УДК 519.22 : 62-192

В.Ю. Дубницкий, И.Г. Скорикова, А.И. Ходырев

Харьковский учебно-научный институт ГВУЗ «Университет банковского дела», Харьков

ОПТИМАЛЬНАЯ АППРОКСИМАЦИЯ ФУНКЦИИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ ПО КРИТЕРИЮ МИНИМУМА ПОТЕРИ ИНФОРМАЦИИ

Предложена методика оценки качества аппроксимации функции плотности распределения вероятности с использованием расхождения (дивергенции) Кульбака. Мерой качества аппроксимации функции плотности распределения предложена величина расхождения (дивергенции) Кульбака. Критерием качества аппроксимации принят минимум дивергенции, определённой по отношению к аппроксимируемому распределению на всём множестве аппроксимирующих функций.

Получены выражения для определения расхождения (дивергенции) Кульбака для пар усеченных распределений. Применение методики показано на решении задачи выбора оптимальной аппроксимации усечённого нормального распределения среди распределений: Лапласа, двойного показательного, логистического, Чампернауна, гамма-распределения, логарифмически нормального.

В результате численного эксперимента установлено, что наилучшей интерполяцией нормального распределения на выбранном интервале усечения (280≤x≤320) будет логарифмически нормальное распределение. Полученный результат экспериментально подтверждает предположение о том, что логарифмически нормальное распределение хорошо аппроксимирует нормальное, если его коэффициент вариации ν<0,25.

Ключевые слова: *аппроксимация функции плотности распределения вероятности, расхождение (дивергенция) Кульбака, усеченные распределения, распределения: Лапласа, двойное показательное, логистическое, Чампернауна, гамма-распределение, логарифмически нормальное.*

Введение

При анализе надёжности механической системы её надёжность определяют как вероятность осуществления события

$$A = P(R > S); \tag{1}$$

где P – вероятность того, что случайная величина R, определяемая как обобщённая несущая способность конструкции, превысит значение случайной величины S, определяемой как обобщённое действующее максимальное напряжение. Законы распределения этих величин и их параметры могут быть известны заранее или определены по выборочным данным.

Решение задачи, сформулированной в виде условия (1), для различных вариантов сочетания законов распределения случайных величин R и S приведено в работах [1…3].

В этих работах не все полученные результаты приведены в явном виде, удобном для дальнейшего анализа. Способом, упрощающим анализ полученных результатов, может быть их аппроксимация более простыми выражениями [4]. В указанной работе отмечено, что одним из способов аппроксимации функции нормального распределения может быть, например, её замена функцией логистического распределения. Однако анализ качества выполненной аппроксимации в ней отсутствует.

Анализ литературы. В работе [5] была определена величина, равная величине потери количества информации при замене распределения, имеющего плотность $f_1(x)$ на распределение, имеющее плотность $f_2(x)$. Эта величина получила авторское название расхождения, позднее её стали называть расхождением или дивергенцией Кульбака и определять по условию, приведенному в работе [5]:

$$\begin{aligned} J(1,2) &= \int_{-\infty}^{\infty} \left[f_1(x) - f_2(x)\right] \ln \frac{f_1(x)}{f_2(x)} dx = \\ &= \int_{-\infty}^{\infty} \left[f_1(x) - f_2(x)\right] \cdot \left[\ln f_1(x) - \ln f_2(x)\right] dx. \end{aligned} \tag{2}$$

Использование расхождения (дивергенции) Кульбака получило применение в задачах распознавания образов и цифровой обработки сигналов [6; 7]. В доступной авторам данного сообщения литературе сведений о применении расхождения (дивергенции) Кульбака для оценки качества аппроксимации функции плотности распределения вероятности не обнаружено. Следует отметить, что в практических расчётах по оценке надёжности механических систем зачастую интерес представляет определение величины $J(1,2)$ при условии ограничений на интервал области определения аргумента. То есть случаи, когда плотности вероятностей представлены усечёнными распределениями. Определение таких распределений дано в работе [8].

