sqrt{\frac{mn(m+n-2)}{m+n}}. \qquad (1)$$

По заданному уровню значимости α и числу степеней свободы $(m+n-2)$ из таблиц распределения Стьюдента (см., например, [2]) находят критическое значение t_{kp}. Если $|t| > t_{kp}$, то гипотезу однородности (отсутствия различия) отклоняют, если же $|t| \leq t_{kp}$, то принимают. (При односторонних альтернативных гипотезах вместо условия $|t| > t_{kp}$ проверяют, что $t > t_{kp}$; эту постановку рассматривать не будем, так как в ней нет принципиальных отличий от обсуждаемой здесь.)

В литературе зачастую описывается только приведенный выше алгоритм. Этого недостаточно для квалифицированного анализа статистических данных. Рассмотрим условия применимости традиционного метода проверки однородности, основанного на использовании статистики t Стьюдента, а также обсудим современные методы проверки однородности двух выборок.

Классические условия применимости критерия Стьюдента. Согласно математико-статистической теории должны быть выполнены два классических условия применимости критерия Стьюдента, основанного на использовании статистики t, заданной формулой (1):

а) результаты наблюдений имеют нормальные распределения:

$$F(x) = N(x; m_1, \sigma_1^2), \quad G(x) = N(x; m_2, \sigma_2^2)$$

с математическими ожиданиями m_1 и m_2 и дисперсиями σ_1^2 и σ_2^2 в первой и во второй выборках соответственно;

б) дисперсии результатов наблюдений в первой и второй выборках совпадают:

$$D(X) = \sigma_1^2 = D(Y) = \sigma_2^2.$$

Если условия а) и б) выполнены, то нормальные распределения $F(x)$ и $G(x)$ отличаются только математическими ожиданиями, а поэтому *обе* гипотезы H_0 и H_0' (см. раздел 3 выше) сводятся к гипотезе

$$H_0'' : m_1 = m_2,$$

а *обе* альтернативные гипотезы H_1 и H_1' сводятся к гипотезе

$$H_1'' : m_1 \neq m_2.$$

Если условия а) и б) выполнены, то статистика t при справедливости H_0'' имеет распределение Стьюдента с $(m+n-2)$ степенями свободы. Только в этом случае описанный выше традиционный метод обоснован безупречно. Если хотя бы одно из условий а) и б) не выполнено, то нет никаких оснований считать, что статистика t имеет распределение Стьюдента, поэтому применение традиционного метода, строго говоря, не обосновано. Обсудим возможность проверки этих условий и последствия их нарушений.

Имеют ли результаты наблюдений нормальное распределение? Как подробно показано в литературе (см., например, сводку [4]), априори нет оснований предполагать нормальность распределения результатов экономических, технико-экономических, технических, медицинских и иных наблюдений. Следовательно, нормальность надо проверять (или использовать непараметрические методы анализа данных). Разработано много статистических критериев для проверки нормальности распределения результатов наблюдений [2]. Однако проверка нормальности – более сложная и трудоемкая статистическая процедура, чем проверка однородности (как с помощью статистики t Стьюдента, так и с использованием непараметрических критериев, рассматриваемых ниже).

Для достаточно надежного установления нормальности требуется весьма большое число наблюдений. В [5] показано, что для того, чтобы гарантировать, что функция распределения результатов наблюдений отличается от некоторой нормальной не более чем на 0,01 (при любом значении аргумента), требуется порядка 2500 наблюдений. В большинстве технических, экономических, медицинских и иных исследований число наблюдений существенно меньше.

Есть и еще одна общая причина отклонений от нормальности: любой результат наблюдения записывается конечным (обычно 2 – 5) количеством цифр, а с математической точки зрения вероятность такого события равна 0. Точнее, для случайной величины с непрерывной плотностью распределения вероятность попадания в счетное множество рациональных чисел равна 0. Следовательно, при статистической обработке данных в организационно-экономических исследованиях распределение результатов наблюдений практически всегда более или менее отличается от нормального распределения.

Последствия нарушения условия нормальности. Если условие а) не выполнено, то распределение статистики t не является распределением Стьюдента. Однако можно показать, используя Центральную предельную теорему теории вероятностей и теоремы о наследовании сходимости [5, гл.4], что при справедливости гипотезы H_0' и условия б) распределение статистики t при росте объемов выборок приближается к стандартному нормальному распределению $\Phi(x) = N(x; 0, 1)$. К этому же распределению приближается распределение Стьюдента при возрастании числа степеней свободы. Другими словами, несмотря на нарушение условия нормальности критерий Стьюдента можно использовать (при определенных условиях!) для проверки гипотезы H_0' при больших объемах выборок. При этом вместо таблиц распределения Стьюдента достаточно пользоваться таблицами стандартного нормального распределения $\Phi(x)$. Это утверждение справедливо для любых функций распределения $F(x)$ и $G(x)$ таких, что $M(X) = M(Y)$, $D(X) = D(Y)$ и выполнены некоторые условия, обычно считающиеся справедливыми в реальных задачах [6, 7]. Если же $M(X) \neq M(Y)$, то нетрудно вычислить, что при больших объемах выборок

$$P(t \leq x) \approx \Phi(x - a_{mn}), \tag{2}$$

где

$$a_{mn} = \frac{\sqrt{mn}[M(X) - M(Y)]}{\sqrt{mD(X) + nD(Y)}}. \tag{3}$$

Формулы (2) – (3) позволяют приближенно вычислять мощность t–критерия (точность возрастает при увеличении объемов выборок m и n).

О проверке условия равенства дисперсий. Иногда условие б) вытекает из методики получения результатов наблюдений, например, когда с помощью одного и того же прибора или методики m раз измеряют характеристику первого объекта и n раз – второго, а параметры распределения погрешностей измерения при этом не меняются. Однако ясно, что в постановках большинства исследовательских и практических задач нет основании априори предполагать равенство дисперсий.

Целесообразно ли проверять равенство дисперсий статистическими методами, например, как это иногда предлагают, с помощью F-критерия Фишера? Этот критерий основан на нормальности распределений результатов наблюдений. А от нормальности неизбежны отклонения (см. выше). Причем хорошо известно, что в отличие от t-критерия распределение F-критерия Фишера сильно меняется при малых отклонениях от нормальности [8]. Кроме того, F-критерий отвергает гипотезу $D(X) = D(Y)$ лишь при большом различии выборочных дисперсий. Так, для данных [2] о двух группах результатов химических анализов отношение выборочных дисперсий равно 1,95, т.е. существенно отличается от 1. Тем не менее, гипотеза о равенстве теоретических дисперсий принимается при применении F-критерия на 1%-м уровне значимости. Следовательно, при проверке однородности применение F-критерия для предварительной проверки равенства дисперсий с целью обоснования возможности использования критерия Стьюдента нецелесообразно.

Итак, в большинстве технических, экономических, медицинских и иных задач условие б) нельзя считать выполненным, а проверять его перед проверкой однородности нецелесообразно.

Последствия нарушения условия равенства дисперсий. Если объемы выборок m и n велики, то можно показать, что распределение статистики t описывается с помощью только математических ожиданий $M(X)$ и $M(Y)$, дисперсий $D(X)$, $D(Y)$ и отношения объемов выборок, а именно:

$$P(t \le x) \approx \Phi(b_{mn}x - a_{mn}), \tag{4}$$

где a_{mn} определено формулой (3),

$$b_{mn}^2 = \frac{\lambda D(X) + D(Y)}{D(X) + \lambda D(Y)}, \quad \lambda = \frac{m}{n}. \tag{5}$$

Если $b_{mn} \neq 1$, то распределение статистики t отличается от распределения, заданного формулой (2), полученной в предположении равенства дисперсий. Когда $b_{mn} = 1$? В двух случаях – при $m = n$ и при $D(X) = D(Y)$. Таким образом, при больших и равных объемах выборок требовать выполнения условия б) нет необходимости. Кроме того, ясно, что если объемы выборок мало различаются, то b_{mn} близко к 1. Так, для данных [2] о двух группах результатов химических анализов имеем $b^*_{mn} = 0,987$, где b^*_{mn} – оценка b_{mn} , полученная заменой в формуле (5) теоретических дисперсий на их выборочные оценки.

Область применимости традиционного метода проверки однородности с помощью критерия Стьюдента. Подведем итоги рассмотрения t-критерия. Он позволяет проверять гипотезу H_0' о равенстве математических ожиданий, но не гипотезу H_0 о том, что обе выборки взяты из одной и той же генеральной совокупности. Классические условия применимости критерия Стьюдента в подавляющем большинстве технических, экономических, медицинских и иных задач не выполнены. Тем не менее, при больших и примерно равных объемах выборок его можно применять. При конечных объемах выборок традиционный метод носит неустранимо приближенный характер.

Критерий Крамера–Уэлча равенства математических ожиданий. Вместо критерия Стьюдента целесообразно для проверки H_0' использовать критерий Крамера-Уэлча, основанный на статистике

$$T = \frac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} = \frac{\sqrt{mn}(\overline{x} - \overline{y})}{\sqrt{ns_x^2 + ms_y^2}}. \quad (6)$$

Критерий Крамера–Уэлча имеет прозрачный смысл – разность выборочных средних арифметических для двух выборок делится на естественную оценку среднего квадратического отклонения этой разности. Естественность указанной оценки состоит в том, что неизвестные статистику дисперсии заменены их выборочными оценками. Из многомерной центральной предельной теоремы и из теорем о наследовании сходимости [5, гл.4] вытекает, что при росте объемов выборок распределение статистики T Крамера–Уэлча сходится к стандартному нормальному распределению с математическим ожиданием 0 и дисперсией 1. Эта сходимость установлена Г. Крамером [9]. Термин "статистика Крамера–Уэлча" введен нами в [10].

Итак, при справедливости H_0' и больших объемах выборок распределение статистики T приближается с помощью стандартного нормального распределения $\Phi(x)$, из таблиц которого и следует брать критические значения. При $m = n$, как следует из формул (1) и (6), $t = T$. При $m \neq n$ этого равенства нет. В частности, при s_x^2 в формуле (1) стоит множитель $(m-1)$, а в формуле (6)– множитель n.

Если $M(X) \neq M(Y)$, то при больших объемах выборок

$$P(T \leq X) \approx \Phi(x - c_{mn}), \quad (7)$$

где

$$c_{mn} = \frac{\sqrt{mn}[M(X) - M(Y)]}{\sqrt{nD(X) + mD(Y)}}. \quad (8)$$

При $m = n$ или $D(X) = D(Y)$, согласно формулам (3) и (8), $a_{mn} = c_{mn}$, в остальных случаях равенства нет.

Из асимптотической нормальности статистики T, формул (7) и (8) следует, что правило принятия решения для критерия Крамера–Уэлча выглядит так:

– если $|T| \le \Phi^{-1}(1-\alpha/2)$, то гипотеза однородности (равенства) математических ожиданий принимается на уровне значимости α,
– если же $|T| > \Phi^{-1}(1-\alpha/2)$, то гипотеза однородности (равенства) математических ожиданий отклоняется на уровне значимости α.

В прикладной статистике наиболее часто применяется уровень значимости $\alpha = 0,05$. Тогда значение модуля статистики T Крамера–Уэлча надо сравнивать с граничным значением $\Phi^{-1}(1-\alpha/2) = 1,96$

Из сказанного выше следует, что применение критерия Крамера–Уэлча при анализе организационно-экономических данных более обосновано, чем применение критерия Стьюдента. Дополнительное преимущество критерия Крамера–Уэлча по сравнению с критерием Стьюдента – не требуется равенства дисперсий $D(X) = D(Y)$. Распределение статистики не является распределением Стьюдента, однако и распределение статистики t, как показано выше, не является таковым в реальных ситуациях.

Распределение статистики T при объемах выборок $m = n = 6, 8, 10, 12$ и различных функциях распределений выборок $F(X)$ и $G(Y)$ изучено методом статистических испытаний (Монте–Карло). Результаты (частично опубликованы в статьях [11 – 13]) показывают, что даже при таких небольших объемах выборок точность аппроксимации предельным стандартным нормальным распределением вполне удовлетворительна. Поэтому представляется целесообразным во всех тех случаях, когда в соответствии с устаревшими литературными источниками предлагается использовать критерий Стьюдента, заменить его на критерий Крамера–Уэлча. Конечно, такая замена потребует переделки ряда нормативно-технических и методических документов, исправления учебников и учебных пособий для вузов.

5. Непараметрические методы проверки однородности

В большинстве управленческих, технических, экономических, медицинских и иных задач анализа данных в прикладных исследованиях представляет интерес не проверка равенства математических ожиданий или иных характеристик распределения, а обнаружение различия генеральных совокупностей, из которых извлечены выборки, т.е. проверка гипотезы H_0. Методы проверки гипотезы H_0 должны позволять обнаружить не только изменение математического ожидания или дисперсии, но и любые иные изменения функции распределения результатов наблюдений при переходе от одной выборки к другой (увеличение разброса, появление асимметрии и т.

д.). Как установлено выше, методы, основанные на использовании статистик t Стьюдента и T Крамера–Уэлча, не позволяют проверять гипотезу H_0. Априорное предположение о принадлежности функций распределения $F(x)$ и $G(x)$ к какому-либо определенному параметрическому семейству (например, семействам нормальных, логарифмически нормальных, распределений Вейбулла–Гнеденко, гамма-распределений и др.) обычно нельзя достаточно надежно обосновать [4, 5]. Поэтому для проверки H_0 следует использовать методы, пригодные при любом виде $F(x)$ и $G(x)$, т.е. непараметрические методы.

Термин «непараметрический метод» означает, что при использовании этого метода нет необходимости предполагать, что функции распределения результатов наблюдений принадлежат каким-либо определенным параметрическим семействам. Современное представление о методах непараметрической статистики дано в статье [14].

Для проверки гипотезы H_0 разработано много непараметрических методов – критерии Смирнова, типа омега-квадрат (Лемана–Розенблатта), Вилкоксона (Манна–Уитни), Ван-дер-Вардена, Сэвиджа, хи-квадрат и др. [5, 15, 16]. Распределения статистик всех этих критериев при справедливости H_0 не зависят от конкретного вида совпадающих функций распределения $F(x) \equiv G(x)$. Следовательно, таблицами точных и предельных (при больших объемах выборок) распределений статистик этих критериев и их процентных точек [5, 16] можно пользоваться при любых непрерывных функциях распределения результатов наблюдений.

Какой из непараметрических критериев применять? Как известно [8], для выбора одного из нескольких критериев необходимо сравнить их мощности, определяемые видом альтернативных гипотез. Сравнению мощностей критериев посвящена обширная литература.

Хорошо изучены свойства критериев при альтернативной гипотезе сдвига

$$H_{1c} : G(x) = F(x - d), \quad d \neq 0.$$

Критерии Вилкоксона, Ван-дер-Вардена и ряд других ориентированы для применения именно в этой ситуации. Если m раз измеряют характеристику одного объекта и n раз – другого, а функция распределения погрешностей измерения произвольна, но не меняется при переходе от объекта к объекту (это более жесткое требование, чем условие равенства дисперсий), то рассмотрение гипотезы H_{1c} оправдано. Однако в большинстве прикладных исследований нет оснований считать, что функции распределения, соответствующие выборкам, различаются только сдвигом.

Полагаем, что ради адекватности математической модели практической ситуации в качестве альтернативной гипотезы надо рассматривать H_1 (отсутствие абсолютной однородности), а не H_{1c} (наличие сдвига).

Двухвыборочный критерий Вилкоксона. Этот критерий (в литературе его называют также критерием Манна–Уитни), как показано в [17, 18], предназначен для проверки гипотезы

$$H_{0m} : P(X < Y) = 1/2,$$

где X– случайная величина, распределенная как элементы первой выборки, а Y – случайная величина, распределенная как элементы второй выборки. Альтернативой является отрицание H_{0m}: вероятность $P(X < Y)$ отлична от 0,5. Это – непараметрическая гипотеза. Но из нее не следует, что функции распределения двух выборок совпадают. Обратное, конечно, верно: если X и Y одинаково распределены, то $P(X < Y) = 1/2$, т.е. медиана распределения разности $X - Y$ равна 0.

Критерий Вилкоксона – один из самых известных инструментов непараметрической статистики (наряду с критериями на основе статистик типа Колмогорова–Смирнова, омега-квадрат и коэффициентами ранговой корреляции). Свойствам этого критерия и таблицам его критических значений уделяется место во многих монографиях и статьях по математической и прикладной статистике (см., например, [2, 5, 15, 16]). Однако в литературе имеются и неточные утверждения относительно возможностей критерия Вилкоксона. Так, отдельные авторы полагают, что с его помощью можно обнаружить любое различие между функциями распределения $F(x)$ и $G(x)$. По мнению других, этот критерий нацелен на проверку равенства медиан распределений, соответствующих выборкам. И то, и другое неверно, как подробно показано в [17, 18].

При альтернативной гипотезе, когда функции распределения выборок $F(x)$ и $G(x)$ не совпадают, распределение статистики Вилкоксона зависит от величины $\alpha = P(X < Y)$. Если α отличается от $1/2$, то мощность критерия Вилкоксона стремится к 1, и он отличает нулевую гипотезу $F \equiv G$ от альтернативной. Если же $\alpha = 1/2$, то это не всегда имеет место. При справедливости альтернативной гипотезы сдвига мощность стремится к 1, различие медиан также всегда обнаруживается.

6. Состоятельные критерии проверки абсолютной однородности двух независимых выборок

Естественно потребовать, чтобы рекомендуемый для массового использования в прикладных исследованиях критерий однородности был состоятельным. Это значит, что для любых отличных друг от друга функций распределения $F(x)$ и $G(x)$ (другими словами, при справедливости альтернативной гипотезы H_1) вероятность отклонения гипотезы H_0 должна стремиться к 1 при увеличении объемов выборок m и n. Из перечисленных выше критериев однородности состоятельными являются только критерии Смирнова и типа омега-квадрат.

