
УДК 517.518:519.213

В.С. Берегун, канд. техн. наук
Национальный технический университет Украины
«Киевский политехнический ин-т»
(Украина, 03056, Киев, ул. Политехническая, 16, корп. 12,
тел. (044) 2049072, e-mail: viktorberegun@i.ua)

Исследование точности аппроксимации симметричных плотностей вероятности ортогональными представлениями по полиномам Эрмита

Исследованы ошибки аппроксимации симметричных плотностей вероятности типовых распределений с помощью полиномов Эрмита. Проведено сравнение результатов для частных и обобщенных сумм рядов по полиномам Эрмита при числе составляющих ряда до двадцати. Выполнена проверка ортогональных представлений на неотрицательность.

Досліджено помилки апроксимації симетричних щільностей імовірності типових розподілів за допомогою поліномів Ерміта. Проведено порівняння результатів для часткових і узагальнених сум рядів по поліномах Ерміта при кількості складових ряду до двадцяти. Виконано перевірку ортогональних подань на невід'ємність.

К л ю ч е в ы е с л о в а: плотность вероятности, аппроксимация, полиномы Эрмита, ортогональное представление, обобщенное суммирование.

Для нахождения неизвестных плотностей вероятности при решении задач радиотехники [1], метрологии [2], функциональных преобразований случайных процессов [3], исследования моделей шумовых сигналов [4] и других часто применяют [3, 5—7] ортогональные представления следующего вида [8]:

$$p(x) = \rho(x) \sum_{k=0}^{\infty} c_k \varphi_k(x),$$

где c_k — коэффициенты разложения; $\varphi_k(x)$ — система ортогональных функций с весовой функцией $\rho(x)$. Коэффициенты разложения находят по формуле

$$c_k = \frac{1}{\|\varphi_k\|^2} \int_{-\infty}^{\infty} p(x) \varphi_k(x) dx, \quad (1)$$

где $\|\varphi_k\|$ — норма функций $\varphi_k(x)$.

© В.С. Берегун, 2016

Для плотностей вероятности, близких к нормальной, используются ряды по полиномам Эрмита. Пусть ξ — стандартная случайная величина с неизвестной плотностью вероятности $p(x)$, $x \in (-\infty; \infty)$, нулевым математическим ожиданием и единичным средним квадратическим отклонением. Тогда для нахождения плотности вероятности применяется ряд вида

$$p(x) = \rho_H(x) \sum_{k=0}^{\infty} c_{Hk} H_k(x), \quad (2)$$

где $H_k(x)$ — полиномы Эрмита [5, 9],

$$H_k(x) = (-1)^k e^{\frac{x^2}{2}} \frac{d^k}{dx^k} \left(e^{-\frac{x^2}{2}} \right),$$

которые определяются на интервале $x \in (-\infty; \infty)$ с весовой функцией $\rho_H(x) = e^{-\frac{x^2}{2}}$. При этом $\|H_k\|^2 = \sqrt{2\pi} k!$

Сумма (2) называется рядом Грама— Шарлье [3, 5, 6, 10]. При практическом применении формулы (2) вместо бесконечного ряда используется частная сумма с конечным числом составляющих до n -го порядка:

$$S_n(x) = \rho_H(x) \sum_{k=0}^n c_{Hk} H_k(x). \quad (3)$$

Приведем несколько первых полиномов Эрмита $H_k(x)$:

$$H_0(x) = 1; \quad H_1(x) = x; \quad H_2(x) = x^2 - 1;$$

$$H_3(x) = x^3 - 3x; \quad H_4(x) = x^4 - 6x^2 + 3;$$

$$H_5(x) = x^5 - 10x^3 + 15x; \quad H_6(x) = x^6 - 15x^4 + 45x^2 - 15.$$

Тогда согласно (1) коэффициенты разложения представляют собой функции начальных моментов α_j случайной величины ξ :

$$c_{H0} = \frac{1}{\sqrt{2\pi}}; \quad c_{H1} = c_{H2} = 0; \quad c_{H3} = \frac{1}{6\sqrt{2\pi}} \alpha_3; \quad c_{H4} = \frac{1}{24\sqrt{2\pi}} [\alpha_4 - 3];$$

$$c_{H5} = \frac{1}{120\sqrt{2\pi}} [\alpha_5 - 10\alpha_3]; \quad c_{H6} = \frac{1}{720\sqrt{2\pi}} [\alpha_6 - 15\alpha_4 + 30].$$