Постановка задачи. Предложения к методике оценки качества аппроксимации усечённых функций плотности распределения вероятности с использованием расхождения (дивергенции) Кульбака.

Полученные результаты

Содержательный смысл использования условия (1) для оценки расхождения Кульбака рассмотрим на примере, приведенном в работе [10, С. 113-114]. В этой работе предложена аппроксимация нормального распределения распределением вида:

$$f(x) = \frac{1}{2\pi}\left(1+\cos x\right), \; -\pi < x < \pi. \qquad (3)$$

Распределение вида (3) назовем распределением C, стандартное нормальное распределение $N(0;1)$ назовём распределением N. Результаты вычисления величины J(N, C) приведены в табл. 1.

Таблица 1

Определение расхождения Кульбака J(N, C)

№ п/п, j	Интегральная вероятность	Квантили распределений		Расхождение J(N, C)	Относительная мера различия		
		Распределение N, P_{NJ}	Распределение C, P_{CJ}		P_{NJ}/P_{CJ}	Децибелы, dB	«В разы»
1	0,6	0,253	0,279	0,00254	0,90681	-0,42484	0,907
2	0,75	0,674	0,732	0,00478	0,920765	-0,35851	0,921
3	0,9	1,282	1,334	0,00207	0,961019	-0,17268	0,961
4	0,95	1,645	1,649	$9{,}71\cdot10^{-5}$	0,997574	-0,01055	0,998
5	0,975	1,96	1,888	0,00269	1,038136	0,162541	1,04
6	0,99	2,326	2,124	0,01835	1,095104	0,394552	1,09
J(N, C)	-	-	-	0,0304	-	-	-

Величину расхождения J(N, C) определяли по условию:

$$J(N,C) = \sum_{j=1}^{6}\left[\left(P_{Nj} - P_{Cj}\right)\ln\frac{P_{Nj}}{P_{Cj}}\right]. \qquad (4)$$

Из данных, приведенных в табл. 1, видно, что квантили распределения N и распределения C не совпадают. Оценки этого несовпадения для различных уровней интегральной вероятности измерены в децибелах, в этой же таблице приведено отличие этих вероятностей «В разы». Подробно эта процедура описана в работе [11]. Наглядная зависимость между этими характеристиками, построенная по данным табл. 1, показана на рис. 1.

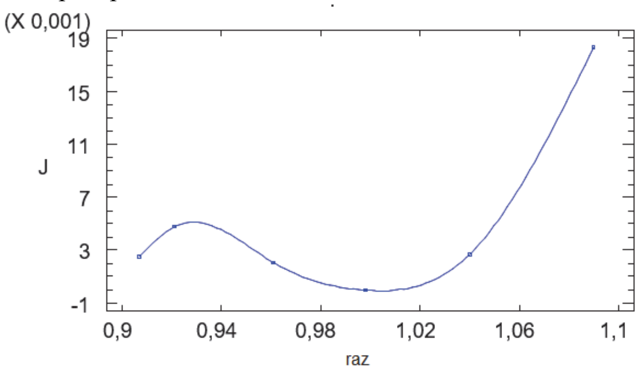

Рис. 1. Зависимость между величиной J, расхождением Кульбака J(N,C) и величиной «raz», отличием «В разы»

На графике, приведенном на рис. 1, видно, что увеличение расхождения Кульбака и увеличение относительного отличия «В разы» однонаправленные величины, что подтверждает содержательный смысл величины расхождения, как меры отличия двух распределений.

Следует обратить внимание на досадную опечатку, допущенную в работе [10, С. 113]. В ней напечатано, что среднеквадратическое отклонение для распределения C, S_c=1/14, тогда как в действительности эта величина равна 1,14, что подтверждает несложный расчёт:

– математическое ожидание m_c определим из условия:

$$m_c = \int_{-\pi}^{\pi} x\left(1+\cos x\right)dx = 0\,; \qquad (5)$$

– среднеквадратическое отклонение s_c находим из условия:

$$s_c = \int_{-\pi}^{\pi}\left[x^2\left(1+\cos x\right)dx\right]^{1/2} = \sqrt{\frac{\pi^2}{3} - 2} \approx 1{,}14\,. \qquad (6)$$

В настоящей работе рассмотрены функции плотности вероятности двух видов. Первый вид – функции плотности распределения, определённые на всей числовой оси. К ним отнесены: нормальное распределение, распределение Лапласа, двойное показательное распределение, логистическое распределение, распределение Чампернауна. К функциям второго вида отнесены функции, плотности вероятности которых определены на положительной полуоси. К ним отнесены: гамма-распределение, распределение Вейбулла, логарифмически нормальное распределение.

