

## Применение теории канонических моментов для оценивания плотности случайной величины, распределенной на отрезке\*

В. И. ДЕНИСОВ, В. С. ТИМОФЕЕВ, В. Ю. ЩЕКОЛДИН  
*Новосибирский государственный технический университет, Россия*  
e-mail: videnis@nstu.ru, netsc@fpm.ami.nstu.ru, raix@ngs.ru

Рассмотрена задача идентификации случайных величин, распределенных на финитных интервалах. Получено разложение функции плотности искомой случайной величины в ряд по ортогональным полиномам на основе теории канонических моментов (в качестве базовой плотности использовано семейство бета-распределений). Предложен алгоритм оценивания функции плотности случайной величины, распределенной на финитном интервале. Представлены результаты применения алгоритма для оценивания ряда распределений.

*Ключевые слова:* случайная величина, оценивание плотности распределения, канонические моменты, ортогональное разложение, бета-распределение.

### Введение

Во многих практических приложениях, связанных со статистической обработкой экспериментальной информации, возникает задача идентификации распределений исследуемых показателей. На сегодня существует широкий спектр методов, позволяющих решать эту задачу при тех или иных условиях, определяемых спецификой предметной постановки задачи, целей исследования, экспериментальных данных и имеющимися в распоряжении исследователя техническими и технологическими средствами.

На практике при проведении исследований, связанных с решением задачи идентификации распределений, используются, как правило, только стандартные способы оценивания распределений, такие как методы типа хи-квадрат Пирсона, максимального правдоподобия, критерии согласия типа критериев Смирнова, Колмогорова, омега-квадрат и т. д., известные еще по работам [1–3]. Однако анализ современного состояния рассматриваемой проблемы [4–7] показывает, что наиболее перспективны специализированные методы, основанные на использовании непараметрических статистик и подходов, базирующихся на теоретических достижениях математической статистики. В этой связи следует отметить работы, посвященные применению теории вероятностных мер и ее приложений, таких как планирование оптимальных экспериментов [5], устойчивое оценивание отдельных характеристик распределений [6], устойчивое оценивание регрессионных моделей в целом [4] и др.

---

\*Работа выполнена при финансовой поддержке Министерства образования и науки РФ (код проекта РНП.2.1.2.43).

© ИВТ СО РАН, 2009.

Один из альтернативных подходов к изучению распределений предоставляет теория канонических моментов, предложенная М. Скибински [8] и развитая впоследствии Х. Детте и В.Дж. Стадденом [5]. Особенность данного подхода состоит в том, что для случайных величин, распределенных на финитном интервале, могут быть применены специальные методы, инвариантные относительно преобразований сдвига и масштаба. В частности, это позволяет рассматривать с одной и той же точки зрения целые семейства случайных величин, обладающих общими свойствами, которые могут быть описаны в терминах канонических моментов.

При решении различных технических задач, связанных с измерениями (наблюдениями) значений случайных величин, соответствующих конкретным техническим показателям, естественно рассмотрение конечных (финитных) интервалов значений этих показателей. Это связано с тем, что, с одной стороны, даже если случайная величина теоретически распределена на неограниченном интервале, ее реально наблюдаемые значения будут конечными. С другой стороны, вследствие своей природы, многие технические показатели изначально не могут принимать бесконечно большие или бесконечно малые значения.

Такие ситуации возникают, например, в задачах статистического контроля качества, теории надежности, в исследованиях медицины, биологии, экологии, сельского хозяйства и др. Следовательно, условие финитности интервала распределения случайной величины является естественно вытекающим из условия решаемой задачи, т.е. скорее необходимым условием получения корректных практических результатов. Но нельзя забывать о том, что при решении экстремальных задач (связанных, например, с теорией катастроф) к выбору финитного интервала распределения нужно подходить с известной осторожностью, основываясь на специфике априорной информации об изучаемом техническом объекте.