Полученные результаты справедливы как для нахождения неизвестных плотностей вероятности, так и для аппроксимации известных. В первом случае в коэффициентах разложения используются оценки моментов,

во втором — их теоретические значения. Следует заметить, что при $c_{Hk} = 0$ для всех $k \geq 1$ выражение (3) определяет плотность вероятности стандартной нормальной случайной величины с нулевым математическим ожиданием и единичным средним квадратическим отклонением.

В работе [11] для улучшения сходимости и точности аппроксимации плотностей вероятности предложено кроме частных сумм (4) использовать обобщенные суммы следующих видов:

суммы Фейера, являющиеся средним арифметическим n первых частных сумм,

$$F_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} S_k(x);$$

суммы Валле—Пуссена, представляющие собой среднее арифметическое частных сумм от $S_n(x)$ до $S_{2n-1}(x)$,

$$V_n(x) = \frac{1}{n} \sum_{k=n}^{2n-1} S_k(x).$$

Постановка задачи. В работе [5] приведено условие сходимости ряда (2) к плотности вероятности — существование интеграла $\int_{-\infty}^{\infty} e^{\frac{x^2}{4}} p(x) dx$.

Следует заметить, что данное условие не выполняется для типовых распределений [12], определенных на бесконечном интервале действительных чисел $(-\infty; \infty)$ (в том числе и для рассмотренных ниже), за исключением нормального распределения. Однако более важным является не сходимость ряда при использовании бесконечного числа составляющих, что практически невозможно, а обеспечение приемлемой точности при использовании небольшого числа составляющих ряда [5, 9]. На практике число составляющих в формуле (3) обычно не более четырех [13—15] или шести [3, 5, 6]. Важной задачей является также обеспечение неотрицательности ортогонального представления [14, 16]. Появление отрицательных значений противоречит свойствам плотности вероятности.

Будем рассматривать симметричные плотности вероятности $p(x)$, наиболее распространенные на практике. Исследуем точность аппроксимации плотности вероятности при увеличении числа составляющих как частных, так и обобщенных сумм, и выполним проверку этих ортогональных представлений на неотрицательность.

Ортогональные представления плотностей вероятности. Симметричные плотности вероятности, определенные на бесконечном интервале

$(-\infty; \infty)$, являются четными функциями и имеют нулевое математическое ожидание [12]. Для таких распределений справедливо следующее:

$$S_0(x) = S_1(x) = S_2(x) = S_3(x) = F_1(x) = F_2(x) = \\ = F_3(x) = F_4(x) = V_1(x) = V_2(x).$$

Все эти суммы совпадают со стандартной нормальной плотностью вероятности. Совпадают также начальные $\tilde{\alpha}_s$ и центральные $\tilde{\mu}_s$ моменты распределений. Кроме того, частные суммы нечетных порядков равны частным суммам предыдущих порядков: $S_5(x) = S_4(x)$, $S_7(x) = S_6(x)$, $S_9(x) = S_8(x)$ и так далее. Исследование проведем для частных сумм до двадцатого порядка $S_{20}(x)$.

В формулах (2), (3) использованы полиномы Эрмита для стандартных плотностей вероятности, поэтому произвольная случайная величина $\tilde{\xi}$ с плотностью вероятности $\tilde{p}(x)$, начальными моментами $\tilde{\alpha}_s$ ($\tilde{\alpha}_1 = 0$) и средним квадратическим отклонением $\tilde{\sigma} = \sqrt{\tilde{\alpha}_2}$ должна быть сведена к стандартной случайной величине ξ . В этом случае преобразование имеет вид $\xi = \tilde{\xi} / \tilde{\sigma}$, стандартная плотность вероятности — $p(x) = \tilde{\sigma} \tilde{p}(\tilde{\sigma} x)$, а моменты — $\alpha_s = \tilde{\alpha}_s / (\tilde{\sigma})^s$.

Приведем значения теоретических плотностей вероятности и их основных числовых характеристик [12].