Необходимые сведения о плотностях распределения этих функций заимствованы из работы [8] и показаны в табл. 2. В этой таблице и далее принято, что греческие буквы соответствуют параметрам функций плотности, латинские буквы – их основным числовым характеристикам: m – математическому ожиданию, s – среднеквадратическому отклонению, v – коэффициенту вариации. Особенности определения зависимости параметров распределения Вейбулла от его начальных характеристик рассмотрены далее.

Таблица 2

Основные характеристики функций плотности распределения вероятности

Функции плотности распределения при $-\infty < x < \infty$		
Тип распределения (Условное обозначение)	Плотность распределения	Зависимость параметров распределения от его начальных характеристик
Нормальное распределение (1)	$f_1(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\mu = m$ $\sigma = s$
Лапласа (2)	$f_2(x) = \frac{1}{2}\lambda \exp\left(-\lambda\lvert x-\mu\rvert\right),$	$\mu = m$ $\lambda = \frac{\sqrt{2}}{s} = \frac{1,41421}{s}$
Двойное показательное распределение (3)	$f_3(x) = \lambda\mu \exp(-\lambda\mu - \mu e^{-\lambda \cdot x}),$	$\lambda = \frac{\pi}{s\sqrt{6}} = \frac{1,28255}{s}$ $\mu = \exp\left(\frac{\pi}{v^{**)}\sqrt{6}} - \gamma^{*)}\right) =$ $= 0,56146 \exp\left(\frac{1.28255}{v}\right)$
Логистическое распределение (4)	$f_4(x) = \frac{\exp\left(\frac{x-\mu}{\lambda}\right)}{\lambda\left[1+\exp\left(\frac{x-\mu}{\lambda}\right)\right]^2} =$ $= \frac{1}{4\lambda \mathrm{ch}^2\left(\frac{x-\mu}{2\lambda}\right)}$	$\mu = m$ $\lambda = \frac{s\sqrt{3}}{\pi} = 0,55133s$
Распределение Чампернауна (5)	$f_5(x) = \frac{\alpha}{\pi \cdot \mathrm{ch}\alpha(x-\mu)}$	$\mu = m$ $\alpha = \frac{\pi}{2s} = \frac{1,57080}{s}$
Функции плотности распределения при $0 \le x < \infty$		
Гамма-распределение (6)	$f_6(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},$ x>0.	$\alpha = 1/v^2$ $\lambda = \frac{m}{s^2}$
Распределение Вейбулла (7)	$f_7(x) = \left(\frac{c}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{c-1} \exp\left[-\left(\frac{x}{\alpha}\right)^c\right]$	Выражение в явном виде отсутствует
Логарифмически нормальное распределение (8)	$f_8(x) = \frac{1}{\alpha x\sqrt{2\pi}} \exp\left\{-\frac{[\ln(x/\mu)]^2}{2\alpha^2}\right\}$	$\mu = \ln\frac{m}{\sqrt{1+v^2}}$ $\alpha = \sqrt{\ln(1+v^2)}$

Примечание: $^{*)}\gamma$ – постоянная Эйлера, γ =0,5772; $^{**)}$v – коэффициент вариации.

Авторы вынуждены также отметить, что в использованной ими работе [8] также обнаружена опечатка. В формуле для определения величины μ в случае логарифмически нормального распределения отсутствует символ натурального логарифма.

Выражения для зависимости параметров гамма-распределения от его начальных характеристик даны в виде, приведенном в работе [12]. Решение задачи определения зависимости параметров распределения от его начальных характеристик для распределения Вейбулла описано в работе [9].