Критерий Смирнова однородности двух независимых выборок. Он был предложен членом-корреспондентом АН СССР Н.В. Смирновым в 1939 г. (см. справочник [2]). Единственное ограничение – функции распределения $F(x)$ и $G(x)$ должны быть непрерывными. Согласно Л.Н. Большеву и Н.В. Смирнову [2] значение эмпирической функции распределения в точке x равно доле результатов наблюдений в выборке, меньших x. Критерий Смирнова основан на использовании эмпирических функций распределения $F_m(x)$ и $G_n(x)$, построенных по первой и второй выборкам соответственно. Значение статистики Смирнова

$$D_{m,n} = \sup_x |F_m(x) - G_n(x)|$$

сравнивают с соответствующим критическим значением (см., например, [2]) и по результатам сравнения принимают или отклоняют гипотезу H_0 о совпадении (однородности) функций распределения. Практически значение статистики $D_{m,n}$ рекомендуется согласно монографии [2] вычислять по формулам

$$D^+_{m,n} = \max_{1\le r\le n}\left[\frac{r}{n} - F_m(y'_r)\right] = \max_{1\le s\le m}\left[G_n(x'_s) - \frac{s-1}{m}\right], \tag{9}$$

$$D^-_{m,n} = \max_{1\le r\le n}\left[F_m(y'_r) - \frac{r-1}{n}\right] = \max_{1\le s\le m}\left[\frac{s}{m} - G_n(x'_s)\right], \tag{10}$$

$$D_{m,n} = \max\left(D^+_{m,n}, D^-_{m,n}\right), \tag{11}$$

где $x'_1 < x'_2 < \ldots < x'_m$ – элементы первой выборки $x_1, x_2, \ldots, x_m$, переставленные в порядке возрастания, а $y'_1 < y'_2 < \ldots < y'_n$ – элементы второй выборки $y_1, y_2, \ldots, y_n$, также переставленные в порядке возрастания. Поскольку функции распределения $F(x)$ и $G(x)$ предполагаются непрерывными, то вероятность совпадения каких-либо выборочных значений равна 0. Рекомендации по нахождению численного значения статистики Смирнова на основе выборочных данных даны в статье [19].

Разработаны алгоритмы и компьютерные программы, позволяющие рассчитывать точные распределения, процентные точки и достигаемый уровень значимости для двухвыборочной статистики Смирнова, составлены подробные таблицы (см., например, методику [20], содержащую описание алгоритмов, тексты программ и подробные таблицы критических значений).

Однако у критерия Смирнова есть и недостатки. Его распределение сосредоточено в сравнительно небольшом числе точек. Ясно, что принимаемые этой статистикой значения пропорциональны величине $1/L$, где L – наименьшее общее кратное объемов выборок m и n. Поэтому функция распределения растет большими скачками. Как следствие, не удается выдержать заданный уровень значимости. Реальный (другими словами, истинный) уровень значимости может значительно, даже в несколько раз отличаться от номинального (подробному обсуждению неклассического феномена существенного отличия реальных уровней значимости непараметрических критериев от номинальных посвящены работы [11, 12]).

При больших объемах выборок можно воспользоваться доказанной Н.В. Смирновым в 1939 г. теоремой: в случае совпадения непрерывных функций распределения элементов двух независимых выборок

$$\lim_{m\to\infty, n\to\infty} P\left\{\sqrt{\frac{mn}{m+n}} D_{m,n} < y\right\} = K(y),$$

где $K(y)$ – функция распределения Колмогорова, заданная формулой

$$K(y) = \sum_{k=-\infty}^{\infty} (-1)^k \exp\{-2k^2 y^2\}.$$

Критерий типа омега-квадрат (Лемана-Розенблатта). Статистика критерия типа омега-квадрат для проверки однородности двух независимых выборок имеет вид:

$$A = \frac{mn}{m+n} \int_{-\infty}^{\infty} (F_m(x) - G_n(x))^2 \, dH_{m+n}(x), \qquad (12)$$

где $H_{m+n}(x)$ – эмпирическая функция распределения, построенная по объединенной выборке. Легко видеть, что

$$H_{m+n}(x) = \frac{m}{m+n}F_m(x) + \frac{n}{m+n}G_n(x).$$

Статистика A типа омега-квадрат была предложена Э. Леманом в 1951 г., изучена М. Розенблаттом в 1952 г., а затем и другими исследователями. Она зависит лишь от рангов элементов двух выборок в объединенной выборке. Статистика A представляется в виде (см., например, [2]):

$$A = \frac{1}{mn(m+n)}\left[m\sum_{i=1}^{m}(r_i - i)^2 + n\sum_{j=1}^{n}(s_j - j)^2\right] - \frac{4mn-1}{6(m+n)},$$

где r_i – ранг x'_i и s_j – ранг y'_j в общем вариационном ряду, построенном по объединенной выборке.

Правила принятия решений при проверке однородности двух выборок на основе статистик Смирнова и типа омега-квадрат, т.е. таблицы критических значений в зависимости от уровней значимости и объемов значимости приведены, например, в таблицах [2]. При достаточно больших объемах выборок правило принятия решения формулируется просто: если наблюдаемое значение статистики меньше соответствующего квантиля предельного распределения, гипотеза однородности принимается, в противном случае отклоняется.

Известно [5], что

$$\lim_{m\to\infty, n\to\infty} P\left\{A < x\right\} = a_1(x)$$

(в обозначениях [2]), где $a_1(x)$ – предельная функция распределения классической статистики омега-квадрат (Крамера–Мизеса–Смирнова), используемой для проверки согласия эмпирического распределения с заданным теоретическим.

Рекомендации по выбору критерия однородности. Для критерия типа омега-квадрат (Лемана–Розенблатта) нет выраженного эффекта различия между номинальными и реальными уровнями значимости. Поэтому мы *рекомендуем для проверки однородности функций распределения (гипотеза H_0) применять статистику A типа омега-квадрат. Если методическое, табличное или программное обеспечение для статистики Лемана–Розенблатта отсутствует, рекомендуем использовать критерий Смирнова. Для проверки однородности математических ожиданий* *(гипотеза H'_0) целесообразно применять критерий Крамера–Уэлча.* По нашему мнению, статистики Стьюдента, Вилкоксона и др. допустимо использовать лишь в отдельных частных случаях, рассмотренных выше.

7. Проблемы разработки и применения непараметрических критериев

Обсудим четыре проблемы, связанные с разработкой и применением непараметрических критериев.

О критериях Колмогорова и Смирнова. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат, разработанные в 1930-е годы, являются основой непараметрической математической статистики. Они продолжают быть образцом при разработке новых методов и широко применяются на практике. Используются термины «статистики типа Колмогорова–Смирнова», «статистики типа омега-квадрат». Встречается в литературе и в описании программных продуктов некорректный термин «статистика Колмогорова–Смирнова». А также и другие некорректности – например, пишут о применении критерия Колмогорова для проверки согласия с семейством нормальных распределений. В последнем из упомянутых случаев речь отнюдь не только о терминологии – часты грубые ошибки при применении непараметрических критериев математической статистики, приводящие к неправильным управленческим решениям [21].

В литературе, особенно переводной, иногда используют термин «критерий Колмогорова–Смирнова» по отношению к процедурам проверки непараметрических статистических гипотез, в частности, проверки однородности двух независимых выборок. Однако анализ публикаций академика А.Н. Колмогорова (1903 – 1987) и члена-корреспондента АН СССР Н.В. Смирнова (1900 – 1966), проведенный в статьях [21, 22], свидетельствует о том, что такого критерия **не существует**. У А.Н. Колмогорова и Н.В. Смирнова нет совместных работ, они никогда не изучали одновременно один и тот же статистический критерий. Право на существование имеет лишь термин "критерий типа Колмогорова–Смирнова". Так, критерий проверки однородности двух независимых выборок, основанный на формулах (9) - (11) – это критерий Смирнова. Можно сказать, что это "критерий типа Колмогорова–Смирнова". Критерий (12) – это критерий Лемана–Розенблатта. Можно сказать, что это "критерий типа омега-квадрат".

Предельная теория непараметрических статистик. Предельные (при безграничном росте объемов выборок) распределения непараметрических статистик, предназначенных для проверки однородности двух независимых выборок, находились различными методами. Один из наиболее общих – "принцип инвариантности"[23]. Нами разработан метод аппроксимации ступенчатыми функциями, с его помощью получен ряд необходимых и достаточных условий [24].

Анализ совпадающих наблюдений. Обычно предполагается, что функции распределения $F(x)$ и $G(x)$ непрерывны и строго возрастают. Из непрерывности этих функций следует, что с вероятностью 1 все $m+n$ результатов наблюдений различны. При рассмотрении реальных статистических данных иногда наблюдаются совпадения результатов наблюдений, но сам факт их наличия – свидетельство нарушений предпосылок только что описанной базовой математической модели. В статье [25] мы предлагаем вероятностно-статистическую модель, объясняющую появление совпадений и дающую алгоритмы анализа совпадений. Эта модель основана на предположении о появлении совпадений данных в результате "слипания"мало различающихся результатов наблюдений. Мы предлагаем добавить малую поправку к каждому элементу совпадающей группы результатов наблюдений и в результате получить выборку без совпадений, для которой рассчитать значение ранговой статистики. Рассмотрев различные варианты поправок, получаем "облако"значений ранговой статистики. Анализ этого "облака" позволяет получить статистические выводы.

Проверка однородности связанных выборок. Не всегда можно применять методы проверки однородности двух независимых выборок объемов m и n соответственно. Если проведено $(m+n)$ измерений объемов продаж в $(m+n)$ торговых точках, то эта модель, как правило, адекватна. Если же, например, x_i и y_i – объемы продаж одного и того же товара до и после определенного рекламного воздействия, то рассматриваемую модель применять нельзя (при этом $m=n$), поскольку очевидно, что анализируемые объемы продаж определяются не только и не столько рекламным воздействием, сколько особенностями конкретной торговой точки (ее расположением, продолжительностью работы, репутацией и т.д.). В подобных случаях используют модель связанных выборок. В ней обычно строят новую выборку $z_i = x_i - y_i$ и используют статистические методы анализа одной выборки, а не двух. Методы проверки однородности для связанных выборок рассматриваются в [26, 27]. Они во многом аналогичны соответствующим методы проверки однородности двух независимых выборок.

8. Заключение

В настоящей статье дана сводка основных результатов, касающихся методов проверки однородности двух независимых выборок. Она позволяет системно анализировать многообразие таких методов с целью выбора наиболее адекватного для обработки конкретных данных. За подробностями и примерами адресуем к обширному списку литературных источников, большинство из которых легко доступны.

Библиографический список

1. *Student.* The probable error of a mean // Biometrika. –1908. –№6 (1). – P.1-25.
2. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
3. *Орлов А.И., Друянова Г.Б.* Непараметрическое оценивание коэффициентов вариации технических характеристик и показателей качества // Надежность и контроль качества. – 1987. –№7. – С.10-16.
4. *Орлов А.И.* Распределения реальных статистических данных не являются нормальными // Научный журнал КубГАУ. – 2016. – №117. – С.71-90.
5. *Орлов А.И.* Прикладная статистика. – М.: Экзамен, 2006. – 671 с.
6. *Орлов А.И.* О проверке однородности двух независимых выборок // Заводская лаборатория. Диагностика материалов. – 2003. – Т.69, №1. – С.55-60.
7. *Орлов А.И.* Проверка статистической гипотезы однородности математических ожиданий двух независимых выборок: критерий Крамера-Уэлча вместо критерия Стьюдента // Научный журнал КубГАУ. – 2015. – №110. – С.197-218.
8. *Боровков А.А.* Математическая статистика. – М.: Наука, 1984. – 472 с.
9. *Крамер Г.* Математические методы статистики (Пер. с англ.). – М.: Мир, 1975. – изд.2. – 648 с.
10. *Орлов А.И.* О применении статистических методов в медико-биологических исследованиях // Вестник Академии медицинских наук СССР. – 1987. – №2. – С. 88-94.
11. *Камень Ю.Э., Камень Я.Э., Орлов А.И.* Реальные и номинальные уровни значимости в задачах проверки статистических гипотез // Заводская лаборатория. – 1986. – Т.52, №12. – С.55-57.

12. *Орлов А.И.* Реальные и номинальные уровни значимости при проверке статистических гипотез // Научный журнал КубГАУ. – 2015. – №114. – С. 42-54.

13. *Орлов А.И.* Метод статистических испытаний в прикладной статистике // Заводская лаборатория. Диагностика материалов. – 2019. – Т.85, №5. – С. 67-79.

14. *Орлов А.И.* Современное состояние непараметрической статистики // Научный журнал КубГАУ. – 2015. – №106. – С. 239-269.

15. *Гаек Я., Шидак З.* Теория ранговых критериев (Пер. с англ.). – М.: Наука, 1971. – 376 с.

16. *Холлендер М., Вульф Д.* Непараметрические методы статистики. – М.: Финансы и статистика, 1983. – 518 с.

17. *Орлов А.И.*. Какие гипотезы можно проверять с помощью двухвыборочного критерия Вилкоксона? // Заводская лаборатория. Диагностика материалов. – 1999. – Т.65, №1. – С.51-55.

18. *Орлов А.И.* Двухвыборочный критерий Вилкоксона – анализ двух мифов // Научный журнал КубГАУ. – 2014. – №104. – С. 91-111.

19. *Орлов А.И.* Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. – 2012. – Т.78, №11. – С.66-70.

20. *Орлов А.И., Миронова Н.Г., Фомин В.Н., Черномордик О.М.* Методика. Проверка однородности двух выборок параметров продукции при оценке ее технического уровня и качества – М.: ВНИИСтандартизации, 1987. – 116 с.

21. *Орлов А.И.* Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении // Научный журнал КубГАУ. – 2014. – №97. – С. 32-45.

22. *Орлов А.И.* О критериях Колмогорова и Смирнова // Заводская лаборатория. Диагностика материалов. – 1995. – Т.61, №7. – С.59-61.

23. *Биллингсли П.* Сходимость вероятностных мер. – М.: Наука, 1977. – 353 с.

24. *Орлов А.И.* Предельная теория непараметрических статистик // Научный журнал КубГАУ. – 2014. – №100. – С. 31-52.

25. *Орлов А.И.* Модель анализа совпадений при расчете непараметрических ранговых статистик // Заводская лаборатория. Диагностика материалов. – 2017. – Т.83, №11. – С. 66-72.

26. *Орлов А.И.* Методы проверки однородности связанных выборок // Заводская лаборатория. Диагностика материалов. – 2004. – Т.70, №7. – С.57-61.

27. *Орлов А.И.* О проверке однородности связанных выборок // Научный журнал КубГАУ. – 2016. – №123. – С. 708-726.

The manifold of methods for testing the homogeneity of two independent samples

A.I. Orlov

Bauman Moscow State Technical University, Institute of high statistical technologies and econometrics; Moscow, Russia; prof-orlov@mail.ru; http://orlovs.pp.ru; +7(916)8305117; 123104, Moscow, the Sytinsky lane, house 7/14, apartment 14.

Abstract. *A summary of the main results concerning the methods for testing the homogeneity of two independent samples is given. It allows you to systematically analyze the manifold of such methods in order to select the most appropriate for processing specific data. Based on the basic probabilistic-statistical model, the main statements of the problem of testing the homogeneity of two independent samples are formulated. A comparative analysis of the Student and Cramer-Welch tests, designed to test the homogeneity of mathematical expectations, is given. From non-parametric methods for testing homogeneity, the criteria of Wilcoxon, Smirnov, Lehmann-Rosenblatt are considered. To verify absolute homogeneity, it is recommended to use the Lehmann-Rosenblatt criterion. The problems of the development and application of non-parametric tests are discussed.*

Key words: *testing of statistical hypotheses, independent samples, homogeneity of characteristics, absolute homogeneity, Cramer-Welch test, Wilcoxon test, Smirnov test, Lehmann-Rosenblatt test.*

УДК 519.2

Стационарные характеристики решений систем линейных дифференциальных уравнений нейтрального типа с кратными запаздываниями и случайными возмущениями

И.Е. Полосков

Пермский государственный национальный исследовательский университет, г. Пермь, Россия; polosk@psu.ru; +7342 2396560, 614990, г. Пермь, ул. Букирева, 15, ПГНИУ, кафедра высшей математики

Аннотация. *Схема С. Гийюзика, предназначенная для получения ковариационных функций и спектральных плотностей решения (в установившемся состоянии) линейного стохастического дифференциального уравнения с одним постоянным запаздыванием распространяется на новую область применения, а именно, на линейные системы стохастических дифференциальных уравнений нейтрального типа с кратными запаздываниями. Полученные аналитические соотношения для матриц спектральных плотностей использованы для численного расчета (в среде пакета* Mathematica*) элементов матриц соответствующих ковариационных функций.*
Ключевые слова: *стохастический анализ, стохастическое обыкновенное дифференциальное уравнение, линейная динамическая система, запаздывание, стационарный шум, спектральная плотность, ковариационная функция.*

1. Введение

В последние десятилетия дифференциальные уравнения с запаздыванием (ДУсЗ) стали мощным инструментом для моделирования пространственно распределенных систем. В этих системах геометрия часто такова, что можно заменить распространяющийся эффект замедленной версией этого эффекта. Таким образом, вместо уравнения в частных производных без запаздывания можно использовать обыкновенное ДУсЗ (ОДУсЗ). Это оправдано, когда запаздывание соразмерно или намного больше, чем другие временные масштабы в системе. Детерминированные ОДУсЗ могут генерировать несколько различных типов асимптотического поведения, таких как наличие точек равновесия, предельных циклов и хаоса. Такие уравнения также могут описывать мультистабильность. Такое поведение позволило использовать ОДУсЗ во многих областях. Например, эти уравнения применялись для моделирования оптических устройств, динамики населения, физиологических систем, нейронных сетей, экономических явлений и химической кинетики [6, 12, 18, 19, 21].

При моделировании систем, которые не оказывают заметного влияния на окружающую среду, стохастические переменные часто используются для учета колебаний окружающей среды, что приводит к стохастическим обыкновенным дифференциальным уравнения с запаздыванием. Модели, поведение которых описывают с помощью СОДУсЗ, применяются во многих областях, таких как физиология и оптика.