Распределение Стьюдента. Плотность вероятности

$$\tilde{p}(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v} \Gamma(v/2)} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}},$$

где $\Gamma(x)$ — гамма-функция; $v > 0$ — параметр формы, число степеней свободы.

Среднее квадратическое отклонение $\tilde{\sigma} = \sqrt{\frac{v}{v-2}}$;
моменты

$$\tilde{\mu}_s = \begin{cases} 0, s - \text{нечетное}, v > s, \\ \frac{1 \cdot 3 \cdot 5 \cdot \dots \cdot (s-1) v^{s/2}}{(v-1)(v-4) \dots (v-s)}, s - \text{четное}, v > s. \end{cases}$$

Распределение логистическое. Плотность вероятности

$$\tilde{p}(x) = \frac{\exp(x/\lambda)}{\lambda [1 + \exp(x/\lambda)]^2},$$

где $\lambda > 0$ — параметр масштаба.

Среднее квадратическое отклонение $\tilde{\sigma} = \frac{\lambda\pi}{\sqrt{3}}$;

моменты

$$\tilde{\mu}_s = \begin{cases} 0, s - \text{нечетное,} \\ (\lambda\pi)^s (2^s - 2) |B_s|, s - \text{четное,} \end{cases}$$

где B_s — числа Бернулли [17].

Распределение Чампернауна. Плотность вероятности

$$\tilde{p}(x) = \frac{\alpha}{\pi \operatorname{ch}(\alpha x)},$$

где $\alpha > 0$ — параметр масштаба.

Среднее квадратическое отклонение $\tilde{\sigma} = \frac{\pi}{2\alpha}$;

моменты

$$\tilde{\mu}_s = \begin{cases} 0, s - \text{нечетное,} \\ \left(\frac{\pi}{2\alpha}\right)^s |E_s|, s - \text{четное,} \end{cases}$$

где E_s — числа Эйлера [18].

Распределение Лапласа. Плотность вероятности

$$\tilde{p}(x) = \frac{\lambda}{2} \exp(-\lambda |x|),$$

где $\lambda > 0$ — параметр масштаба.

Среднее квадратическое отклонение $\tilde{\sigma} = \sqrt{2/\lambda}$;

моменты

$$\tilde{\mu}_s = \begin{cases} 0, s - \text{нечетное,} \\ \frac{s!}{\lambda^s}, s - \text{четное.} \end{cases}$$

Примем следующие значения параметров распределений: Стьюдента — $\nu = 21$, логистического — $\lambda = 1$, Чампернауна — $\alpha = 1$, Лапласа — $\lambda = 1$. При $\nu = 21$ у распределения Стьюдента существуют все моменты до двадцатого порядка, для остальных распределений значения параметров масштаба влияют только на масштаб плотности вероятности и не изменяют стандартной плотности вероятности.

На рис. 1 приведены графики указанных выше стандартных плотностей вероятности. Для определения точности аппроксимации плотнос-