Математическое ожидание m случайной величины, распределённой по закону Вейбулла, связано с параметром распределения a условием:

$$\alpha = m \cdot \left[\Gamma\left(\frac{1}{c} + 1 \right) \right]^{-1}. \qquad (7)$$

Интерполяционные формулы для определения параметра формы c распределения Вейбулла в зависимости от коэффициента вариации v с указанием области их применения ранее были получены авторами данного сообщения в работе [9] и приведены в табл. 3.

Таблица 3

Интерполяционные формулы для определения параметра формы c распределения Вейбулла в зависимости от коэффициента вариации v

Вид модели	Границы применимости модели
$c = (0,8641 + 1,4767 \ln v)$	$1,053 \le \vartheta \le 15,83$
$c = (-0.049 + 1.0523v)^{-1}$	$0,363 \le \vartheta \le 1$
$c = \exp\left(3,7915 - 4,5365\sqrt{v}\right)$	$0,12 \le \vartheta \le 0.316$
$c = \exp(0,0741 - 1,051 \ln v)$	$0,06 \le \vartheta < 0,12$

Если вариантов аппроксимирующих распределений несколько, то для выбора наилучшей аппроксимации из r возможных следует применить правило:

$$L = \min_j J(1, j) = \min_j \int_{-\infty}^{\infty} \left[f_1(x) - f_j(x) \right] \ln \frac{f_1(x)}{f_j(x)} dx; \quad j = 2, 3, \ldots r. \qquad (8)$$

Плотность усеченного распределения в общем случае, в соответствии с работой [8], имеет вид:

$$f(x; a, b) = \begin{cases} 0, & x < a, x > b; \\ \dfrac{f(x)}{F(b) - F(a)}, & a \le x \le b. \end{cases} \qquad (9)$$

Поставленная задача решалась при условии, что интервалы усечения исходного и аппроксимирующего распределений равны.

Рассмотрим первый, основной, случай. Отрезок усечения двусторонний и определён условием: $a \le x \le b$.

Используя условия (2) и (9), получим выражение расхождения Кульбака в виде:

$$J(1, 2) = \int_a^b \left[\frac{f_1(x)}{F_1(b) - F_1(a)} - \frac{f_2(x)}{F_2(b) - F_2(a)} \right] \ln \left[\frac{f_1(x)}{F_1(b) - F_1(a)} \Big/ \frac{f_2(x)}{F_2(b) - F_2(a)} \right] dx. \qquad (10)$$

Примем, что:

$$k_1 = \frac{F_2(b) - F_2(a)}{F_1(b) - F_1(a)}; \qquad (11)$$

тогда:

$$J(1, 2) = \int_a^b \left[\frac{f_1(x)}{F_1(b) - F_1(a)} - \frac{f_2(x)}{F_2(b) - F_2(a)} \right] \ln \left[k_1 \frac{f_1(x)}{f_2(x)} \right] dx. \qquad (12)$$

Рассмотрим второй случай, при котором интервал усечения определён условием: $-\infty < x \le b$. Так как функция распределения вероятности $F(-\infty) = 0$, то функция плотности вероятности, усечённая на заданном интервале, примет вид:

$$f(x; a, b) = \begin{cases} 0, & x > b; \\ \dfrac{f(x)}{F(b)}, & -\infty < x \le b. \end{cases} \qquad (13)$$

Величину расхождение Кульбака в этом случае определим по условию:

$$J(1, 2) = \int_{-\infty}^b \left[\frac{f_1(x)}{F_1(b)} - \frac{f_2(x)}{F_2(b)} \right] \ln \left[k_2 \frac{f_1(x)}{f_2(x)} \right] dx; \qquad (14)$$

где

$$k_2 = \frac{F_2(b)}{F_1(b)}. \qquad (15)$$

Рассмотрим третий случай, при котором интервал усечения определён условием: $a \le x \le \infty$. Учитывая, что $F(\infty) = 1$, определим плотность распределения в виде:

$$f(x; a, b) = \begin{cases} 0, & x < a, \\ \dfrac{f(x)}{1 - F(a)}, & a \le x < \infty. \end{cases} \qquad (16)$$