Известно, что случайные возмущения оказывают значительное влияние на системы, описываемые ОДУ без запаздывания, приводя к качественным изменениям поведения системы. То же самое ожидается и от использования СОДУсЗ, и это может быть проверено численно. Однако аналитические инструменты для СОДУсЗ недостаточны, поскольку методы, которые используются для изучения СОДУ без запаздывания в общем не могут быть непосредственно применены к СОДУсЗ. Значительная работа была проделана в отношении анализа экспоненциальной устойчивости СОДУсЗ и поиска условий существования гладких плотностей вероятности. Однако, как правило, нет способа оценить плотности вероятностей векторов состояния, изменение которых описывается СОДУсЗ.

Чтобы преодолеть трудности учета эффекта запаздывания, используют несколько подходов. Самый простой – это анализ стационарных режимов в линейных системах.

Несколько задач для СОДУ с одним постоянным запаздыванием были решены С. Гийюзиком (S. Guillouzic), Т.Д. Франком и У. Кюхлером [7–11, 15–17, 20]. В частности, С. Гийюзик предложил интересную и очень полезную схему расчета ковариационных функций и спектральных плотностей в установившемся состоянии, а также параметров одномерных стационарных гауссовских распределений. Кроме того, Т.Д. Франк использовал эту схему для получения двухточечной плотности вероятности искомого решения в моменты времени, разделенные запаздыванием. Заметим, что другой способ вывода ФПК–уравнения для одномерной линейной системы СОДУ с постоянными коэффициентами, с одним или несколькими постоянными запаздываниями и стационарным белым шумом на входе предложен в работе [13, 14].

Ранее в ряде наших работ [2–4] были рассмотрены некоторые задачи вычисления стационарных характеристик (спектральных плотностей и ковариационных функций) систем линейных стохастических дифференциальных уравнений (обыкновенных, нейтральных, в частных производных) с постоянными (кратными) запаздываниями. Сначала схема С. Гийюзика была распространена на отдельные уравнения: 1) с кратными запаздываниями; 2) нейтрального типа; 3) высших порядков. Далее эта схема была применена для анализа установившихся колебаний в линейной стохастической системе, которая возмущается стационарными шумами, дифференцируемыми в среднем квадратическом. Целью исследования являлось построение матрицы спектральных плотностей вектора состояния рассматриваемой системы. В последней из указанных работ решалась задача построения спектральной плотности стационарного случайного поля – решения гиперболического уравнения с двумя постоянными коэффициентами и стационарным случайным входом. Ниже схема С. Гийюзика распространяется на еще одну новую область применения, а именно, на линейные системы стохастических дифференциальных уравнений нейтрального типа с кратными запаздываниями. Полученные аналитические соотношения для матриц спектральных плотностей были использованы для численного расчета (в среде пакета Mathematica [5]) элементов матриц соответствующих ковариационных функций.

2. Постановка задачи

Рассмотрим следующую систему линейных ОДУ нейтрального типа с q кратными запаздываниями и случайными возмущениями:

$$\boldsymbol{X}'(t) = \sum_{\ell=0}^{q} \mathcal{A}_\ell\, \boldsymbol{X}(t-\tau_\ell) + \sum_{k=1}^{q} \mathcal{B}_k\, \boldsymbol{X}'(t-\tau_k) + \mathcal{H}\, \boldsymbol{V}(t), \tag{1}$$

где t – время; $\tau_k = \tau\, k$ – запаздывания, $k = 1, 2, ..., q$, $q \in \mathbb{N}$, $\tau = \text{const} > 0$; $\boldsymbol{X}(t) = \{X_1(t), X_2(t), ..., X_n(t)\}^{\mathsf{T}}$ – вектор состояния; $\boldsymbol{V}(t) = \{V_i(t)\} \in \mathbb{R}^m$ – центрированный стационарный векторный случайный процесс с независимыми компонентами, известными автоковариационными функциями $\mathcal{C}_{V_jV_j}(\theta)$ и спектральными плотностями $\mathcal{S}_{V_jV_j}(\omega)$ процессов $V_j(t)$ ($j = 1, 2, ..., m$), являющимися компонентами диагональных матриц $\mathcal{C}_{VV}(\theta)$ и $\mathcal{S}_{VV}(\omega)$ соответственно и связанных формулами Винера–Хинчина [1]:

$$\mathcal{C}_{V_jV_j}(\theta) = \int\limits_{-\infty}^{+\infty} e^{\boldsymbol{i}\omega\theta}\, \mathcal{S}_{V_jV_j}(\omega)\, d\omega, \quad \mathcal{S}_{V_jV_j}(\omega) = \frac{1}{2\,\pi} \int\limits_{-\infty}^{+\infty} e^{-\boldsymbol{i}\omega\theta}\, \mathcal{C}_{V_jV_j}(\theta)\, d\theta;$$

$\mathcal{A}_\ell = \{a_{\ell ij}\}$, $\mathcal{B}_k = \{b_{kij}\}$ и $\mathcal{H} = \{h_{ij}\}$ – постоянные матрицы соответствующей размерности; T – символ транспонирования матрицы (вектора); $\mathbb{E}[\cdot]$ – оператор математического ожидания.

Предположим, что коэффициенты уравнений (1) таковы, что при $t \to +\infty$ после переходного режима в системе, описываемой этими уравнениями, устанавливается стационарный режим, при котором $\boldsymbol{X}(t)$, как и $\boldsymbol{V}(t)$, будет стационарным СП с нулевым вектором математических ожиданий $\boldsymbol{m}_X = \mathbb{E}[\boldsymbol{X}(t)] = \boldsymbol{0}$, неизвестными матрицами ковариационных функцией $\boldsymbol{\mathcal{C}}_{XX}(\omega) = \{\mathcal{C}_{X_iX_k}(\omega)\}$ и спектральных плотностей $\boldsymbol{\mathcal{S}}_{XX}(\omega) = \{\mathcal{S}_{X_iX_k}(\omega)\}$, а оба эти векторных процесса – стационарно связанными, которые характеризуются матрицами ковариационных функцией $\boldsymbol{\mathcal{C}}_{XV}(\omega) = \{\mathcal{C}_{X_iV_j}(\omega)\}$ и спектральных плотностей $\boldsymbol{\mathcal{S}}_{XV}(\omega) = \{\mathcal{S}_{X_iV_j}(\omega)\}$, где

$$\mathcal{C}_{X_iX_k}(\theta) = \int\limits_{-\infty}^{+\infty} e^{\boldsymbol{i}\omega\theta}\,\boldsymbol{\mathcal{S}}_{X_iX_k}(\omega)\,d\omega, \quad \boldsymbol{\mathcal{S}}_{X_iX_k}(\omega) = \frac{1}{2\,\pi}\int\limits_{-\infty}^{+\infty} e^{-\boldsymbol{i}\omega\theta}\,\mathcal{C}_{X_iX_k}(\theta)\,d\theta,$$

$$\mathcal{C}_{X_iV_j}(\theta) = \int\limits_{-\infty}^{+\infty} e^{\boldsymbol{i}\omega\theta}\,\boldsymbol{\mathcal{S}}_{X_iV_j}(\omega)\,d\omega, \quad S_{X_iV_j}(\omega) = \frac{1}{2\,\pi}\int\limits_{-\infty}^{+\infty} e^{-\boldsymbol{i}\omega\theta}\,C_{X_iV_j}(\theta)\,d\theta,$$

$$i, k = 1, 2, ..., n, \quad j = 1, 2, ..., m.$$

Вследствие утверждений корреляционного анализа [1] и сделанных предположений ясно, что для полного описания гауссовского стационарного векторного случайного процесса $\boldsymbol{X}(t)$ достаточно вычислить матрицу спектральных плотностей $\boldsymbol{\mathcal{S}}_{XX}(\omega)$, а затем через нее и все остальные необходимые характеристики этого векторного случайного процесса.

3. Получение необходимых характеристик

Ниже для решения поставленной задачи дополнительно будем использовать следующие соотношения:

$$\mathcal{C}_{X_iX_k}(t_1, t_2) = \mathbb{E}\big[\big\{X_i(t_1) - m_{X_i}(t_1)\big\}\big\{X_k(t_2) - m_{X_k}(t_2)\big\}\big]\,;$$

$$\mathcal{C}_{X_iV_j}(t_1, t_2) = \mathbb{E}\big[\big\{X_i(t_1) - m_{X_i}(t_1)\big\}\big\{V_j(t_2) - m_{V_j}(t_2)\big\}\big]\,;$$

$$\mathcal{C}_{V_jV_j}(t_1, t_2) = \mathbb{E}\big[\big\{V_j(t_1) - m_{V_j}(t_1)\big\}\big\{V_j(t_2) - m_{V_j}(t_2)\big\}\big]\,;$$

$$\frac{\partial}{\partial t_1} = -\frac{\partial}{\partial \theta}; \qquad \frac{\partial}{\partial t_2} = \frac{\partial}{\partial \theta};$$

$$e^{\pm\boldsymbol{i}\omega\tau}\,\mathcal{S}_{X_iX_k}(\omega) = \frac{1}{2\,\pi}\int\limits_{-\infty}^{+\infty} e^{-\boldsymbol{i}\omega\theta}\,\mathcal{C}_{X_iX_k}(\theta \pm \tau)\,d\theta;$$

$$\boldsymbol{i}\,\omega\,\mathcal{S}_{X_iX_k}(\omega) = \frac{1}{2\,\pi}\int\limits_{-\infty}^{+\infty} e^{-\boldsymbol{i}\omega\theta}\,\mathcal{C}'_{X_iX_k}(\theta)\,d\theta, \quad r \in \mathbb{N},$$

где $\boldsymbol{\mathcal{C}}_{XX}(t_1,t_2) = \{\mathcal{C}_{X_iX_k}(t_1,t_2)\}$, $\boldsymbol{\mathcal{C}}_{XV}(t_1,t_2) = \{\mathcal{C}_{X_iV_j}(t_1,t_2)\}$, $\boldsymbol{\mathcal{C}}_{VV}(t_1,t_2) = \operatorname{diag}\big(\mathcal{C}_{V_jV_j}(t_1,t_2)\big)$ – матрицы авто- и кросс-ковариационных функций векторных случайных процессов $\boldsymbol{X}(t)$ и $\boldsymbol{V}(t)$ в переходном режиме ($\theta = t_2 - t_1$ в стационарном). Кроме указанных, далее используются еще следующие символы: $\boldsymbol{\mathcal{K}}_{ZZ} = \{\mathcal{K}_{Z_iZ_j}\}$ $(\mathcal{K}_{Z_iZ_j} = \mathbb{E}\big[(Z_i - m_{Z_i})(Z_j - m_{Z_j})\big])$ – матрица ковариаций;

$$\mathcal{N}(\boldsymbol{z};\boldsymbol{m}_Z,\boldsymbol{\mathcal{K}}_{ZZ}) = \frac{1}{(2\,\pi)^{r/2}\,|\boldsymbol{\mathcal{K}}_{ZZ}|^{1/2}}\,\exp\Big[-\frac{1}{2}\,(\boldsymbol{z}-\boldsymbol{m}_Z)^{\mathsf{T}}\,\boldsymbol{\mathcal{K}}_{ZZ}^{-1}\,(\boldsymbol{z}-\boldsymbol{m}_Z)\Big]$$

– невырожденная многомерная нормальная плотность случайного вектора $\boldsymbol{Z}$ $(i,\, j = 1,\, 2,\, ...,\, r$.

Применим идею С. Гийюзика для выполнения общей части выкладок по решению поставленной задачи. Для этого сначала из обеих частей уравнений (1) вычтем векторно-матричное уравнение для функции математического ожидания, имеющее следующий вид:

$$\boldsymbol{m}'_X(t) = \sum_{\ell=0}^{q}\mathcal{A}_\ell\,\boldsymbol{m}_X(t-\tau_\ell) + \sum_{k=1}^{q}\mathcal{B}_k\,\boldsymbol{m}'_X(t-\tau_k), \tag{2}$$

что дает уравнение для центрированного векторного процесса $\boldsymbol{X}(t)$:

$$\mathring{\boldsymbol{X}}'(t) = \sum_{\ell=0}^{q}\mathcal{A}_\ell\,\mathring{\boldsymbol{X}}(t-\tau_\ell) + \sum_{k=1}^{q}\mathcal{B}_k\,\mathring{\boldsymbol{X}}'(t-\tau_k) + \mathcal{H}\,\boldsymbol{V}(t) \tag{3}$$

$(\mathring{\boldsymbol{V}}(t) = \boldsymbol{V}(t))$. Далее обе части уравнений (3) для моментов времени $t = t_1$ и $t = t_2$ умножим на $\mathring{\boldsymbol{X}}^{\mathsf{T}}(t_2)$ (справа) и $\boldsymbol{V}(t_1)$ (слева после транспонирования системы уравнений) соответственно, а затем усредним. В результате будем иметь следующие уравнения для матриц (пока еще) нестационарных ковариационных функций $\boldsymbol{\mathcal{C}}_{XX}(t_1,t_2)$ и $\boldsymbol{\mathcal{C}}_{VX}(t_1,t_2)$:

$$\begin{aligned}\frac{\partial\boldsymbol{\mathcal{C}}_{XX}(t_1,t_2)}{\partial t_1} &= \sum_{\ell=0}^{q}\mathcal{A}_\ell\,\boldsymbol{\mathcal{C}}_{XX}(t_1-\ell\,\tau,t_2) + \\ &\qquad + \sum_{k=1}^{q}\mathcal{B}_k\,\frac{\partial\boldsymbol{\mathcal{C}}_{XX}(t_1-k\,\tau,t_2)}{\partial t_1} + \mathcal{H}\,\boldsymbol{\mathcal{C}}_{VX}(t_1,t_2),\end{aligned}$$

$$\begin{aligned}\frac{\partial\boldsymbol{\mathcal{C}}_{VX}(t_1,t_2)}{\partial t_2} &= \sum_{\ell=0}^{q}\boldsymbol{\mathcal{C}}_{VX}(t_1,t_2-\ell\,\tau)\,\mathcal{A}_\ell^{\mathsf{T}} + \\ &\qquad + \sum_{k=1}^{q}\frac{\partial\boldsymbol{\mathcal{C}}_{VX}(t_1,t_2-k\,\tau)}{\partial t_2}\,\mathcal{B}_k^{\mathsf{T}} + \boldsymbol{\mathcal{C}}_{VV}(t_1,t_2)\,\mathcal{H}^{\mathsf{T}}.\end{aligned}$$

В установившемся режиме с учетом соотношений, приведенных выше, эти уравнения можно записать так:

$$-\boldsymbol{\mathcal{C}}'_{XX}(\theta) = \sum_{\ell=0}^{q} \mathcal{A}_\ell\, \boldsymbol{\mathcal{C}}_{XX}(\theta + \ell\,\tau) - \sum_{k=1}^{q} \mathcal{B}_k\, \boldsymbol{\mathcal{C}}'_{XX}(\theta + k\,\tau) + \mathcal{H}\, \boldsymbol{\mathcal{C}}_{VX}(\theta),$$

$$\boldsymbol{\mathcal{C}}'_{VX}(\theta) = \sum_{\ell=0}^{q} \boldsymbol{\mathcal{C}}_{VX}(\theta - \ell\,\tau)\, \mathcal{A}_\ell^{\mathsf{T}} + \sum_{k=1}^{q} \boldsymbol{\mathcal{C}}'_{VX}(\theta - k\,\tau)\, \mathcal{B}_k^{\mathsf{T}} + \boldsymbol{\mathcal{C}}_{VV}(\theta)\, \mathcal{H}^{\mathsf{T}}.$$

Если теперь левые и правые части последних уравнений умножить на $e^{-\boldsymbol{i}\omega\theta}/(2\pi)$, а затем проинтегрировать по θ от $-\infty$ до $+\infty$, то получим систему линейных алгебраических уравнений для неизвестных матриц спектральных плотностей $\boldsymbol{\mathcal{S}}_{XX}(\omega)$ и $\boldsymbol{\mathcal{S}}_{VX}(\omega)$:

$$-\boldsymbol{i}\,\omega\, \boldsymbol{\mathcal{S}}_{XX}(\omega) = \sum_{\ell=0}^{q} e^{\boldsymbol{i}\omega\ell\tau}\, \mathcal{A}_\ell\, \boldsymbol{\mathcal{S}}_{XX}(\omega) - \boldsymbol{i}\,\omega \sum_{k=1}^{q} e^{\boldsymbol{i}\omega k\tau}\, \mathcal{B}_k\, \boldsymbol{\mathcal{S}}_{XX}(\omega) + \mathcal{H}\, \boldsymbol{\mathcal{S}}_{VX}(\omega),$$

$$\boldsymbol{i}\,\omega\, \boldsymbol{\mathcal{S}}_{VX}(\omega) = \sum_{\ell=0}^{q} e^{-\boldsymbol{i}\omega\ell\tau}\, \boldsymbol{\mathcal{S}}_{VX}(\omega)\, \mathcal{A}_\ell^{\mathsf{T}} + \boldsymbol{i}\,\omega \sum_{k=1}^{q} e^{-\boldsymbol{i}\omega k\tau}\, \boldsymbol{\mathcal{S}}_{VX}(\omega)\, \mathcal{B}_k^{\mathsf{T}} + \boldsymbol{\mathcal{S}}_{VV}(\omega)\, \mathcal{H}^{\mathsf{T}},$$

или

$$\Big[\boldsymbol{i}\,\omega \sum_{k=1}^{q} e^{\boldsymbol{i}\omega k\tau}\, \mathcal{B}_k - \boldsymbol{i}\,\omega\, \mathcal{I} - \sum_{\ell=0}^{q} e^{\boldsymbol{i}\omega\ell\tau}\, \mathcal{A}_\ell\Big]\, \boldsymbol{\mathcal{S}}_{XX}(\omega) = \mathcal{H}\, \boldsymbol{\mathcal{S}}_{VX}(\omega),$$

$$\boldsymbol{\mathcal{S}}_{VX}(\omega)\, \Big[-\boldsymbol{i}\,\omega \sum_{k=1}^{q} e^{-\boldsymbol{i}\omega k\tau}\, \mathcal{B}_k^{\mathsf{T}} + \boldsymbol{i}\,\omega\, \mathcal{I} - \sum_{\ell=0}^{q} e^{-\boldsymbol{i}\omega\ell\tau}\, \mathcal{A}_\ell^{\mathsf{T}}\Big] = \boldsymbol{\mathcal{S}}_{VV}(\omega)\, \mathcal{H}^{\mathsf{T}}$$