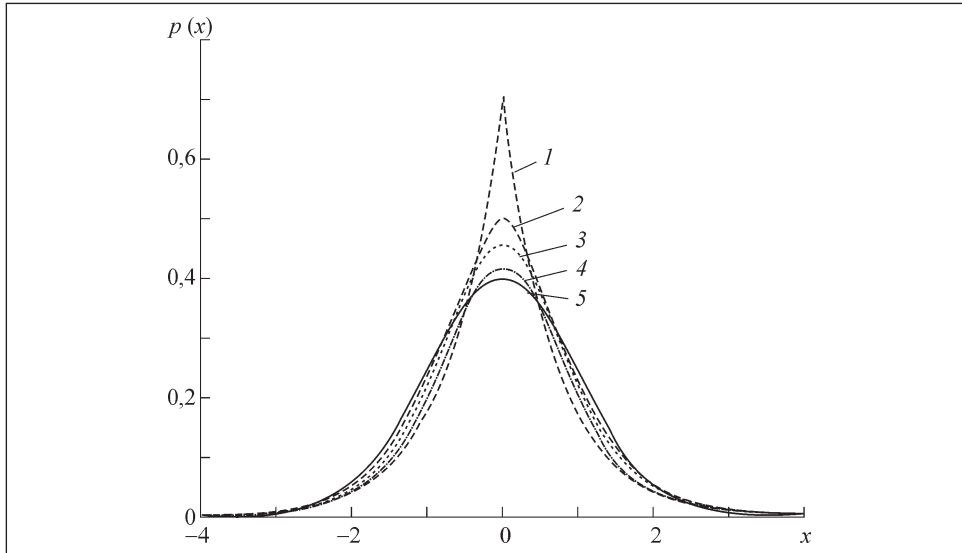


Рис. 1. Графики плотностей вероятности: 1 — Лапласа; 2 — Чампернауна; 3 — логистическое; 4 — Стьюдента; 5 — нормальное

тей вероятности ортогональными представлениями во всех случаях найдем интегральную ошибку:

$$r_i = \frac{1}{2} \int_{-\infty}^{\infty} |p(x) - p_i(x)| dx,$$

где $p(x)$ — стандартная теоретическая плотность вероятности; $p_i(x)$ — соответствующая аппроксимация. При аппроксимации плотности вероятности другой плотностью максимальное значение интегральной ошибки равно единице. В случае аппроксимации частными или обобщенными суммами значение интегральной ошибки может превышать единицу.

В табл. 1 приведены значения интегральной ошибки r_i для частных сумм $S_n(x)$, в табл. 2 — для сумм Фейера $F_n(x)$, в табл. 3 — для сумм Валле—Пуссена $V_n(x)$. Курсивным шрифтом в табл. 1—3 указаны значения ошибок при наличии отрицательных значений в ортогональном представлении. Значения ошибки, превышающие единицу, смещены к левому краю. Полужирным шрифтом указаны минимальные значения ошибок среди всех сумм для каждого распределения.

Среди частных сумм наилучшую аппроксимацию плотностей вероятности обеспечивает сумма $S_4(x)$ для всех рассмотренных распределений, зависящая только от четвертого момента. Следовательно, использование

Таблица 1

$S_n(x)$	Значение t_i для распределений			
	Стьюдента	Логистическое	Чампернауна	Лапласа
S_0	0,0174	0,0543	0,0885	0,1413
S_4	0,0034	0,0181	0,0377	0,0992
S_6	0,0066	0,0570	0,1411	0,2839
S_8	0,0077	0,1090	0,3578	0,8597
S_{10}	0,0163	0,3297	1,4169	4,1699
S_{12}	0,0451	1,1332	6,4576	23,328
S_{14}	0,1836	4,6152	34,829	154,62
S_{16}	1,1589	21,506	215,36	1176
S_{18}	13,3801	113,12	1504,3	10110
S_{20}	503,6837	662,82	11714	96923

Таблица 2

$F_n(x)$	Значение t_i для распределений			
	Стьюдента	Логистическое	Чампернауна	Лапласа
F_5	0,0133	0,0405	0,0658	0,1106
F_6	0,0106	0,0313	0,0510	0,0914
F_7	0,0096	0,0328	0,0605	0,1150
F_8	0,0090	0,0347	0,0693	0,1352
F_9	0,0076	0,0238	0,0365	0,0651
F_{10}	0,0065	0,0172	0,0301	0,0957
F_{11}	0,0063	0,0304	0,1102	0,3284
F_{12}	0,0062	0,0542	0,2183	0,6463
F_{13}	0,0053	0,0459	0,3146	1,2441
F_{14}	0,0057	0,1211	0,7497	2,8148
F_{15}	0,0091	0,2020	1,671	7,8737
F_{16}	0,0196	0,4764	3,7359	17,022
F_{17}	0,0511	0,8462	9,3753	54,168
F_{18}	0,1125	1,9886	20,789	116,38
F_{19}	0,6026	4,1746	60,584	428,03
F_{20}	1,2411	9,6047	132,63	911,48