Расхождение Кульбака примет вид:

$$J(1,2)=\int_a^{\infty}\left[\frac{f_1(x)}{1-F_1(a)}-\frac{f_2\ (x)}{1-F_2(a)}\right]\ln\left[k_3\frac{f_1(x)}{f_2(x)}\right]dx\ , \quad (17)$$

где

$$k_3=\frac{1-F_2(b)}{1-F_1(b)}\,. \quad (18)$$

Применение предложенной методики рассмотрим на численном промере. Пусть случайная величина X имеет математическое ожидание m=200 и среднеквадратическое отклонение S=40, следовательно, её коэффициент вариации ν=0,2. Численные значения параметров законов распределения, использованных в данной работе, вычислены по выражениям, указанным в табл. 2 и табл. 3. Их численные значения приведены в табл. 4, 5.

Таблица 4

Численные значения параметров законов распределения вероятности, которые определены при $-\infty < x < \infty$

Типы законов распределения, использованные в работе									
Нормальный		Лапласа		Двойной показательный		Логистический		Чампернауна	
μ	σ	μ	λ	μ	λ	μ	λ	μ	α
200	40	200	0,03535	342,24787	0,03206	200	22,05315	200	0,03927

Таблица 5

Численные значения параметров законов распределения вероятности, которые определены при $0 \le x < \infty$

Типы законов распределения, использованные в работе					
Вейбулла		Гамма-распределение		Логарифмически нормальное	
α	c	α	λ	α	μ
215.93098	5,82824	25	0,125	0,19804	5,278

При решении примера принимали для распределений, определённых при $-\infty < x < \infty$, в качестве аппроксимируемого распределения – нормальное, аппроксимирующими служили остальные распределения, указанные в табл. 1. Вид функций распределения вероятности, использованных в работе, приведен в табл. 6.

Таблица 6

Функции распределения вероятности, использованные в работе*)

Тип функции распределения	Вид функции распределения	Тип функции распределения	Вид функции распределения
Нормальное	$F(x)=\Phi(x)= =\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-t^2/2)dt$	Чампернауна	$F(x)=\frac{2}{\pi}\operatorname{arctg}\exp\left[\alpha\left(x-\mu\right)\right]$
Лапласа	$F(x)=\begin{cases}\frac{1}{2}e^{-\lambda(x-\mu)}, & x\le\mu\\ 1-\frac{1}{2}e^{-\lambda(x-\mu)}, & x>\mu\end{cases}$	Гамма-распределение	$F(x)=I\left(x/\beta;\alpha\right)= =\left(\beta^{\alpha}\Gamma\left(\alpha\right)\right)^{-1}\int_0^x x^{\alpha-1}e^{-x/\beta}dx$
Двойное показательное	$F(x)=\exp\left(-\mu e^{-\lambda x}\right)$	Вейбулла	$F(x)=1-\exp\left[-\left(\frac{x}{\alpha}\right)^c\right]$
Логистическое	$F(x)=\frac{1}{2}\left[1+\operatorname{th}\left(\frac{x-\mu}{2\lambda}\right)\right]$	Логарифмически нормальное	$F(x)=\Phi\left(\frac{\ln\left(x/\mu\right)}{\alpha}\right)$

Примечание: *) В табл. 6 для гамма- распределения принято, что β=1/λ.

Следует отметить следующее. Общепринятые выражения для функций распределения и эквивалентные им, но принятые в конкретном программном продукте, могут различаться. Это хорошо видно из сравнения выражений для функции гамма-распределения, приведенных в работах [4] и [12]. Пренебрежение этим обстоятельством приводит, как правило, к ошибочным результатам. В частности, в системе Excel для вычисления значений гамма-распределения синтаксис выражения следующий:

ГАММАРАСП(x, альфа; бета; 1)

Следует иметь в виду, что в этом случае *бета* =1/*λ*.