($\mathcal{I}$ –единичная матрица), из которой в предположении обратимости соответствующих матриц находим

$$\boldsymbol{\mathcal{S}}_{XX}(\omega) = \mathcal{G}_1(\omega)\, \boldsymbol{\mathcal{S}}_{VV}(\omega)\, \mathcal{G}_2(\omega), \tag{4}$$

где

$$\mathcal{G}_1(\omega) = \Big[\boldsymbol{i}\,\omega \sum_{k=1}^{q} e^{\boldsymbol{i}\omega k\tau}\, \mathcal{B}_k - \boldsymbol{i}\,\omega\, \mathcal{I} - \sum_{\ell=0}^{q} e^{\boldsymbol{i}\omega\ell\tau}\, \mathcal{A}_\ell\Big]^{-1} \mathcal{H},$$

$$\mathcal{G}_2(\omega) = \mathcal{H}^{\mathsf{T}} \Big[-\boldsymbol{i}\,\omega \sum_{k=1}^{q} e^{-\boldsymbol{i}\omega k\tau}\, \mathcal{B}_k^{\mathsf{T}} + \boldsymbol{i}\,\omega\, \mathcal{I} - \sum_{\ell=0}^{q} e^{-\boldsymbol{i}\omega\ell\tau}\, \mathcal{A}_\ell^{\mathsf{T}}\Big]^{-1}.$$

После получения $\boldsymbol{\mathcal{S}}_{XX}(\omega)$ с помощью покомпонентного применения формулы для ковариационных функций можно вычислить матрицу $\boldsymbol{\mathcal{C}}_{XX}(\theta)$, а следовательно, и стационарную матрицу ковариаций $\boldsymbol{\mathcal{K}}_0 = \boldsymbol{\mathcal{C}}_{XX}(0)$, что позволяет построить как одноточечные, так и различные многоточечные (со сдвигом на τ) стационарные нормальные плотности распределения подмножеств компонент центрированного случайного вектора

$$\mathbb{X}^{[\ell]} = \mathrm{col}\big(\boldsymbol{X}_0, \boldsymbol{X}_1, ..., \boldsymbol{X}_\ell\big), \quad \boldsymbol{X}_k = \boldsymbol{X}(t + k\,\tau), \quad k = 0, 1, ..., \ell, \quad \ell \geqslant 0,$$

в частности, в качестве матрицы ковариации $\boldsymbol{\mathcal{K}}_{ZZ}$, где $\boldsymbol{Z} = \mathbb{X}^{[\ell]}$, будем иметь:

$$\boldsymbol{\mathcal{K}}_{ZZ} = \begin{bmatrix} \boldsymbol{\mathcal{K}}_{ZZ}^{[11]} & ... & \boldsymbol{\mathcal{K}}_{ZZ}^{[12]} \\ ... & ... & ... \\ \boldsymbol{\mathcal{K}}_{ZZ}^{[21]} & ... & \boldsymbol{\mathcal{K}}_{ZZ}^{[22]} \end{bmatrix}$$

где

$$\boldsymbol{\mathcal{K}}_{ZZ}^{[11]} = \begin{bmatrix} \boldsymbol{\mathcal{K}}_0 & \boldsymbol{\mathcal{C}}_{XX}(\tau) & \boldsymbol{\mathcal{C}}_{XX}(2\,\tau) \\ \boldsymbol{\mathcal{C}}_{XX}(\tau) & \boldsymbol{\mathcal{K}}_0 & \boldsymbol{\mathcal{C}}_{XX}(\tau) \\ \boldsymbol{\mathcal{C}}_{XX}(2\,\tau) & \boldsymbol{\mathcal{C}}_{XX}(\tau) & \boldsymbol{\mathcal{K}}_0 \end{bmatrix}, \quad \boldsymbol{\mathcal{K}}_{ZZ}^{[22]} = \begin{bmatrix} \boldsymbol{\mathcal{K}}_0 & \boldsymbol{\mathcal{C}}_{XX}(\tau) \\ \boldsymbol{\mathcal{C}}_{XX}(\tau) & \boldsymbol{\mathcal{K}}_0 \end{bmatrix},$$

$$\boldsymbol{\mathcal{K}}_{ZZ}^{[12]} = \begin{bmatrix} \boldsymbol{\mathcal{C}}_{XX}((\ell-1)\,\tau) & \boldsymbol{\mathcal{C}}_{XX}(\ell\,\tau) \\ \boldsymbol{\mathcal{C}}_{XX}((\ell-2)\,\tau) & \boldsymbol{\mathcal{C}}_{XX}((\ell-1)\,\tau) \\ \boldsymbol{\mathcal{C}}_{XX}((\ell-3)\,\tau) & \boldsymbol{\mathcal{C}}_{XX}((\ell-2)\,\tau) \end{bmatrix},$$

$$\boldsymbol{\mathcal{K}}_{ZZ}^{[21]} = \begin{bmatrix} \boldsymbol{\mathcal{C}}_{XX}((\ell-1)\,\tau) & \boldsymbol{\mathcal{C}}_{XX}((\ell-2)\,\tau) & \boldsymbol{\mathcal{C}}_{XX}((\ell-3)\,\tau) \\ \boldsymbol{\mathcal{C}}_{XX}(\ell\,\tau) & \boldsymbol{\mathcal{C}}_{XX}((\ell-1)\,\tau) & \boldsymbol{\mathcal{C}}_{XX}((\ell-2)\,\tau) \end{bmatrix}.$$

4. Пример

Рассмотрим систему второго порядка с двумя кратными запаздываниями, структурные матрицы модели которой имеют вид:

$$\mathcal{A}_0 = \begin{bmatrix} a_{011} & a_{012} \\ a_{021} & a_{022} \end{bmatrix}, \quad \mathcal{A}_1 = \begin{bmatrix} a_{111} & a_{112} \\ a_{121} & a_{122} \end{bmatrix}, \quad \mathcal{A}_2 = \begin{bmatrix} a_{211} & 0 \\ 0 & a_{222} \end{bmatrix},$$

$$\mathcal{B}_1 = \begin{bmatrix} b_{111} & 0 \\ 0 & b_{122} \end{bmatrix}, \quad \mathcal{B}_2 = \begin{bmatrix} b_{211} & 0 \\ 0 & b_{222} \end{bmatrix}, \quad \mathcal{H} = \begin{bmatrix} h_{11} & 0 \\ 0 & h_{22} \end{bmatrix}, \quad \boldsymbol{\mathcal{S}}_{VV} = \begin{bmatrix} s_{11} & 0 \\ 0 & s_{22} \end{bmatrix}.$$

Расчеты производились при следующих значениях параметров:

$$a_{011} = -5.0; \quad a_{012} = 1.0; \quad a_{021} = -2.0; \quad a_{022} = -3.0;$$
$$a_{111} = -1.0; \quad a_{112} = 0.5; \quad a_{121} = -0.5; \quad a_{122} = -1.0;$$
$$a_{211} = 0.2; \quad a_{212} = 0.0; \quad a_{221} = 0.0; \quad a_{222} = 0.1;$$
$$b_{111} = 0.5; \quad b_{112} = 0.0; \quad b_{121} = 0.0; \quad b_{122} = 0.5;$$
$$b_{211} = 0.2; \quad b_{212} = 0.0; \quad b_{221} = 0.0; \quad b_{222} = 0.1;$$

$$h_{11} = 1.0; \quad h_{12} = 0.0; \quad h_{21} = 0.0; \quad h_{22} = 1.5;$$
$$s_{11} = 1.0; \quad s_{22} = 0.5; \quad \tau \in \{1.0, 1.5, 2.0\}.$$

При получении $\mathcal{C}_{X_1X_1}(\theta)$, $\mathcal{C}_{X_1X_2}(\theta)$ и $\mathcal{C}_{X_2X_2}(\theta)$ через $\mathcal{S}_{X_1X_1}(\omega)$, $\mathcal{S}_{X_1X_2}(\omega)$ и $\mathcal{S}_{X_2X_2}(\omega)$ применялись формулы Винера–Хинчина. При этом учитывалась четность действительных функций $\mathcal{S}_{X_1X_1}(\omega)$ и $\mathcal{S}_{X_2X_2}(\omega)$, а также четность действительной части комплексной функции $\mathcal{S}_{X_1X_2}(\omega)$ и нечетность мнимой ее части, что позволило использовать указанные выше формулы в следующей форме:

$$\mathcal{C}_{X_iX_i}(\theta) = 2 \int\limits_0^{+\infty} \mathcal{S}_{X_iX_i}(\omega) \cos\omega\,\theta d\omega, \quad i = 1, 2; \tag{5}$$

$$\mathcal{C}_{X_1X_2}(\theta) = 2 \int\limits_0^{+\infty} \big[\mathrm{Re}\left\{\mathcal{S}_{X_1X_2}(\omega)\right\} \cos\omega\theta - \mathrm{Im}\left\{\mathcal{S}_{X_1X_2}(\omega)\right\} \sin\omega\theta \big]\, d\omega, \tag{6}$$

что при учете свойств функций $\mathcal{C}_{X_1X_1}(\theta)$, $\mathcal{C}_{X_1X_2}(\theta)$ и $\mathcal{C}_{X_2X_2}(\theta)$ при численных расчетах позволило ограничиться таблицей значений этих функций на заданной сетке только для $\theta \geqslant 0$.

Для приближенного вычисления интегралов в формулах (5), (6) применялась замена бесконечного интервала интегрирования $[0, +\infty)$ конечным $[0, L]$ ($L \gg 1$), который разбивался на более мелкие части $[L_{m-1}, L_m]$: $[0, L] = \bigcup_{m=1}^{M} [L_{m-1}, L_m]$, и встроенная функция `NIntegrate` пакета Mathematica с параметром метода `GaussKronrodRule` с автоматическим выбором количества разбиений каждого отрезка интегрирования $[L_{m-1}, L_m]$. Для получения большей точности расчеты велись с 40 значащими цифрами.

Результаты вычислений представлены на рис. 1–7, где на рис. 2, 5, 7 отображено поведение ковариационных функций $\mathcal{C}_{X_1X_1}(\theta)$, $\mathcal{C}_{X_1X_2}(\theta)$ и $\mathcal{C}_{X_2X_2}(\theta)$, а на рис. 1, 6, 3, 4 – спектральных плотностей $\mathcal{S}_{X_1X_1}(\omega)$, $\mathcal{S}_{X_2X_2}(\omega)$ и $\mathcal{S}_{X_1X_2}(\omega)$ (действительная и мнимая части). На всех этих рисунках запаздываниям $\tau = 1.0$, 1.5 и 2.0 соответствуют непрерывные, штриховые и точечные кривые соответственно.

Стационарные дисперсии $\mathcal{D}^{[s]}_{X_1} = \mathcal{C}_{X_1X_1}(0)$, $\mathcal{D}^{[s]}_{X_2} = \mathcal{C}_{X_2X_21}(0)$ и ковариации $\mathcal{K}^{[s]}_{X_1X_2} = \mathcal{C}_{X_1X_2}(0)$ при этом будут равны:

Характеристика	$\tau = 1.0$	$\tau = 1.5$	$\tau = 2.0$
$\mathcal{D}^{[s]}_{X_1}$	0.8695669982	0.8708367623	0.8715484372
$\mathcal{K}^{[s]}_{X_1X_2}$	0.0119059989	0.0096128103	0.0089303070
$\mathcal{D}^{[s]}_{X_2}$	1.6013575257	1.6094588646	1.6112101630