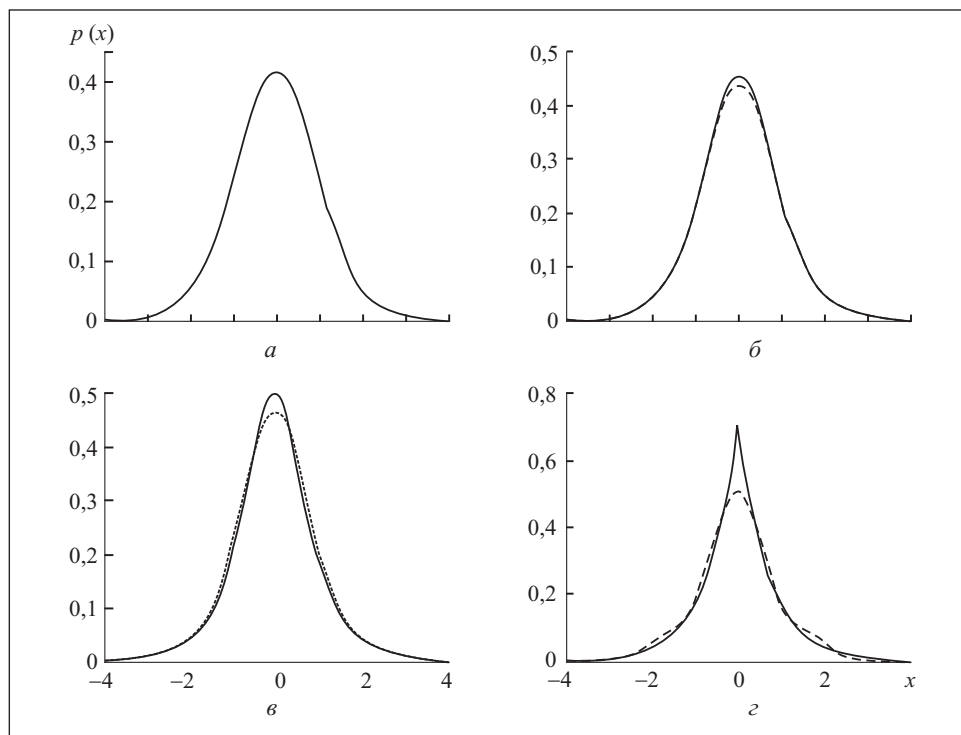


Рис. 2. Графики плотностей вероятности (сплошная линия) и их наилучших аппроксимаций (штриховая линия) для распределений: *a* — Стьюдента; *б* — логистического; *в* — Чампернауна; *з* — Лапласа

Таблица 3

$V_n(x)$	Значение r_i для распределений			
	Стьюдента	Логистическое	Чампернауна	Лапласа
V_3	0,0038	0,0115	0,0277	0,0787
V_4	0,0027	0,0270	0,0702	0,1555
V_5	0,0009	0,0230	0,0905	0,2366
V_6	0,0047	0,0913	0,4015	1,213
V_7	0,0097	0,2546	1,5334	5,7148
V_8	0,0376	0,9421	7,4428	33,964
V_9	0,2254	3,981	41,588	232,8
V_{10}	2,4822	19,208	265,26	1822,9

большого числа составляющих частных сумм не уменьшит интегральной ошибки.

Среди сумм Фейера наилучшую аппроксимацию плотностей вероятности обеспечивают суммы: $F_{13}(x)$ — для распределения Стьюдента, $F_{10}(x)$ — для распределений Чампернауна и логистического и $F_9(x)$ — для распределения Лапласа.

Среди сумм Валле—Пуссена наилучшую аппроксимацию плотностей вероятности обеспечивают суммы: $V_3(x)$ — для распределений Лапласа, Чампернауна и логистического и $V_5(x)$ — для распределения Стьюдента.

Используя обобщенные суммы, по данным, приведенным в табл. 1—3, можно вычислить значения уменьшения интегральной ошибки по формуле

$$\delta = \frac{r_S - r_{\min}}{r_S} \cdot 100\%,$$

где r_S — интегральная ошибка, соответствующая частной сумме $S_4(x)$, а r_{\min} — минимальная интегральная ошибка для данного распределения. В результате расчета получены следующие значения δ для распределений: Стьюдента — 73,5%; логистического — 36,5%; Чампернауна — 26,5%; Лапласа — 34,4%.

На рис. 2 приведены графики теоретических плотностей вероятности и их наилучших аппроксимаций, для которых значения интегральных ошибок указаны полужирным шрифтом в табл. 1—3. Из рис. 2 видно, что для распределения Стьюдента графики практически совпадают, а для остальных распределений наибольшее отклонение наблюдается при $x = 0$.