Для вычисления распределения Вейбулла синтаксис выражения следующий:

ВЕЙБУЛЛ (x; альфа; бета; 1)

Следует иметь в виду, что в этом случае *альфа* =c и *бета* =α.

Результаты вычисления расхождения Кульбака для выбранных пар распределений, условные обозначения которых показаны в табл. 2, выполнены методом численного интегрирования условия (12) и приведены в табл. 7.

Таблица 7

Расхождения (дивергенции) Кульбака

Величина расхождения (дивергенции) Кульбака для пар распределений						
J (1,2)	J (1,3)	J (1,4)	J (1,5)	J (1,6)	J (1,7)	J (1,8)
164,689	228,8559	565,0577	161,3026	163,9974	167,226	1,22434

Используя критерий (8) следует считать, что в данном случае наилучшей интерполяцией нормального распределения на выбранном интервале усечения ($280 \le x \le 320$) будет логарифмически нормальное распределение. Полученный результат экспериментально подтверждает предположение, сделанное в работе [10, С. 113-114] о том, что логарифмически нормальное распределение хорошо аппроксимирует нормальное, если его коэффициент вариации v<0,25.

Выводы

1. Предложена методика оценки качества аппроксимации функции плотности распределения вероятности с использованием расхождения (дивергенции) Кульбака.
2. Мерой качества аппроксимации функции плотности распределения предложена величина расхождения (дивергенции) Кульбака.
3. Критерием качества аппроксимации принят минимум дивергенции, определённой по отношению к аппроксимируемому распределению на всём множестве аппроксимирующих функций.
4. Получены выражения для определения расхождения (дивергенции) Кульбака для пар усеченных распределений.
5. Применение методики показано на решении задачи выбора оптимальной аппроксимации усечённого нормального распределения одним из принятых законов распределения: Лапласа, двойного показательного, логистического, Чампернауна, гамма-распределения, логарифмически нормального.
6. В результате численного эксперимента установлено, что наилучшей интерполяцией нормального распределения на выбранном интервале усечения ($280 \le x \le 320$) будет логарифмически нормальное распределение.
7. Полученный результат экспериментально подтверждает ранее сделанные предположения о том, что логарифмически нормальное распределение хорошо аппроксимирует нормальное, если его коэффициент вариации v<0,25.

Список литературы

1. *Арасланов А.М. Расчет элементов конструкций заданной надёжности при случайных воздействиях / А.М. Арасланов. – Москва: «Машиностроение», 1987. – 128 с.*
2. *Капур К. Надёжность и проектирование систем / К. Капур, Л. Ламберсон. – Москва: «МИР», 1980. – 604 с.*
3. *Переверзев Е.С. Надёжность и испытание технических систем / Е.С. Переверзев. – К.: «Наукова думка», 1990. – 327 с.*
4. *Дубницкий В.Ю. Аппроксимация функции нормального распределения функцией логистического распределения и её применение для определения надёжности механических систем / В.Ю. Дубницкий, Г.В. Фесенко, И.А. Черепнев // Вісник Харківського Національного технічного університету сільського господарства імені Петра Василенка. Технічні Науки. «Механізація сільськогосподарського виробництва». – 2017. – Вип. 180. – С. 168-181.*
5. *Кульбак С. Теория информации и статистика / С. Кульбак. – Москва.: Изд. «Наука», 1967. – 408 с.*
6. *Фукунага К. Введение в статистическую теорию распознавания образов / К Фукунага. – Москва: «Наука», 1979. – 368 с.*
7. *Купченко Л.Ф. Критерий согласованности оптимальной обработки сигналов в оптико-электронных системах с динамической спектральной фильтрацией / Л.Ф. Купченко, А.С. Рыбьяк // Системи озброєння і військова техніка.– 2015. – № 1 (41). – С. 120-123.*
8. *Вадзинский Р.Н. Справочник по вероятностным распределениям / Р.Н. Вадзинский. – Санкт-Петербург.: Изд. «Наука», 2001. – 295 с.*
9. *Дубницкий В.Ю. Решение в неявном виде обратной задачи моделирования непрерывной случайной величины. / В.Ю. Дубницкий, И.Г. Скорикова // Системи обробки інформації. – Х.: ХУПС, 2015. – Вип. 3(128). – С. 47-52.*
10. *Джонсон Н.Л. Одномерные непрерывные распределения. В 2 ч. Ч.1 / Н.Л. Джонсон, С. Коц, Н. Балакришнан. – Москва: БИНОМ. Лаборатория знаний, 2010-2012. – 703 с.*
11. *Дубницкий В.Ю. Определение относительной оценки тяжести хвоста распределения – уровня хвоста /*