Во многих исследованиях, связанных со случайными величинами, распределенными на финитных интервалах, используется бета-распределение. Оно удобно в первую очередь тем, что имеет частными случаями такие широко применяемые на практике распределения, как равномерное, распределение арксинуса,  $F$ -распределение Снедекора—Фишера, гамма-распределение, которое является пределом бета-распределения, и др.

С другой стороны, известен метод оценивания функции плотности распределения, основанный на построении ортогонального разложения, частичные суммы которого с той или иной степенью точности приближают искомую плотность. Первые исследования такого рода относятся к работам К.В.Л. Шарлье [9] начала XX века, в которых было предложено разложение Грама—Шарлье для нормальной плотности. Позднее подход был обобщен в [2] для плотностей произвольного вида.

В представляемой статье предлагается методика оценивания плотностей случайных величин, распределенных на отрезках, на основе синтеза теорий канонических моментов и ортогональных разложений плотностей.

## 1. Постановка задачи

Пусть по результатам эксперимента получены статистические данные  $\{x_1, x_2, \dots\}$ , суть измерения значений некоторой случайной величины  $\xi$ , распределенной на финитном интервале. Без потери общности [5] будем считать, что областью допустимых значений  $\xi$  является отрезок  $[0, 1]$ , поскольку для случайных величин, распределенных на произ-

вольном отрезке, скажем,  $[a, b]$ , при помощи аффинного преобразования  $\tilde{x}_i = \frac{x_i - a}{b - a}$  реализаций случайной величины можно получить результаты, полностью аналогичные излагаемым ниже.

Задача идентификации распределения случайной величины  $\xi$  по значениям полной последовательности ее моментов представляет собой известную задачу Хаусдорфа [10]. В силу финитности задача Хаусдорфа разрешима, следовательно, может быть однозначно определена функция плотности  $f_\xi(x)$ .

Основное внимание в работе будет уделено построению ортогонального разложения плотности в ряд по полиномам Якоби, порожденным бета-распределением. При этом коэффициенты искомого разложения будут выражаться в терминах канонических моментов оцениваемой случайной величины  $\xi$ .

## 2. Построение ортогонального разложения функции плотности

Пусть  $\mu$  — случайная величина, имеющая бета-распределение на отрезке  $[0, 1]$ , плотность которой имеет вид

$$w_{\alpha, \beta}(x) = \frac{1}{B(\alpha + 1, \beta + 1)}(1 - x)^\alpha x^\beta, \quad (1)$$

где  $\alpha, \beta > -1$ .

Последовательность полиномов  $\{P_k^{(\alpha, \beta)}(x)\}_{k \geq 0}$  — ортогональна относительно плотности (1), если

$$\int_0^1 P_k^{(\alpha, \beta)}(x) P_l^{(\alpha, \beta)}(x) w_{\alpha, \beta}(x) dx = 0, \quad k \neq l, \quad (2)$$

где  $P_k^{(\alpha, \beta)}(x)$  — полином степени  $k$ . Полиномы в (2) являются ортогональными полиномами Якоби. Хорошо известно, что на основе системы ортогональных полиномов могут быть построены разложения произвольных функций (удовлетворяющих некоторым условиям регулярности) в ряды. В рассматриваемой задаче нас будут интересовать разложения функций плотности вида

$$f_\xi(x) = w_{\alpha, \beta}(x) \sum_{k=0}^{\infty} r_k P_k^{(\alpha, \beta)}(x), \quad (3)$$

где  $w_{\alpha, \beta}(x)$  — базовая плотность разложения.