Выводы

Среди всех рассмотренных видов ортогональных представлений симметричных плотностей вероятности по полиномам Эрмита наилучшую аппроксимацию обеспечивают частные суммы, а именно: сумма Фейера для распределения Лапласа и суммы Валле—Пуссена для распределений Стьюдента, Чампернауна и логистического. При этом уменьшение интегральной ошибки в результате использования обобщенных сумм составило от 26 до 73%. С увеличением порядка сумм значение интегральной ошибки превышает единицу и наблюдаются отрицательные значения представлений.

СПИСОК ЛИТЕРАТУРЫ

1. Шелухин О.И. Негауссовские процессы в радиотехнике. — М.: Радио и связь, 1999. — 287 с.
2. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. — Л.: Энергоатомиздат, 1991. — 304 с.

3. Деч Р. Нелинейные преобразования случайных процессов / Пер. с англ. Б.А. Смиренина; под ред. Б.Р. Левина. — М.: Сов. радио, 1965. — 208 с.
4. Красильников А.И. Модели шумовых сигналов в системах диагностики теплоэнергетического оборудования. — Киев: Ин-т техн. теплофизики НАН Украины, 2014. — 112 с.
5. Крамер Г. Математические методы статистики / Пер. с англ. А.С. Монины и А.А. Петрова; под ред. А.Н. Колмогорова. — М.: Мир, 1975. — 648 с.
6. Кендалл М, Стьюарт А. Теория распределений / Пер. с англ. В.В. Сазонова, А.Н. Ширяева; под ред. А.Н. Колмогорова. — М.: Наука, 1966. — 588 с.
7. Деврой Л., Дьерфи Л. Непараметрическое оценивание плотности. L_1 -подход / Пер. с англ. А.Б. Цыбакова; под ред. М.Б. Малютова. — М.: Мир, 1988. — 408 с.
8. Красильников О.И., Березун В.С. Систематизация ортогональных подань щільності імовірності випадкових процесів // Електроніка та системи управління. — 2010. — № 3, (25). — С. 28—35.
9. Суетин П.К. Классические ортогональные полиномы. — М.: Физматлит, 2005. — 480 с.
10. Сенатов В.В. Центральная предельная теорема: Точность аппроксимации и асимптотические разложения. — М.: Книжный дом «Либроком», 2009. — 352 с.
11. Красильников А.И., Березун В.С. Использование методов обобщенного суммирования при аппроксимации плотности вероятности // Электрон. моделирование. — 2015. — 37, № 2. — С. 87—99.
12. Вадзинский Р.Н. Справочник по вероятностным распределениям. — СПб.: Наука, 2001. — 295 с.
13. Малахов А.Н. Кумулянтный анализ случайных негауссовых процессов и их преобразований. — М.: Сов. радио, 1978. — 376 с.
14. Jondeau E., Rockinger M. Gram-Charlier densities // Journal of Economic Dynamics & Control. — 2001. — Vol. 25. — P. 1457—1483.
15. Запелалов А.С., Ратнер Ю.Б. Аналитическая модель плотности вероятностей уклонов морской поверхности // Морской гидрофизический журнал. — 2003. — № 1. — С. 3—17.
16. Березун В.С., Красильников О.И. Дослідження областей невід'ємності при ортогональних поданнях щільності імовірностей // Електроніка і зв'язь. — 2010. — № 3 (56). — С. 73—78.
17. The Bernoulli Number Page. — [Электронный ресурс]. — Режим доступа: <http://bernoulli.org/>
18. The On-Line Encyclopedia of Integer Sequences. — [Электронный ресурс]. — Режим доступа: <http://oeis.org/A028296>

V.S. Beregun

RESEARCH OF APPROXIMATION ACCURACY
OF SYMMETRICAL PROBABILITY DENSITY FUNCTION BY ORTHOGONAL
REPRESENTATIONS USING HERMITE POLYNOMIALS

Approximation errors of symmetrical probability density functions of typical distributions using the Hermite polynomials have been described. Comparison of results for the partial and generalized sums of series for the Hermite polynomials is carried out at the number of series components to twenty. Orthogonal representations have been checked for non-negativity.