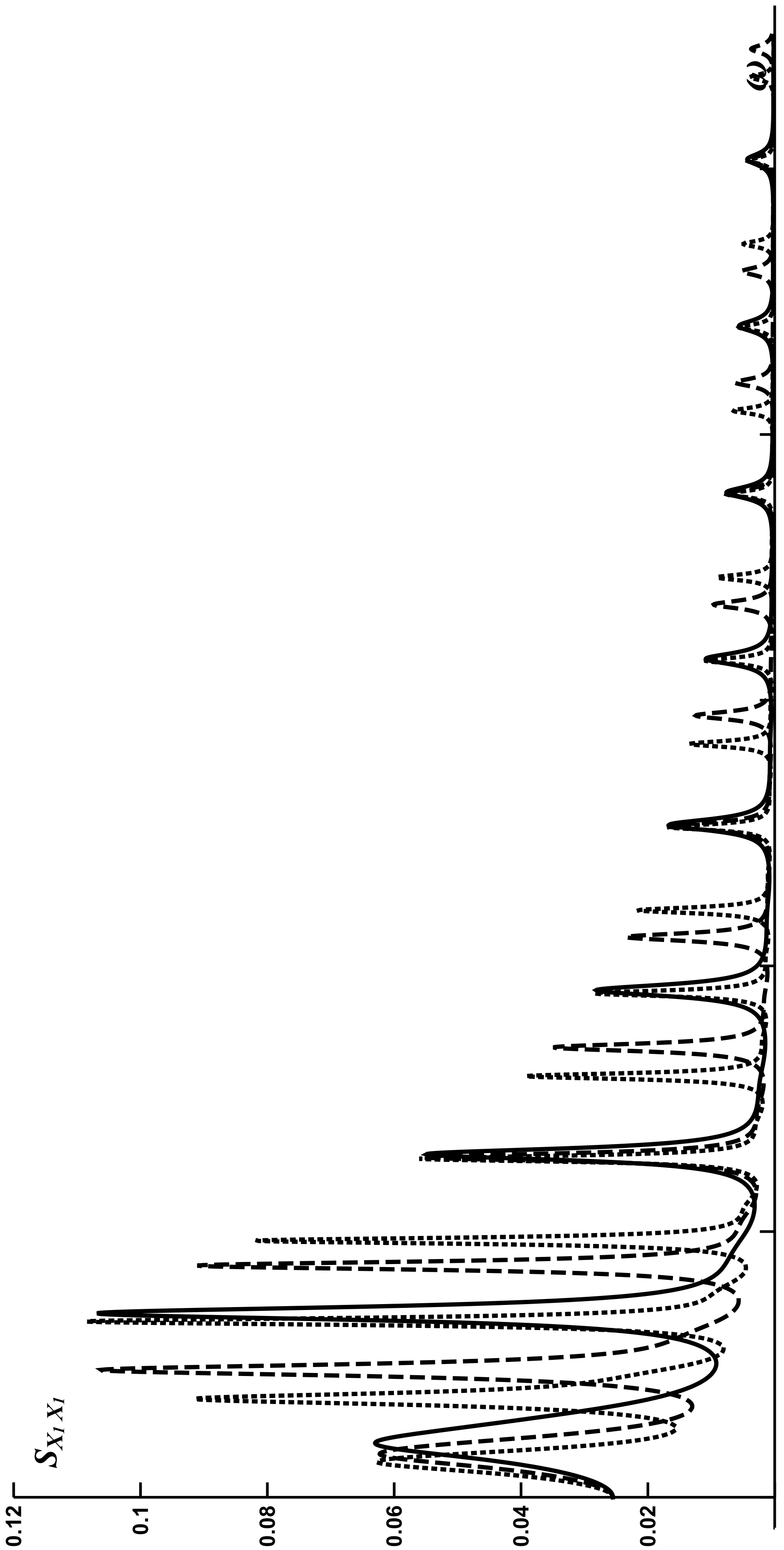


Рис. 1

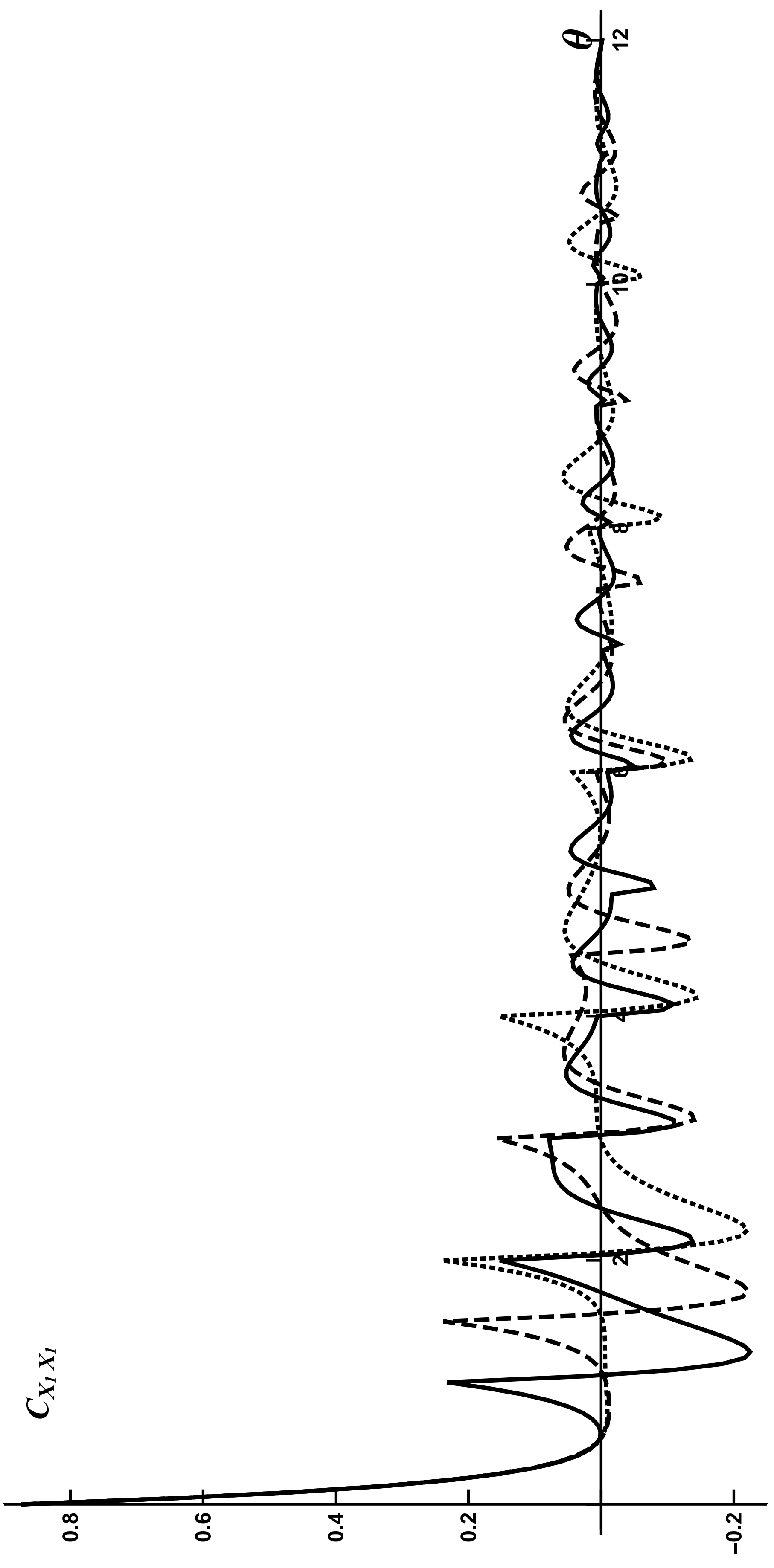


Рис. 2

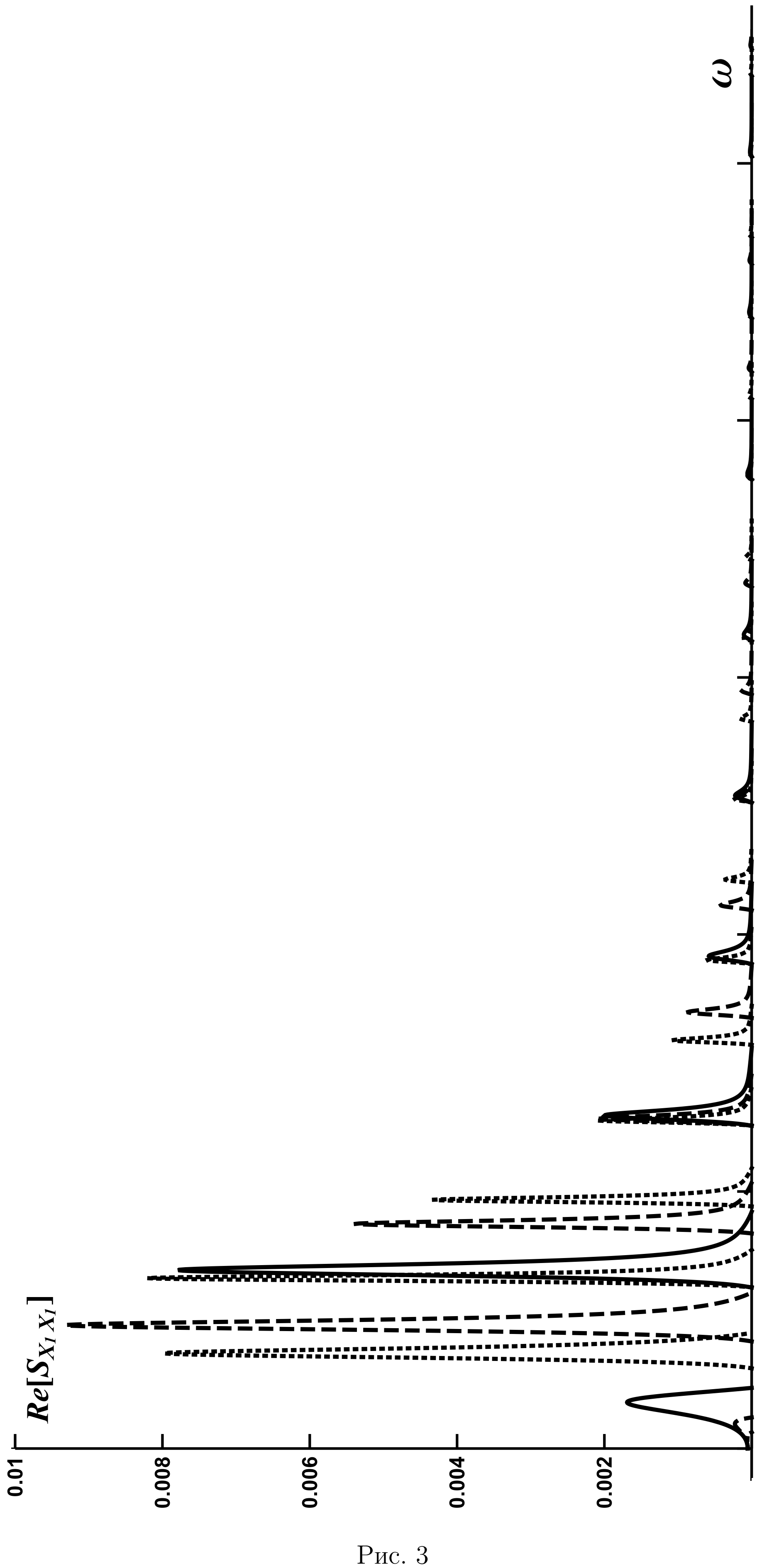


Рис. 3

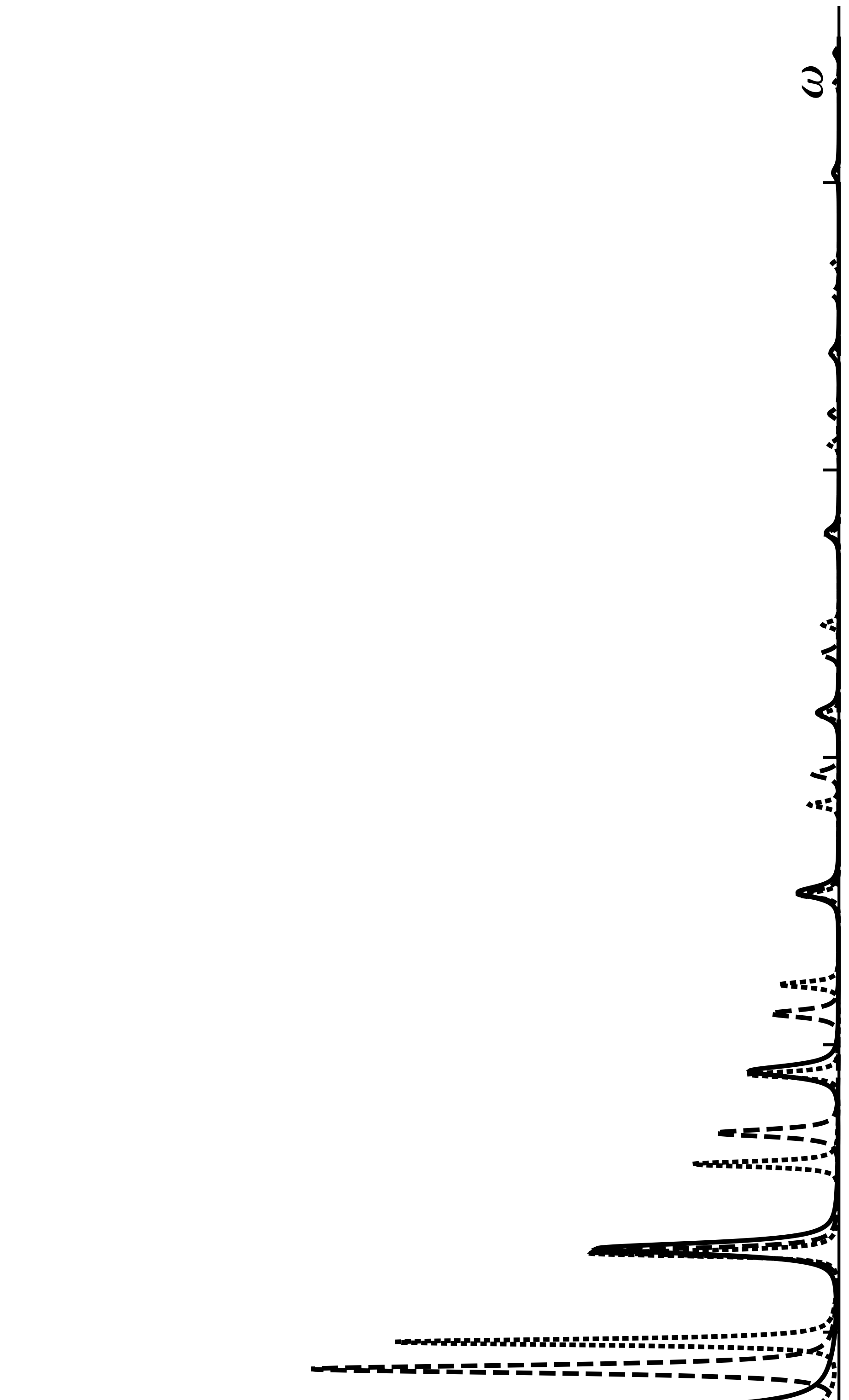


Рис. 4

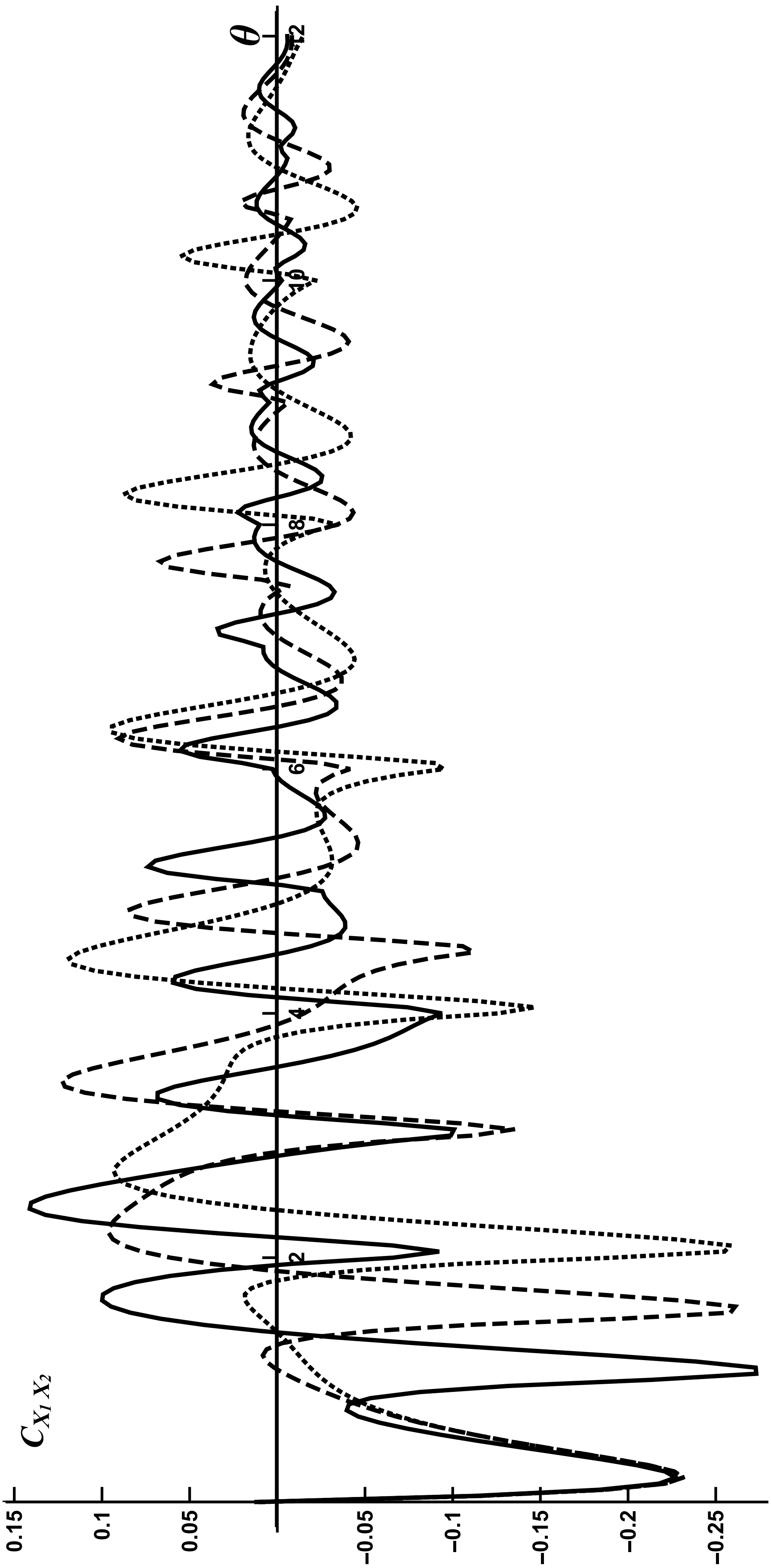


Рис. 5

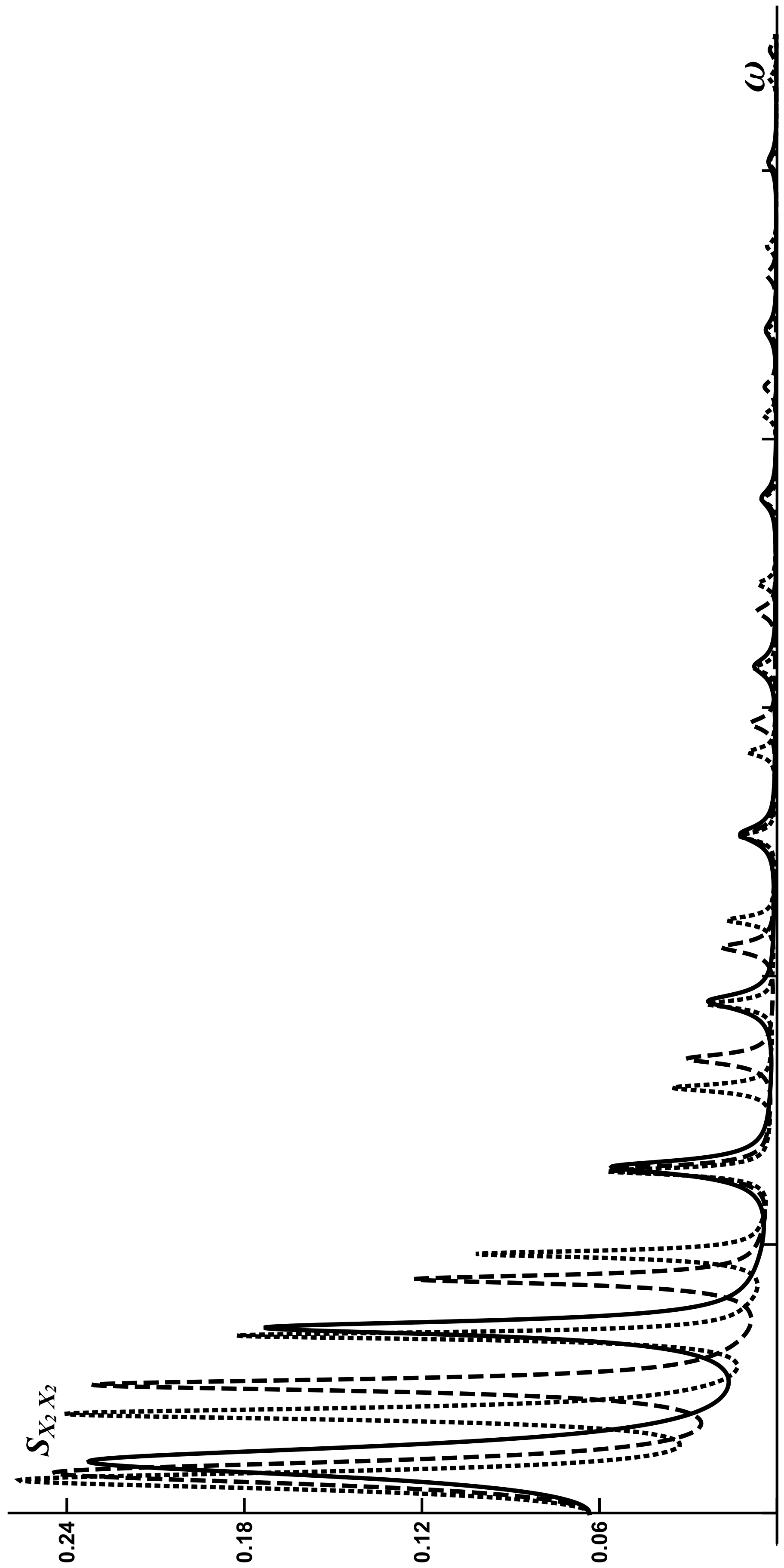


Рис. 6

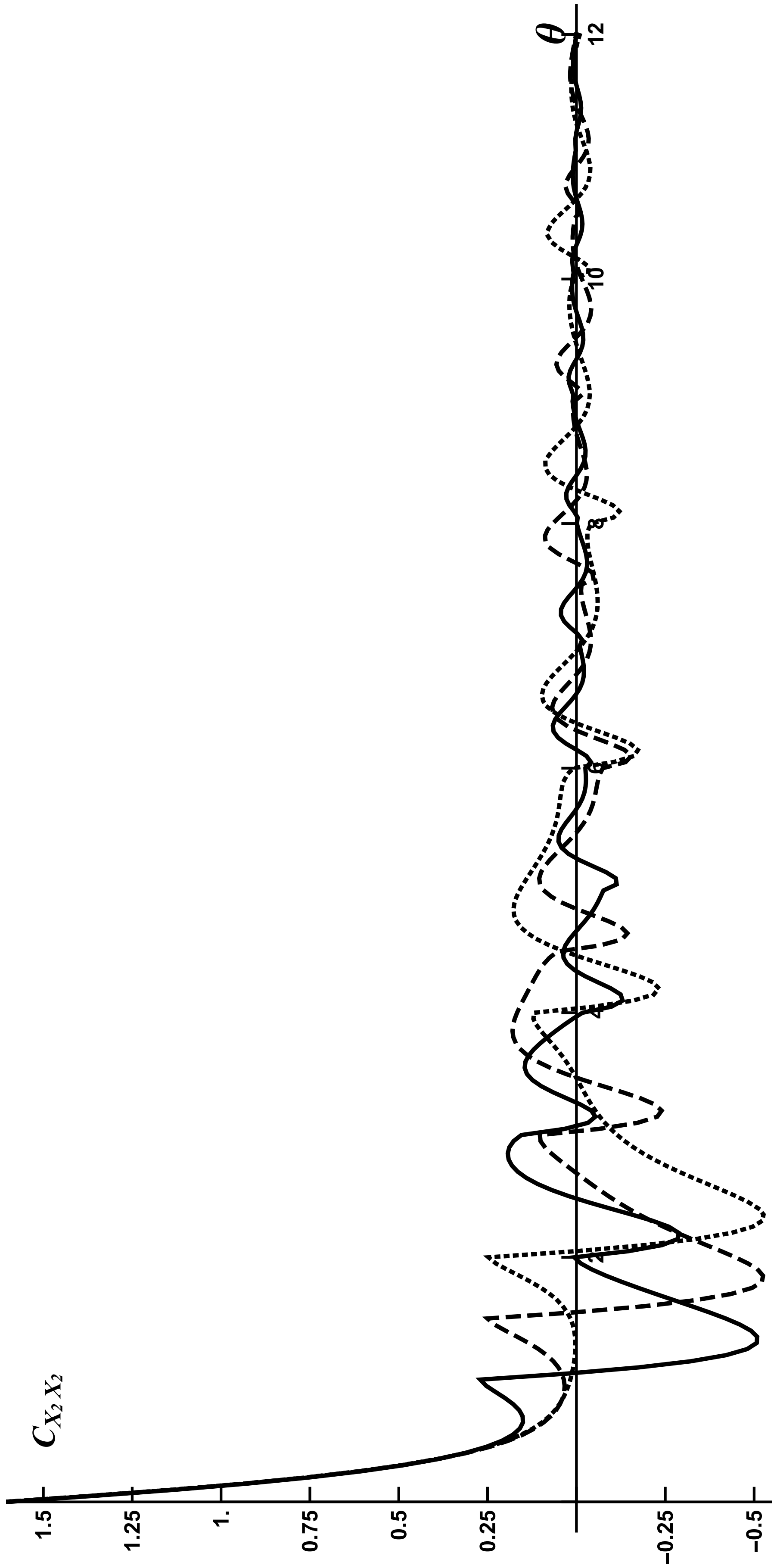


Рис. 7

Из таблицы несложно увидеть, что с ростом запаздывания стационарные дисперсии увеличиваются, а ковариация уменьшается.