В.Ю. Дубницкий, А.И. Ходырев // Системи обробки інформації. – Х.: ХУПС, 2015. – Вип. 7 (132). – С. 83-92.

12. Вадзинский Р. Статистические вычисления в среде Excel. Библиотека пользователя. / Р. Вадзинский. – Санкт-Петербург.: Изд. «Питер», 2008. –608 с.

Поступила в редколлегию 24.04.2017

Рецензент: д-р техн. наук проф. Г.А. Кучук, Национальный технический университет «ХПИ», Харьков.

ОПТИМАЛЬНА АПРОКСИМАЦІЯ ФУНКЦІЇ ЩІЛЬНОСТІ РОЗПОДІЛУ ЗА КРИТЕРІЄМ МІНІМУМУ ВТРАТИ ІНФОРМАЦІЇ

В.Ю. Дубницький, І.Г. Скорікова, О.І. Ходирєв

Запропоновано методику оцінки якості апроксимації функції щільності розподілу ймовірності з використанням розбіжності (дивергенції) Кульбака. Мірою якості апроксимації функції щільності розподілу запропоновано величину розбіжності (дивергенції) Кульбака. За критерій якості апроксимації прийнято мінімум дивергенції, визначеної по відношенню до розподілу, що апроксимується, на всій множині апроксимуючих функцій.

Отримано вирази для визначення розбіжності (дивергенції) Кульбака для пар зрізаних розподілів. Використання методики показано на розв'язанні задачі вибору оптимальної апроксимації зрізаного нормального розподілу серед розподілів: Лапласа, подвійного показникового, логістичного, Чампернауна, гамма-розподілу, логарифмічно нормального розподілу.

В результаті чисельного експерименту встановлено, що найкращою інтерполяцією нормального розподілу на вибраному інтервалі зрізу ($280 \le x \le 320$ *) буде логарифмічно нормальний розподіл. Отриманий результат експериментально підтверджує раніш зроблене припущення про те, що логарифмічно нормальний розподіл добре апроксимує нормальний, якщо його коефіцієнт варіації v<0,25.*

Ключові слова: *апроксимація функції щільності розподілу ймовірності, розбіжність (дивергенція) Кульбака, зрізані розподіли, розподіли: Лапласа, подвійний показниковий, логістичний, Чампернауна, гамма-розподіл, логарифмічно нормальний розподіл.*

OPTIMAL APPROXIMATION OF DENSITY FUNCTION BY MINIMUM INFORMATION LOSS CRITERION

V. Dubnitskiy, I. Skorikova, A. Khodyrev

A method proposed for approximation quality estimate of probability distribution density using Kullback divergence. The proposed measure of density function approximation quality is the value of Kullback divergence which is determined relative to distribution to be approximated over the totality of approximating functions.

Expressions found for determination of Kullback divergence for pairs of truncated distributions. Application of the method is shown by solution of optimal approximation selection problem for truncated normal distribution among the following distributions: Laplace distribution, double exponential distribution, logistic distribution, Champernowne distribution, gamma distribution, lognormal distribution.

After a numerical experiment it was found that the best normal distribution interpolation at selected truncation interval ($280 \le x \le 320$ *) will be lognormal distribution. The obtained result serves as experimental confirmation of the previous assumption that lognormal distribution quite well approximates normal distribution provided its coefficient of variation is v<0,25.*

Keywords: *approximation of probability distribution density, Kullback divergence, truncated distributions, Laplace distribution, double exponential distribution, logistic distribution, Champernowne distribution, gamma distribution, lognormal distribution.*