Key words: probability density function, approximation, Hermite polynomials, orthogonal representation, generalized summation.

REFERENCES

1. Shelukhin, O.I. (1998), *Negaussovskie protsessy v radiotekhnike* [Non-gaussian processes in radio engineering], Radio i sviaz, Moscow, Russia.
2. Novitskii, P.V. and Zograf, I.A. (1991), *Otsenka pogreshnostei rezultatov izmerenii* [Error estimation in measurement results], Energoatomizdat, St. Petersburg, Russia.
3. Dech, R. (1965), *Nelineinye preobrazovaniia sluchainykh protsessov* [Nonlinear transformations of random processes], Translated by Smirenin, B.A., Sov. radio, Moscow, Russia.
4. Krasilnikov, A.I. (2014), *Modeli shumovykh signalov v sistemakh diagnostiki teploenergeticheskogo oborudovaniia* [Models of noise signals in systems of diagnostics of the heat-and-power equipment], Institute of Engineering Thermophysics of NAS of Ukraine, Kyiv, Ukraine.
5. Kramer, G. (1975), *Matematicheskie metody statistiki* [Mathematical methods of statistics], Translated by Monin, A.S. and Petrov, A.A., Mir, Moscow, Russia.
6. Kendall, M. and Stuart, A. (1966), *Teoriia raspredelenii* [Distribution theory], Translated by Sazonov, V.V. and Shiriaev, A.N., Nauka, Moscow, Russia.
7. Devroy, L. and Derfi, L. (1988), *Neparametricheskoe otsenivanie plotnosti. L_1 -podkhod* [Nonparametric density estimation. The L_1 view], Translated by Tsybakov, A.B., Mir, Moscow, Russia.
8. Krasyl'nikov, O.I. and Beregun, V.S. (2010), "Ordering of orthogonal representations of probability density functions of stochastic processes", *Elektronika ta systemy upravlinnia*, no. 3 (25), pp. 28-35.
9. Suetin, P.K. (2005), *Klassicheskie ortogonalnye polinomy* [Classical orthogonal polynomials], Fizmatlit, Moscow, Russia.
10. Senatov, V.V. (2009), *Tsentralnaia predelnaia teorema: Tochnost' approksimatsii i asimptoticheskie razlozheniia* [Central limit theorem: approximation accuracy and asymptotic decompositions], Knizhnyi dom "Librokom", Moscow, Russia.
11. Krasil'nikov, A.I. and Beregun, V.S. (2015), "Using of the generalized summation methods at approximation of probability density function", *Elektronnoe modelirovanie*, Vol. 37, no. 2, pp. 87-99.
12. Vadzinskii, R.N. (2001), *Spravochnik po veroiatnostnym raspredeleniiam* [Directory on probabilistic distributions], Nauka, St. Petersburg, Russia.
13. Malakhov, A.N. (1978), *Kumuliantnyi analiz sluchainykh negaussovykh protsessov i ikh preobrazovaniia* [Cumulant analysis of random non-gaussian processes and their transformations], Sov. radio, Moscow, Russia.
14. Jondeau, E. and Rockinger, M. (2001), "Gram-Charlier densities", *Journal of Economic Dynamics & Control*, Vol. 25, pp. 1457-1483.
15. Zapevalov, A.S. and Ratner, B.Yu. (2003), "Analytical model of probability density function of sea surface slopes", *Morskoi gidrofizicheskii zhurnal*, no. 1, pp. 3-17.
16. Beregun, V.S. and Krasyl'nikov, O.I. (2010), "Research of non-negativity areas at orthogonal representations of probability density function", *Elektronika i sviaz*, no. 3 (56), pp. 73-78.
17. The Bernoulli number page, available at <http://bernoulli.org/>
18. The on-line encyclopedia of integer sequences, available at: <http://oeis.org/A028296>.

Поступила 15.02.16;
после доработки 28.04.16

БЕРЕГУН Виктор Сергеевич, канд. техн. наук, доцент кафедры акустики и акустоэлектроники факультета электроники Национального технического университета Украины «Киевский политехнический ин-т», который окончил в 2004 г. Область научных исследований — вероятностные характеристики и методы обработки негауссовых сигналов и их функциональных преобразований в акустических информационных системах.