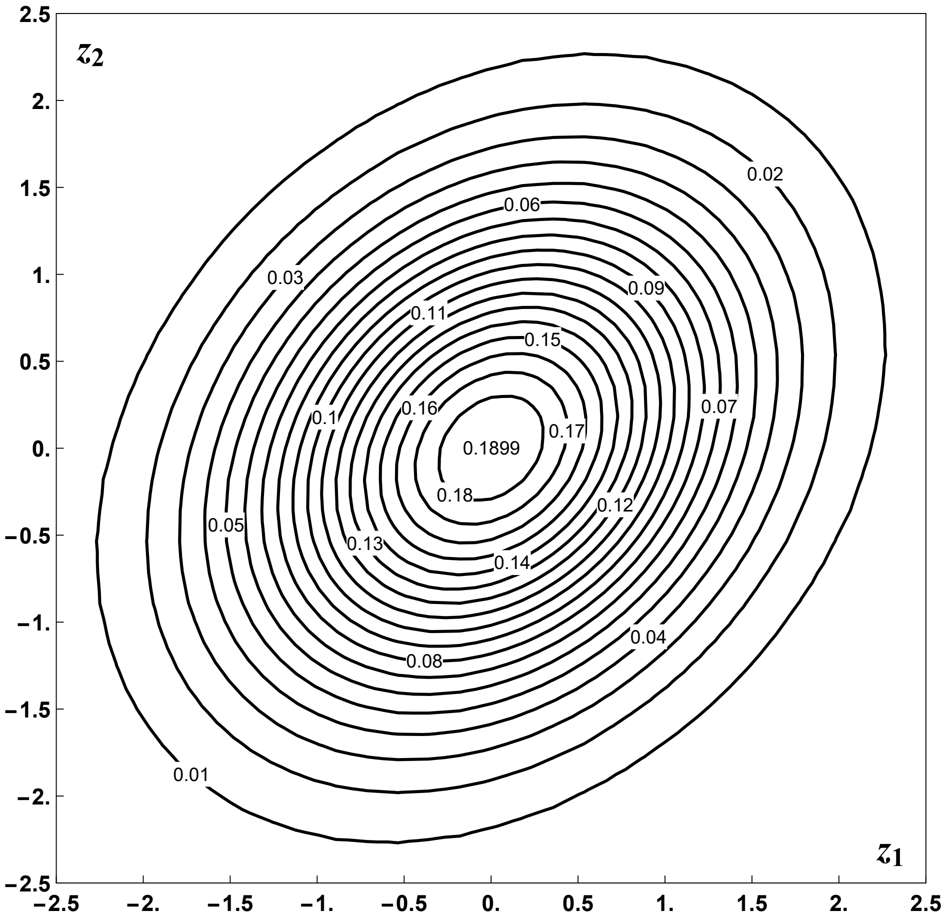


Рис. 8

Линии уровня совместной нормальной плотности сечений $Z_1 = X_1(t)$ и $Z_2 = X_1(t+\tau)$ случайного процесса X_1 в стационарном состоянии изображены на рис. 8.

5. Заключение

В работе схема С. Гийюзика, предназначенная для вычисления стационарных характеристик линейных стохастических систем первого порядка с одним запаздыванием, распространена на новый класс систем, описываемых линейными дифференциальными уравнениями нейтрального типа с кратными постоянными запаздываниями и возмущаемых векторными стационарными случайными процессами. Изложенная в статье процедура применена на получения стационарных характеристик системы второго порядка с двумя запаздываниями. Для проведения численно-аналитических расчетов применялся пакет MATHEMATICA.

Библиографический список

1. *Волков И.К., Зуев С.М., Цветкова Г.М.* Случайные процессы. – 3-е изд. – М.: Изд-во МГТУ им. Баумана, 2006. – 448 с.

2. *Полосков И.Е.* О расширении области применения схемы Гийюзика поиска стационарных распределений для линейных стохастических систем с запаздыванием // Проблемы механики и управления. Нелинейные динамические системы: межвуз. сб. науч. тр. / Перм. ун-т. – Пермь, 2013. – Вып. 45. – С. 92–102.

3. *Полосков И.Е.* О применении схемы Гийюзика для расчета матрицы спектральных плотностей вектора состояния линейной стохастической системы с многими запаздываниями // Вестник Пермского ун-та. Математика. Механика. Информатика. – 2015. – Вып. 1 (28). – С. 39–44.

4. *Полосков И.Е.* Построение спектральной плотности решения линейного стохастического дифференциального уравнения в частных производных с постоянными запаздываниями // Вестник Пермского ун-та. Математика. Механика. Информатика. – 2018. – Вып. 1 (40). – С. 36–45.

5. *Abell M.L, Braselton J.P.* Mathematica by example. – Amsterdam e.a.: Elsevier Inc., 2009. – XI, 564 p.

6. *Erneux T.* Applied delay differential equations. – New York: Springer, 2009. – XII, 204 p.

7. *Frank T.D., Beek P.J.* Stationary solutions of linear stochastic delay differential equations: Applications to biological systems // Physical Review E. – 2001. – Vol. 64, № 2 (021917). –12 p.

8. *Frank T.D.* Multivariate Markov processes for stochastic systems with delays: Application to the stochastic Gompertz model with delay // Physical Review E. – 2002. – Vol. 66, № 1 (011914). – 8 p.

9. *Frank T.D., Friedrich R.* Fokker–Planck perspective on stochastic delay systems: Exact solutions and data analysis of biological systems // Physical Review E. – 2003. – Vol. 68, № 2 (021912). – 10 p.

10. *Frank T.D.* Stationary distributions of stochastic processes described by a linear neutral delay differential equation // Journal of Physics A: Mathematical and General. – 2005. – Vol. 38, № 28. – P. L485–L490.

11. *Frank T.D.* Delay Fokker–Planck equations, Novikov's theorem, and Boltzmann distributions as small delay approximations // Physical Review E. – 2005. – Vol. 72, № 1 (011112). – 8 p.

12. *Fridman E.* Introduction to time-delay systems: Analysis and control. – Basel: Birkhäuser, 2014. – XVIII, 362 p.

13. *Giuggioli L., McKetterick T.J., Kenkre V.M., Chase M.* Fokker–Planck description for a linear delayed Langevin equation with additive Gaussian noise // Journal of Physics A: Mathematical and Theoretical. – 2016. – Vol. 49, № 32 (384002). – 29 p.

14. *Giuggioli L., Neu Z.* Fokker–Planck representations of nonMarkov Langevin equations: application to delayed systems // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. – 2019. – Vol. 377, № 2153. – Article ID 20180131. – 23 p.

15. *Guillouzic S., L'Heureux I., Longtin A.* Small delay approximation of stochastic differential delay equations // Physical Review. – 1999. – Vol. E59, № 4. – P. 3970–3982.

16. *Guillouzic S., L'Heureux I., Longtin A.* Rate processes in a stochastically driven delayed overdamped // Physical Review. – 2000. – Vol. E61, № 5. – P. 4906–4914.

17. *Guillouzic S.* Fokker–Planck approach to stochastic delay differential equations. – Thesis for the degree of Doctor of Philosophy. – Ottawa: University of Ottawa, 2000. – 100 p.

18. *Kolmanovskii V., Myshkis A.* Applied theory of functional differential equations. Mathematics and its Applications. – Dordrecht: Springer, 1992. – XVI, 234 p.

19. *Kolmanovskii V., Myshkis A.* Introduction to the theory and applications of functional differential equations. – Dordrecht: Springer, 1999. – XVI, 648 p.

20. *Küchler U., Mensch B.* Langevin's stochastic differential equation extended by a time-delayed term // Stochastics Rep. – 1992. – Vol. 40. – P. 23–42.

21. *Smith H.* An introduction to delay differential equations with applications to the life sciences. – New York: Springer, 2011. – XI, 172 p.

Stationary Characteristics for Solutions of Systems of Linear Neutral Differential Equations with Multiple Delays and Random Fluctuations

Igor E. Poloskov[1]

[1] Perm State National Research University;
polosk@psu.ru; +7(342) 2396560, Russia, Perm, Bukirev str., 15

Abstract. *The scheme of S. Guillouzic designed to obtain covariance functions and spectral densities of solutions* (*in the steady state*) *of a linear stochastic differential equation with one constant delay, extends to a new field of application, namely, linear systems of stochastic neutral differential equations with multiple delays. The obtained analytical relations for the spectral density matrices are used for numerical calculation* (*in the environment of the* Mathematica *package*) *of the corresponding matrices of the covariance functions.*

Key words: *stochastic analysis, stochastic ordinary differential equation, linear dynamical system, delay, stationary noise, spectral density, covariance function.*

УДК 519.2

Построение толерантных интервалов в случае двухпараметрического гамма-распределения

М. Н. Федорук, В. В. Чичагов

Пермский государственный национальный исследовательский университет, г. Пермь, Россия; chichagov@psu.ru; +7(342) 2396214; 614990, г. Пермь, ул. Букирева, 15

Аннотация. *Рассмотрена задача построения правосторонних толерантных интервалов в случае двухпараметрического гамма-распределения. Наряду с тремя известными методами решения этой задачи (фидуциальный, использование обращенного разложения Корниша-Фишера, нормальной аппроксимации) предложено еще два новых метода (с помощью интервальной оценки и асимптотики оценок максимального правдоподобия).*
Ключевые слова: *двухпараметрическое гамма-распределение, толерантный интервал, фидуциальный метод, разложение Корниша-Фишера, оценка максимального правдоподобия.*

1. Введение

Гамма-распределение является одной из современных моделей прикладной статистики, часто используется в качестве модели неотрицательных данных со скошенным вправо распределением. Поскольку семейство гамма распределений включает распределения, которым соответствует убывающая, постоянная или возрастающая интенсивность отказов, гамма модель является достаточно гибкой и позволяет описывать различные данные типа времени жизни в теории надежности и анализа живучести. Например, установлено, что гамма-распределение хорошо подходит для описания данных о времени жизни [1], [2] и данных, отражающих процесс деградации качества изделий [3], [4]. Также гамма распределение находит широкое применение в других областях, таких как метеорология и гидрология [5]–[6] для моделирования количества суточных осадков гидрологических данных, генетика, климатологии и иных науках об окружающей среде. В последнее время гамма-распределение считается подходящей моделью для отношения сигнал /шум в беспроводных каналах [7].

Поскольку существуют различные параметризации двухпараметрического гамма-распределения, определимся с этим понятием.

Определение 1. Случайная величина ξ имеет гамма-распределение $Gam(\alpha, \lambda)$ с параметром формы $\alpha > 0$ и параметром интенсивности $\lambda > 0$, если распределение задано функцией плотности распределения

$$f_\xi(x|\alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x},\ x > 0.$$

Ей соответствует функция распределения

$$F_\xi(x|\alpha, \lambda) = \int_0^x \frac{\lambda^\alpha t^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda t} dt, \quad x \geq 0.$$

Нетрудно видеть, что $F_\xi\left(x|\alpha, \lambda\right)$ – интегральная и $F_\xi^{-1}\left(x|\alpha, \lambda\right)$ – квантильная функции гамма-распределения допускают представление в виде

$$F_\xi\left(x|\alpha, \lambda\right) = \int_0^x \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \exp\left\{-\lambda x\right\} dx = \int_0^{\lambda x} \frac{u^{\alpha-1}}{\Gamma(\alpha)} \exp\left\{-u\right\} du,$$

$$F_\xi^{-1}\left(p|\alpha, \lambda\right) = \frac{1}{\lambda} F_0^{-1}\left(p|\alpha, 1\right) = \frac{\alpha}{\lambda} \frac{F_0^{-1}\left(p|\alpha, 1\right)}{\alpha},$$

где $F_0\left(x|\alpha, 1\right)$ – функция распределения $Gam(\alpha, 1), 0 < p < 1$.

Интегральная функция гамма-распределения обладает свойством монотонности по параметрам.

Лемма. *Функция распределения $F_\xi(x|\alpha, \lambda)$ случайной величины ξ, имеющей гамма-распределение $Gam(\alpha, \lambda)$, является убывающей функцией относительно $\alpha > 0$ и возрастающей функцией относительно $\lambda > 0$.*

Доказательство. Воспользуемся свойством воспроизводимости гамма-распределения: если ξ_1, ξ_2 – независимые случайные величины, имеющие соответственно распределения $Gam(\alpha_1, \lambda)$ и $Gam(\alpha_2, \lambda)$, то их сумма имеет также гамма-распределение $Gam(\alpha_1 + \alpha_2, \lambda)$. Для этой пары случайных величин появление события $\{\xi_1 + \xi_2 < x\}$ влечет появление события $\{\xi_1 < x\}$. Поэтому верно неравенство

$$F_{\xi_1+\xi_2}(x|\alpha_1 + \alpha_2, p) = \mathbf{P}\left(\xi_1 + \xi_2 \leq x\right\} \leq \mathbf{P}\left(\xi_1 \leq x\right\} = \mathbf{F}_{\xi_1}(x|\alpha_1, p),$$

означающее убывание функции $F_\xi(x|\alpha, \lambda)$ по параметру α. Отсюда следует справедливость утверждения леммы 1, поскольку возрастание функции распределении $F_\xi(x|\alpha, \lambda)$ относительно λ очевидно.

Замечание 1. Используя графический анализ отношения $F_0^{-1}(\beta|\alpha, 1)/\alpha$

при каждом из значений $\beta = 0,8,\ 0,9,\ 0,95$ можно убедиться в том, что оно определяет выпуклую кверху функцию относительно аргумента $\alpha > 0$, точка максимума которой меньше $0,8$.

Вопросы построения точечных и интервальных статистических оценок для параметров двухпараметрического гамма-распределения изучены достаточно полно [8]–[11]. Хорошо известно (см. например, [10]), что $\tilde{\alpha}, \tilde{\lambda}$ – оценки максимального правдоподобия (**ОМП**) неизвестных параметров α, λ по случайной выборке $X_1, \ldots, X_n$ из генеральной совокупности ξ определяются соотношениями

$$\tilde{\lambda} = \frac{\tilde{\alpha}}{\bar{X}}, \quad \ln \frac{\tilde{X}}{\bar{X}} + \ln \tilde{\alpha} - \psi\left(\tilde{\alpha}\right) = 0, \tag{1}$$

где

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \tilde{X} = \left(\prod_{i=1}^{n} X_i\right)^{1/n}.$$

При этом **ОМП** $\tilde{\alpha}$ является функцией от статистики $T = \ln W$, где $W = \tilde{X}/\bar{X}$, а **ОМП** $\tilde{\lambda}$ – функция от статистики $\left(T, \bar{X}\right)$. Также установлено, что статистика $\left(T, \bar{X}\right)$ является достаточной и полной, при этом T и $\bar{X}$ – взаимно независимые случайные величины.

В данной работе продолжены исследования, посвященные решению задачи построения толерантных интервалов для гамма-распределения. При этом, не нарушая общности полученных результатов, мы ограничиваемся рассмотрением только случая правосторонних толерантных интервалов.

Определение 2. (β, γ)-правосторонним толерантным интервалом для случайной величины ξ, имеющей гамма-распределение $Gam(\alpha, \lambda)$, назовем интервал $\left[0, L\left(\bar{X}, \tilde{X}\right)\right)$, правой границей которого является некоторая статистика $L = L\left(\bar{X}, \tilde{X}\right)$, удовлетворяющий для любых $\alpha > 0, \lambda > 0$ соотношению

$$\mathbf{P}\left\{\mathbf{P}\left(\xi < L\left(\bar{X}, \tilde{X}\right) | \bar{X}, \tilde{X}\right) \geq \beta\right\} = \mathbf{P}\left\{F_\xi\left(L\left(\bar{X}, \tilde{X}\right) | \alpha, \lambda\right) \geq \beta\right\} = \gamma.$$

или, что то же самое, соотношению

$$\mathbf{P}\left\{x_\beta[Gam(\alpha, \lambda)] \leq L\left(\bar{X}, \tilde{X}\right)\right\} = \gamma,$$

где $x_\beta[Gam(\alpha, \lambda)]$ – квантиль уровня β гамма-распределения $Gam(\alpha, \lambda)$.

Примеры применения такого типа толерантных интервалов для осуществления экологического мониторинга окружающей среды с целью выявления неблагоприятного воздействия потенциального источника загрязнения и контроля качества продукта представлены, в частности, в работах [12] и [13].

В разделах 2–4 данной работы рассмотрено 5 различных подходов к построению толерантных интервалов: фидуциальный метод [14], использование разложение Корниша-Фишера квантильной функции [11], нормальная аппроксимация гамма-распределения [15]–[17], использование интервальных оценок для отдельных сомножителей квантиля гамма распределения, асимптотическое распределение **ОМП** для квантиля гамма распределения.

2. Обзор ранее предложенных методов построения толерантного интервала

Фидуциальный подход. Основная идея фидуциального вывода состоит в изменении роли параметров и наблюдаемых данных. Если имеется процесс генерации данных

$$\mathbf{Z} = H(\boldsymbol{\theta}, \boldsymbol{\Upsilon}), \tag{2}$$

где $\boldsymbol{\theta}$ – вектор параметров, а $\boldsymbol{\Upsilon}$ – случайный вектор с полностью известным законом распределения, не зависящим от параметров, роль данных и параметров $\boldsymbol{\theta}$ можно поменять после наблюдения данных $\boldsymbol{\Upsilon}$, решив структурное уравнение (2), т.е. получив выражение

$$\theta = G\left(\mathbf{Z}, \boldsymbol{\Upsilon}\right). \tag{3}$$

Теперь $\boldsymbol{\theta}$ можно трактовать как случайную переменную. Распределение, определяемое (3), называют фидуциальным.

В статье [14] представлен следующий алгоритм построения толерантного интервала фидуциальным методом.
Шаг 1. По выборке $X_1, \ldots, X_n$ вычисляются **ОМП** $\tilde{\alpha}, \tilde{\lambda}$ с помощью (1).
Шаг 2. Определяются значения статистик

$$\tilde{v} = \frac{2\left[\ln n - \psi\left(n\tilde{\alpha}\right) + \psi\left(\tilde{\alpha}\right)\right]^2}{\psi^{(1)}\left(\tilde{\alpha}\right)/n - \psi^{(1)}\left(n\tilde{\alpha}\right)},\ \tilde{c} = -\frac{-2n\tilde{\alpha}\left(\ln n - \psi\left(n\tilde{\alpha}\right) + \psi\left(\tilde{\alpha}\right)\right)}{\tilde{v}}.$$

где $\psi\left(\alpha\right) = \frac{d\ln\Gamma(\alpha)}{d\alpha}$ – дигамма функция, а $\psi^{(i)}(\alpha)$ – i-я производная от $\psi(\alpha),\ i = 1, 2, 3.$
Шаг 3. Сначала моделируются бутстреп-выборки $\chi_{0:1}, \ldots, \chi_{0:B}$ и $\chi_{1:1}, \ldots, \chi_{1:B}$ из генеральных совокупностей $\tilde{c}\chi^2\left(\tilde{v}\right)$ и $\chi^2\left(n\tilde{\alpha}\right)$ соответственно, а затем по формулам

$$\tilde{\alpha}_b = \frac{\chi_{0:b}}{2n\ln\left(\bar{X}/\tilde{X}\right)}, \quad \tilde{\lambda}_b = \frac{\chi_{1:b}}{n\bar{X}}$$

вычисляются $\left\{\tilde{\alpha}_b, \tilde{\lambda}_b,\ i = 1, \ldots, B\right\}$ – бутстреп-оценки параметров α, λ.

Шаг 4. Для каждой пары бутстреп-оценок $\tilde{\alpha}_b, \tilde{\lambda}_b$ находим $\tilde{\tau}_b(\beta)$ – квантиль уровня β распределения $Gam(\tilde{\lambda}_b, \tilde{\alpha}_b)$, $b = 1, \ldots, B$. Тогда верхняя (β, γ)-толерантная граница $L_{u:1}$ гамма-распределения $Gam(\alpha; \lambda)$ по выборке $X_1, \ldots, X_n$ определяется как квантиль уровня γ выборки $\{\tilde{\tau}_b(\beta), b = 1, \ldots, B\}$.

Метод, использующий обращенное разложение Корниша–Фишера [11]. Пусть $F_T(t|\alpha)$ – функция распределения статистики $T = \ln\left(\tilde{X}/\bar{X}\right)$. Как уже отмечалось выше, статистика $\left(T, \bar{X}\right)$ является достаточной и полной, при этом T и $\bar{X}$ взаимно независимы.

Обращенное разложение Корниша–Фишера $Q(\alpha, \beta)$ для квантильной функции, включающее в себя кумулянты до 5-го порядка включительно, определяется выражением

$$\begin{aligned} Q(\alpha, \beta) = z_\beta &+ \frac{\rho_3(\alpha)}{6}\left(z_\beta^2 - 1\right) + \frac{\rho_4(\alpha)}{24}\left(z_\beta^3 - 3z_\beta\right) - \frac{\rho_3^2(\alpha)}{36}\left(2z_\beta^3 - 5z_\beta\right) + \\ &+ \frac{\rho_5(\alpha)}{120}\left(z_\beta^4 - 6z_\beta^2 + 3\right) - \frac{\rho_3(\alpha)\rho_4(\alpha)}{24}\left(z_\beta^4 - 5z_\beta^2 + 2\right) + \\ &+ \frac{\rho_3^3(\alpha)}{324}\left(12z_\beta^4 - 53z_\beta^2 + 17\right), \end{aligned}$$

где $\rho_i(\alpha) = \kappa_i(\alpha)/\kappa_2^{i/2}(\alpha), i = 3, 4, 5$, а z_β – квантиль уровня β стандартного нормального распределения, $\kappa_i(\alpha)$ – i-й кумулянт распределения случайной величины T. В случае гамма-распределения $Gam(\alpha, \lambda)$

$$\kappa_1(\alpha) = \ln n + \psi(\alpha) - \psi\left(n\alpha\right), \quad \kappa_i(\alpha) = \frac{1}{n^{i-1}}\psi^{(i-1)}(\alpha) - \psi^{(i-1)}\left(n\alpha\right),$$

где $\psi(\alpha) = \frac{d\ln\Gamma(\alpha)}{d\alpha}$ – дигамма функция, а $\psi^{(i)}(\alpha)$ – i-я производная от $\psi(\alpha)$, $i = 2, 3, 4, 5$.

Путем анализа графика функции $h(\alpha) = \kappa_1(\alpha) + \kappa_2^{1/2}(\alpha)Q\left(\alpha, \beta\right)$ в [11] установлено, что если $n \geq 5$ и $z_\beta \leq 4$, то $h(\alpha)$ – строго возрастающая функция при заданном β.

Обозначим через $g(T, U)$ – решение уравнения

$$T = \kappa_1(\alpha) + \kappa_2^{1/2}(\alpha)Q(\alpha, U) \tag{4}$$

относительно α, где U – реализация равномерного распределения $U\left(0; 1\right)$.

В статье [11] предложен следующий алгоритм построения толерантного интервала с использованием обращенного разложения Корниша-Фишера.

Шаг 1. По выборке $X_1, \ldots, X_n$ вычисляются статистики T и $\bar{X}$.

Шаг 2. Моделируем $U_1, \ldots, U_B$ – случайную выборку из генеральной совокупности $U\,(0;1)$ и формируем бутстреп-выборку $\{g_b, b = 1, \ldots, B\}$ из решений уравнений

$$T = \kappa_1(\alpha) + \kappa_2^{1/2}(\alpha) Q\,(\alpha, U_b)\,, \quad b = 1, \ldots, B,$$

относительно α.

Шаг 3. Моделируем независимые случайные величины $V_1, \ldots, V_B$, имеющие хи-квадрат распределения $\chi^2\,(2ng_1)\,, \ldots, \chi^2\,(2ng_B)$ соответственно, а затем вычисляем последовательность значений

$$w_b = V_b / \left(2n\bar{X}\right), \quad b = 1, \ldots, B.$$

Шаг 4. Формируем бутстреп-выборку $\tau_b = F_{\chi^2}^{-1}\,(\beta|2g_b)\,/\,(2w_b)$, $b = 1, \ldots, B$.

Тогда верхнюю границу $L_{u:2}$ правостороннего (β, γ)-толерантного интервала гамма-распределения $Gam(\alpha, \lambda)$ по выборке $X_1, \ldots, X_n$ определим как квантиль уровня γ выборки $\{\tau_b(\beta), b = 1, \ldots, B\}$.

Использование нормальной аппроксимации гамма распределения. Wilson и Hilfertys в [18] предложили нормальную аппроксимацию для кубичного корня случайной величины, имеющей распределение хи-квадрат. Пусть Y_α – случайная величина, имеющая распределение $Gam\,(\alpha, 1)$ или, что то же самое, распределенная как $\chi^2_{2\alpha}/2$. Для получения аппроксимации нормальным распределением был использован принцип соответствия моментов нормального распределения и распределения переменной хи-квадрат в некоторой степени δ. Среднее и дисперсия Y_α^δ соответственно равны

$$\mu_\delta = \frac{\Gamma(\alpha + \delta)}{\Gamma(\alpha)}, \quad \sigma_\delta^2 = \frac{\Gamma(\alpha + 2\delta)}{\Gamma(\alpha)} - \mu_\delta^2.$$

Wilson и Hilfertys [18] выбрали $\delta = 1/3$, в этом случае $\sqrt[3]{Y}$ аппроксимируется нормальным распределениям $N\left(\mu_{1/3}, \sigma^2_{1/3}\right)$. Hawkins и Wixley [19] считают, что при малых значениях α аппроксимацию можно улучшить, используя $\delta = 1/4$. В [17] приведен следующий алгоритм построения толерантного интервала, использующий нормальную аппроксимацию гамма-распределения.

Шаг 1. Применяя Wilson–Hilferty аппроксимацию, полагаем $Y_i = \sqrt[3]{X_i}, i = 1, \ldots, n$, и определяем

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Шаг 2. Вычисляем верхнюю толерантную границу для нормального распределения

$$R = \bar{Y} + x_\gamma \left[T_{n-1}\left(z_\beta\sqrt{n}\right)\right]\frac{s_Y}{\sqrt{n}}, \tag{5}$$

где z_β – квантиль уровня β стандартного нормального распределения, а $x_\gamma\left[T_{n-1}\left(z_\beta\sqrt{n}\right)\right]$ – квантиль уровня γ нецентрального распределения Стьюдента с числом степеней свободы $n-1$ и параметром нецентральности $z_\beta\sqrt{n}$. Тогда $L_{u:3} = R^3$ – приближенная верхняя граница правостороннего (β,γ)-толерантного интервала для гамма-распределения.

3. Использование доверительных интервалов для параметров гамма-распределения

Монотонность функции гамма-распределения $F_\xi\left(x|\alpha,\lambda\right)$ относительно параметров α,λ согласно лемме из введения позволяет для решения задачи построения верхнего предела толерантного интервала использовать решение задачи построения доверительного интервала для функции от двух параметров, в качестве которой в данном случае выступает значение квантильной функции $x_\beta(\alpha,\lambda) = F_\xi^{-1}\left(\beta|\alpha,\lambda\right)$. В соответствии с этим подходом следует определить $\Theta_\gamma\left(\bar{X},T\right)$ – доверительное множество для параметров α и λ с уровнем доверия $\gamma=\gamma_1\gamma_2\,(0<\gamma_1<1,\ 0<\gamma_2<1)$, для которого с вероятностью γ выполняется неравенство

$$x_\beta(\alpha,\lambda) \le \sup_{(\alpha,\lambda)\in\Theta_\gamma\left(\bar{X},T\right)} F_\xi^{-1}\left(\beta|\alpha,\lambda\right). \tag{6}$$

В (6), как и ранее, $\bar{X}$ – среднее по выборке $X_1,\ldots,X_n$, а $T=\ln\left(\tilde{X}/\bar{X}\right)$.

Поскольку случайная величина $n\lambda\bar{X}$ имеет гамма-распределение $Gam\,(n\alpha,1)$, статистики T и $\bar{X}$ взаимно независимы, доверительное множество $\Theta_\gamma\left(\bar{X},T\right)$, учитывая представление квантильной функции

$$F_\xi^{-1}\left(\beta|\alpha,\lambda\right) = \frac{1}{\lambda}F_0^{-1}\left(\beta|\alpha,1\right),$$

естественно определить в следующем виде:

$$\Theta_\gamma\left(\bar{X},T\right) = \left\{(\alpha,\lambda) : \alpha > \Upsilon(T), \lambda n\bar{X} > x_{1-\gamma_2}\left[Gam\,(\alpha n,1)\right]\right\},$$

где $\Upsilon(T)$ – граница левостороннего доверительного интервала $(\Upsilon(T),\infty)$ с уровнем доверия γ_1 для параметра α, а $x_{1-\gamma_2}\left[Gam\,(\alpha n,1)\right]$ – квантиль уровня $1-\gamma_2$ распределения $Gam\,(n\alpha,1)$.

В этом случае с учетом замечания 1 из введения верхний предел правостороннего (β,γ)-толерантного интервала будет определяться следующим выражением

$$L_{u:4} = n\bar{X} \sup_{\alpha > \Upsilon(T)} \frac{x_\beta\left[Gam\left(\alpha,1\right)\right]}{x_{1-\gamma_2}\left[Gam\left(n\alpha,1\right)\right]}. \tag{7}$$

Справедливость формулы (7) следует из соотношений

$$\begin{aligned}
\gamma_2 &= \mathbf{P}\left\{\lambda n\bar{X} > x_{1-\gamma_2}\left[Gam(\alpha n,1)\right]\right\} = \mathbf{P}\left\{\frac{1}{\lambda} < \frac{n\bar{X}}{x_{1-\gamma_2}\left[Gam(\alpha n,1)\right]}\right\} = \\
&= \mathbf{P}\left\{x_\beta\left(\alpha,\lambda\right) < n\bar{X}\frac{x_\beta\left[Gam(\alpha,1)\right]}{x_{1-\gamma_2}\left[Gam(\alpha n,1)\right]}\right\}, \\
\gamma &= \mathbf{P}\left\{\frac{1}{\lambda} < \frac{n\bar{X}}{x_{1-\gamma_2}\left[Gam(\alpha n,1)\right]}, \alpha > \Upsilon(T)\right\} = \\
&= \mathbf{P}\left\{x_\beta(\alpha,\lambda) < n\bar{X}\frac{x_\beta\left[Gam(\alpha,1)\right]}{x_{1-\gamma_2}\left[Gam(\alpha n,1)\right]}, \alpha > \Upsilon(T)\right\} \le \\
&\le \mathbf{P}\left\{x_\beta(\alpha,\lambda) < n\bar{X} \sup_{\alpha > \Upsilon(T)} \frac{x_\beta\left[Gam(\alpha,1)\right]}{x_{1-\gamma_2}\left[Gam(\alpha n,1)\right]}\right\}.
\end{aligned}$$

Интервальную оценку для параметра α в виде неравенства $\alpha > \Upsilon(T)$ можно получить, используя результаты из [20] или [11].

Точный доверительный интервал для параметра формы (по материалам работы [20]). Пусть $U_{ij}, i = 1,\ldots,B,\ j = 1,\ldots,n$ – независимые равномерные на интервале $(0;1)$ случайные величины, а $Q(u,k) = F_\zeta^{-1}(u|k)$ – квантильная функция гамма-распределения $Gam(k;1)$.

Положим

$$X_{ij}(k) = Q\left(U_{ij},k\right), \quad \bar{X}_i(k) = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$$

и определим

$$\hat{T}_i(k) = \prod_{j=1}^{n} \frac{\left[X_{ij}(k)\right]^{1/n}}{\bar{X}_i(k)} = W\left(X_{i1}(k),\ldots,X_{in}(k)\right). \tag{8}$$

Тогда правосторонний и левосторонний доверительные интервалы для параметра α с надежностью γ имеют вид $\left(0,\hat{\alpha}_{[\ell]}\right)$ и $\left(\hat{\alpha}_{[B-\ell+1]},\infty\right)$ соответственно, где $\ell = \gamma\left(B+1\right)$ (выбор B целесообразно осуществлять таким образом, чтобы ℓ оказалось натуральным числом), $\hat{\alpha}_{[\ell]}$ и $\hat{\alpha}_{[B-\ell+1]}$ – решения уравнений $\hat{T}_{(B-\ell+1)}(k) = w_{obs}$ и $\hat{T}_{(\ell)}(k) = w_{obs}$ относительно аргумента

k. Здесь $w_{obs} = W(X_1, \ldots, X_n)$, $\hat{T}_{(i)}(k)$ – i-я порядковая статистика выборки $\hat{T}_1(k), \ldots, \hat{T}_B(k)$.

Примечание 1. В этом случае в формуле (7) следует положить $\Upsilon(T) = \hat{T}_{(B-\ell+1)}\left(\hat{k}\right)$, где $\hat{T}_{(B-\ell+1)}(\cdot)$ – $(B-\ell+1)$-я порядковая статистика выборки $\left\{\hat{T}_b(k),\ b = 1, \ldots, B\right\}$, вычисляемая при каждом значении k по формуле (8), $\hat{k}$ – решение уравнения $\hat{T}_{(B-\ell+1)}(k) = \tilde{X}/\bar{X}$ относительно k.

Приближенный доверительный интервал для параметра формы (по материалам работы [11].) В статье [11] предложен следующий алгоритм построения доверительного интервала для параметра α с использованием обращенного разложения Корниша–Фишера.

Шаг 1. По выборке $X_1, \ldots, X_n$ вычисляется статистика $T = \ln\left(\tilde{X}/\bar{X}\right)$.

Шаг 2. Приближенный доверительный интервал $(\alpha_{AL}, \alpha_{AU})$ с надежностью p получим, решая уравнения

$$T = \kappa_1(\alpha_{AL}) + \sqrt{\kappa_2(\alpha_{AL})}Q\left(\alpha_{AL}, \frac{1+p}{2}\right),$$
$$T = \kappa_1(\alpha_{AU}) + \sqrt{\kappa_2(\alpha_{AU})}Q\left(\alpha_{AL}, \frac{1-p}{2}\right).$$

Примечание 2. Чтобы получить левосторонний доверительный интервал (α_{AL}, ∞) с надежностью γ_1, надо решить уравнение

$$T = \kappa_1(\alpha_{AL}) + \sqrt{\kappa_2(\alpha_{AL})}Q(\alpha_{AL}, \gamma).$$

В этом случае в формуле (7) следует положить $\Upsilon(T) = \alpha_{AL}$.

4. Асимптотический толерантный интервал, основывающийся на ОМП квантиля гамма-распределения

Нам потребуется следующее утверждение.

Теорема. *Пусть $\tilde{\theta}_1, \tilde{\theta}_2$ – статистически независимые оценки максимального правдоподобия параметров $\theta_1 \in \Theta_1 \subset \mathbf{R}$, $\theta_2 \in \Theta_2 \subset \mathbf{R}$ по случайной выборке объема n, элементы которой имеют некоторое двухпараметрическое распределение; $I_1(\theta_1, \theta_2), I_2(\theta_1, \theta_2)$ – информация Фишера о параметрах θ_1, θ_2, содержащаяся в одном наблюдении. Тогда, если последовательности оценок $\left\{\sqrt{n}\left(\tilde{\theta}_1 - \theta_1\right)\right\}, \left\{\sqrt{n}\left(\tilde{\theta}_2 - \theta_2\right)\right\}$ асимптотически нормальны при $n \to \infty$ с параметрами $\left(0; I_1^{-1}(\theta_1, \theta_2)\right)$ и $\left(0; I_2^{-1}(\theta_1, \theta_2)\right)$ соответственно, функция $G = G(\theta_1, \theta_2) = \Upsilon_1(\theta_1) \cdot \Upsilon_2(\theta_2)$*

дифференцируема в окрестности точки (θ_1, θ_2), $\Upsilon_1'(\theta_1) \neq 0$ *или* $\Upsilon_2'(\theta_2) \neq 0$, *то при* $n \to \infty$

$$\left\{\sqrt{n}\left(\tilde{G} - G\right)\right\} \Rightarrow \mathbf{N}\left(0; G^2\sigma_G^2\right), \quad \left\{\sqrt{n}\ln\frac{\tilde{G}}{G}\right\} \Rightarrow \mathbf{N}\left(0; \sigma_G^2\right),$$

где

$$\tilde{G} = \Upsilon_1\left(\tilde{\theta}_1\right)\cdot\Upsilon_2\left(\tilde{\theta}_2\right), \quad \sigma_G^2 = \sigma_G^2(\theta_1, \theta_2) = \sum_{i=1}^{2}\frac{1}{I_i(\theta_1, \theta_2)}\left[\frac{\partial}{\partial\theta_i}\ln\Upsilon_i(\theta_i)\right]^2.$$

Доказательство. Справедливость утверждения при сделанных предположениях следует из соотношений

$$\tilde{G} - G = G\left[\frac{\partial\ln\Upsilon_1(\theta_1)}{\partial\theta_1}\left(\tilde{\theta}_1 - \theta_1\right) + \frac{\partial\ln\Upsilon_2(\theta_1)}{\partial\theta_2}\left(\tilde{\theta}_2 - \theta_2\right)\right] + o_P\left(\frac{1}{\sqrt{n}}\right),$$
$$\ln\tilde{G} - \ln G = \frac{\partial\ln\Upsilon_1(\theta_1)}{\partial\theta_1}\left(\tilde{\theta}_1 - \theta_1\right) + \frac{\partial\ln\Upsilon_2(\theta_1)}{\partial\theta_2}\left(\tilde{\theta}_2 - \theta_2\right) + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Следствие 1. *Пусть выполнены условия теоремы, а* ***ОМП*** $\tilde{\theta}_1, \tilde{\theta}_2$ *состоятельны. Если* z_γ *– квантиль уровня* γ *стандартного нормального распределения, то при* $n \to \infty$ *с асимптотической вероятностью* γ *выполняются неравенства*

$$G < \tilde{G}\left(1 + \frac{z_\gamma\tilde{\sigma}_G}{\sqrt{n}}\right), \quad G < \tilde{G}\exp\left\{\frac{z_\gamma\tilde{\sigma}_G}{\sqrt{n}}\right\}, \quad G < \frac{\tilde{G}}{1 - z_\gamma\tilde{\sigma}_G/\sqrt{n}}, \tag{9}$$

где

$$\tilde{\sigma}_G = \sqrt{\sum_{i=1}^{2} I_i^{-1}(\theta_1, \theta_2)\left[\frac{\partial}{\partial\theta_i}\ln\Upsilon_i(\theta_i)\right]^2}\Bigg|_{\theta_i = \tilde{\theta}_i}. \tag{10}$$

Справедливость первых двух неравенств из (9) следует немедленно из утверждения теоремы и состоятельности при $n \to \infty$ оценок максимального правдоподобия $\tilde{\theta}_1, \tilde{\theta}_2$. Проверим справедливость третьего из неравенств при $n \to \infty$:

$$\mathbf{P}\left[G < \frac{\tilde{G}}{1 - z_\gamma\tilde{\sigma}_G/\sqrt{n}}\right] = \mathbf{P}\left[G\left(1 - z_\gamma\tilde{\sigma}_G/\sqrt{n}\right) \leq \tilde{G}\right] =$$
$$= \mathbf{P}\left[\sqrt{n}\left(G - \tilde{G}\right) \leq Gz_\gamma\tilde{\sigma}_G\right] \to \Phi\left(\frac{Gz_\gamma\sigma_G}{\sqrt{G^2\sigma_G^2}}\right) = \gamma.$$

Следствие 2. *Пусть $\tilde{\alpha}, \tilde{\lambda}$ –* ***ОМП*** *неизвестных параметров α, λ по случайной выборке $X_1, \ldots, X_n$ из генеральной совокупности ξ, имеющей гамма-распределение $Gam(\alpha, \lambda)$, $\alpha > 0, \lambda > 0$. Тогда в качестве верхнего предела правостороннего (β, γ)-толерантного интервала при достаточно большом объеме выборки n можно использовать одно из следующих выражений:*

$$L_{u:5} = \frac{\bar{X}\, u_\beta}{\tilde{\alpha}} \left(1 + \frac{z_\gamma \tilde{\sigma}}{\sqrt{n}}\right),\ L_{u:6} = \frac{\bar{X}\, u_\beta}{\tilde{\alpha}} \exp\left\{\frac{z_\gamma \tilde{\sigma}}{\sqrt{n}}\right\},\ L_{u:7} = \frac{\bar{X}\, u_\beta}{\tilde{\alpha}} \frac{1}{1 - z_\gamma \tilde{\sigma}/\sqrt{n}}, \tag{11}$$

где

$$\sigma^2(\alpha) = \frac{1}{\alpha} + \frac{1}{\psi'(\alpha) - 1/\alpha} \left[\frac{\beta\psi(\alpha)}{u_\beta f_0(u_\beta|\alpha, 1)} - \int\limits_0^{u_\beta} \frac{\ln t\, f_0(t|\alpha, 1)}{u_\beta f_0(u_\beta|\alpha, 1)} dt - \frac{1}{\alpha}\right]^2,$$

$$\tilde{\sigma} = \sqrt{\sigma^2(\tilde{\alpha})}, \quad u_\beta = x_\beta\left[Gam(\tilde{\alpha}, 1)\right],$$

$f_0(t|\alpha, 1)$ – плотность гамма-распределения $Gam(\alpha; 1)$, $\alpha > 0$.

Доказательство. Воспользуемся следствием 1 при $\theta_1 = \alpha/\lambda, \theta_2 = \alpha$, полагая

$$\Upsilon_1(\theta_1) = \theta_1, \quad \Upsilon_2(\theta_2) = \frac{x_\beta}{\alpha},$$

где $x_\beta = x_\beta\left[Gam(\alpha, 1)\right]$ – квантиль уровня β гамма-распределения $Gam(\alpha, 1)$. При таком выборе параметров θ_1, θ_2 и функций $\Upsilon_1(\theta_1), \Upsilon_2(\theta_2)$ **ОМП** $\tilde{\theta}_1 = \bar{X}, \tilde{\theta}_2 = \tilde{\alpha}$, а функция G совпадает с квантильной функцией, т.е.

$$G(\theta_1, \theta_2) \equiv F_\xi^{-1}(\beta|\alpha, \lambda) = \frac{1}{\lambda} F_0^{-1}(\beta|\alpha, 1) = \frac{\alpha}{\lambda} \frac{F_0^{-1}(\beta|\alpha, 1)}{\alpha} = \theta_1 \frac{x_\beta}{\alpha}.$$

В случае гамма-распределения случайной величины $\eta \sim Gam(\alpha, \alpha/\theta_1)$ плотность распределения и логарифмическая функция правдоподобия одного наблюдения имеют вид

$$f_\eta(x|\alpha, \theta_1) = \left(\frac{\alpha}{\theta_1}\right)^\alpha \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-\alpha x/\theta_1},\ x > 0,$$

$$\ln f_\eta(\eta|\alpha, \theta_1) = \alpha(\ln\alpha - \ln\theta_1) + (\alpha - 1)\ln\eta - \ln\Gamma(\alpha) - \frac{\alpha}{\theta_1}\eta$$

соответственно. При этом информация Фишера, содержащаяся в одном наблюдении η о параметрах θ_1 и α определяется выражениями

$$I_1(\theta_1, \alpha) = -\mathbf{E}\left[\frac{\alpha}{\theta_1^2} - \frac{2\alpha}{\theta_1^3}\eta\right] = -\frac{\alpha}{\theta_1^2} + \frac{2\alpha}{\theta_1^3}\theta_1 = \frac{\alpha}{\theta_1^2},$$

$$I_2(\theta_1, \alpha) = -\mathbf{E}\left[\frac{1}{\alpha} - \frac{\partial^2 \ln\Gamma(\alpha)}{\partial\alpha^2}\right] = \psi'(\alpha) - \frac{1}{\alpha}.$$

Вычислим производную от квантильной функции $x_\beta = x_\beta\left[Gam\left(\alpha, 1\right)\right]$ гамма-распределения $Gam(\alpha, 1)$, как производную обратной функции:

$$\frac{\partial x_\beta}{\partial \alpha} = \frac{\partial F_0^{-1}\left(\beta|\alpha, 1\right)}{\partial \alpha} = -\frac{\partial F_0\left(x_\beta|\alpha, 1\right)}{\partial \alpha} : \frac{\partial F_0\left(x_\beta|\alpha, 1\right)}{\partial x_\beta} =$$

$$= -\frac{\partial F_0\left(x_\beta|\alpha, 1\right)}{\partial \alpha}\frac{1}{f_0\left(x_\beta|\alpha, 1\right)} = -\int\limits_0^{x_\beta} \frac{1}{f_0(x_\beta|\alpha, 1)}\frac{\partial f_0(t|\alpha, 1)}{\partial \alpha}dt =$$

$$= -\frac{1}{f_0(x_\beta|\alpha, 1)}\int\limits_0^{x_\beta} \frac{\partial \ln f_0(t|\alpha, 1)}{\partial \alpha} f_0(t|\alpha, 1)dt =$$

$$= -\frac{1}{f_0(x_\beta|\alpha, 1)}\int\limits_0^{x_\beta} \frac{\partial}{\partial \alpha}\left[(\alpha-1)\ln t - t - \ln\Gamma(\alpha)\right] f_0(t|\alpha, 1)dt =$$

$$= \frac{\beta\psi(\alpha)}{f_0(x_\beta|\alpha, 1)} - \frac{1}{f_0(x_\beta|\alpha, 1)}\int\limits_0^{x_\beta} \ln t\, f_0(t|\alpha, 1)dt.$$

Отсюда следует, что справедливы формулы

$$\frac{\partial}{\partial \theta_1}\ln \Upsilon_1\left(\theta_1\right) = \frac{1}{\theta_1},$$

$$\frac{\partial}{\partial \alpha}\ln \Upsilon_2(\alpha) = \frac{1}{x_\beta f_0(x_\beta|\alpha, 1)}\left[\beta\psi(\alpha) - \int\limits_0^{x_\beta} \ln t\, f_0(t|\alpha, 1)dt\right] - \frac{1}{\alpha},$$

$$\sigma_G^2\left(\theta_1, \alpha\right) = \frac{\theta_1^2}{\alpha}\frac{1}{\theta_1^2} + \left(\psi'(\alpha) - \frac{1}{\alpha}\right)^{-1} \times$$

$$\times\left[\frac{\beta\psi(\alpha)}{x_\beta f_0(x_\beta|\alpha, 1)} - \frac{1}{x_\beta f_0(x_\beta|\alpha, 1)}\int\limits_0^{x_\beta} \ln t\, f_0(t|\alpha, 1)dt - \frac{1}{\alpha}\right]^2,$$

из которых с учетом результата следствия 1 следует справедливость следствия 2.

В ближайшем будущем планируется провести компьютерное сравнение точности рассмотренных в работе методов построения толерантных интервалов и сопоставить результаты этих исследований с результатами, представленными в работах [11], [14], [20].

Библиографический список

1. *Fernandez A. J.* Fernandez A. J., Optimal reliability demonstration test plans for k-out-of-n systems of gamma distributed components// IEEE Transactions on Reliability. – 2011. – Vol. 60, №4. – P. 833–844.

2. *Fan T.-H., Yu C.-H.* Statistical inference on constant stress accelerated life tests under generalized gamma lifetime distributions// Quality and Reliability Engineering International. – 2013. – Vol. 29, №5. – P. 631–638.

3. *Ye Z.-S., Xie M., Tang L.-C., Shen Y.* Degradation-based burn-in planning under competing risks// Technometrics. – 2012. – Vol. 54, №2. – P. 159–168.

4. *Ye Z.-S., Xie M., Tang L.-C., and Chen N.* Semiparametric estimation of gamma processes for deteriorating products// Technometrics. – 2014. – Vol. 56, №4. – P. 504–513.

5. *Stephenson D., Kumar K. R., Doblas-Reyes F., Royer J., Chauvin F., Pezzulli S.* Extreme daily rain fall events and their impact on ensemble forecasts of the indian monsoon// Monthly Weather Review. – 1999. Vol. 127, №9. – P. 1954–1966.

6. *Aksoy H.* Use of gamma distribution in hydrological analysis// Turkish Journal of Engineering and Environmental Sciences. – 2000. – Vol. 24, №6. – P. 419–428.

7. *Atapattu S., Tellambura C., Jiang H.* A mixture gamma distribution to model the SNR of wireless channels// IEEE Transactions on Wireless Communications. – 2011. – Vol. 10, №12. – P. 4193–4203.

8. ГОСТ 11.011-83. Правила определения оценок и доверительных границ для параметров гамма-распределения. – М.: Издательство стандартов, 1985.

9. *Johnson N.L., Kotz S., Balakrishnan N.* Continuous univariate distributions. Volume 1. – New York: John Wiley & Sons, 1994.

10. *Zaigraev A., Podraza-Karakulska A.* On estimation of the shape parameter of the gamma distribution// Statistics & Probability Letters. – 2008. – Vol. 78. – P. 286–295.

11. *Wang B.X., Wu F.* Inference on the gamma distribution// Technometrics (Online) Journal. – 2017. – homepage: http://www.tandfonline.com/loi/utch20: ACCEPTED MANUSCRIPT. – 25 P.

12. *Bhaumik D. K. and Gibbons R. D.* One-sided approximate prediction intervals for at least p of m observations from a gamma population at each of r locations// Technometrics. – 2006. – Vol. 48, №1. – P. 112–119.

13. *Fernandez A. J.* Tolerance limits for k-out-of-n systems with exponentially distributed component lifetimes// Reliability, IEEE Transactions on. – 2010. – Vol. 59, №2. – P. 331–337.

14. *Piao C., Zhi-Sheng Y.* Tolerance Limits for Gamma Distribution Based on Generalized Fiducial Method// Department of Industrial & Systems Engineering, National University of Singapor, 2015. – P. 122–125.

15. *Aryal S., Bhaumik D., Mathew T., and Gibbons R.* Approximate tolerance limits and prediction limits for the gamma distribution// J Appl Stat Sci. – 2007. – Vol. 16. – P. 103–111.

16. *Krishnamoorthy K., Mathew T., and Mukherjee S.* Normal-based methods for a gamma distribution// Technometrics. – 2008. – Vol. 50, №1. – P. 69–78.

17. *Krishnamoorthy K.. Mathew T.* Statistical Tolerance Regions: Theory, Applications and Computation. – New York: Wiley, 2009.

18. *Wilson E.B., Hilferty M.M.* The distribution of chi-squares// Proceedings of the National Academy of Sciences. – 1931. – Vol. 17. – P. 684–688.

19. *Hawkins D.M., Wixley R.A.* A note on the transformation of chi-squared variables to normality// The American Statistican. – 1986. – Vol. 40. – P. 296–298.

20. *Iliopoulos G.* Exact confidence intervals for the shape parameter of the gamma distribution// Journal of Statistical Computation and Simulation. – 2015. – http://dx.doi.org/10.1080/00949655.2015.1080705, – P. 1–8 Published online: 01 Sep 2015.

Construction of tolerance intervals in the case of a two-parameter gamma distribution

M. H. Fedoruk, V. V. Chichagov

Perm State University, Perm, Russia;
chichagov@psu.ru; +7(342) 2396214; 15, Bukireva str., Perm, Russia, 614990

Abstract. *The problem of constructing right-hand tolerant intervals in the case of a two-parameter gamma distribution is considered. Along with three well-known methods for solving this problem (fiducial, using the inverted Cornish-Fisher expansion, normal approximation), two more new methods are proposed (using the interval estimate and the asymptotic behavior of the maximum likelihood estimates).*

Key words: *asymptotic expected deficiency, exponential family, maximum likelihood estimation, uniformly minimum variance unbiased estimator.*

СОДЕРЖАНИЕ

Научное издание

Статистические методы оценивания и проверки гипотез

Межвузовский сборник научных трудов
Вып. 29

Издается в авторской редакции
Компьютерная вёрстка: *П. В. Шеховцова*

Подписано в печать 17.12.2019. Формат 60×84/16.
Усл. печ. л. 6,86. Тираж 45 экз. Заказ 200

Издательский центр
Пермского государственного
национального исследовательского университета.
614990 г. Пермь, ул. Букирева, 15

Типография ПГНИУ.
614990 г. Пермь, ул. Букирева, 15