Согласно [11, 12], определение коэффициентов  $r_k$  разложения (3) может быть проведено путем интегрирования этого выражения с соответствующим полиномом степени  $k$ . Тогда

$$r_k = \int_0^1 P_k^{(\alpha, \beta)}(x) f_\xi(x) dx, \quad k = 0, 1, 2, \dots \quad (4)$$

С другой стороны, полиномы Якоби могут быть представлены линейными комбинациями системы степеней  $\{1, x, x^2, \dots\}$ , а именно в виде

$$P_0^{(\alpha, \beta)}(x) = 1, \quad P_k^{(\alpha, \beta)}(x) = \sum_{i=0}^k a_{ki} x^i. \quad (5)$$

При подстановке (5) в (4) возникают интегралы вида  $\int x^i f_\xi(x) dx$ , которые определяют начальные моменты  $m_i$  случайной величины  $\xi$ . Таким образом, коэффициенты (4) будут однозначно определяться последовательностью моментов искомой случайной величины, а именно

$$r_0 = 1, \quad r_k = a_{k0} + \sum_{i=2}^k a_{ki} m_i, \quad k = 1, 2, \dots \quad (6)$$

Отметим, что на практике вместо соотношения (6) удобнее использовать его аналог для системы ортонормированных полиномов Якоби и, следовательно, для центрированной случайной величины  $\tilde{\xi} = \xi - m_1$ . В этом случае (6) имеет вид

$$r_0 = 1, \quad r_1 = a_{10}, \quad r_2 = a_{20} + a_{22} \mu_2, \\ r_k = a_{k0} + \sum_{i=3}^k a_{ki} \mu_i, \quad k = 3, 4, \dots, \quad (7)$$

где  $\mu_i$  — центральные моменты случайной величины  $\xi$ .

Заметим, что для перехода к ортонормированным полиномам Якоби  $\{\tilde{P}_k^{(\alpha, \beta)}(x)\}_{k \geq 0}$  необходимо, следуя [13], провести преобразование

$$\tilde{P}_k^{(\alpha, \beta)}(x) = \sqrt{\frac{k!(\alpha + \beta + 2k + 1)}{\Gamma(\alpha + k + 1)\Gamma(\beta + k + 1)}} P_k^{(\alpha, \beta)}(x).$$

При этом соотношение (2) принимает более общий вид:

$$\int_0^1 \tilde{P}_k^{(\alpha, \beta)}(x) \tilde{P}_l^{(\alpha, \beta)}(x) w_{\alpha, \beta}(x) dx = \delta_{kl},$$

где  $\delta_{kl}$  — символ Кронекера.

Согласно [5] последовательность обычных моментов случайной величины находится во взаимно-однозначном соответствии с последовательностью канонических моментов этой случайной величины, которая задается следующим образом. Если известно значение  $m_k$  для искомой случайной величины  $\xi$  и значения  $m_k^-$  и  $m_k^+$ , определяемые наименьшим и наибольшим возможными значениями этого же момента у всех случайных величин, первые  $k - 1$  предыдущих моментов которых совпадают, то канонический момент порядка  $k$  случайной величины  $\xi$  определяется соотношением

$$p_k = \frac{m_k - m_k^-}{m_k^+ - m_k^-}.$$

На практике вычисление последовательности канонических моментов строго по определению представляется весьма сложной аналитической задачей [5]. Поэтому применяются специальные процедуры, в частности, Q-D-алгоритм, исследованный, например, в [14]. Он основывается на представлении канонических моментов в терминах определителей Ганкеля  $\overline{H}_k$  и  $\underline{H}_k$  [15]:

$$m_k - m_k^- = \frac{\underline{H}_k}{\underline{H}_{k-2}}, \quad m_k^+ - m_k = \frac{\overline{H}_k}{\overline{H}_{k-2}}, \quad p_k = \frac{\underline{H}_k \overline{H}_{k-2}}{\underline{H}_k \overline{H}_{k-2} + \underline{H}_{k-2} \overline{H}_k},$$

где  $\underline{H}_{-1} = \underline{H}_0 = \overline{H}_{-1} = \overline{H}_0 = 1$ , а также

$$\underline{H}_{2k} = \begin{vmatrix} m_0 & \cdots & m_k \\ \vdots & \ddots & \vdots \\ m_k & \cdots & m_{2k} \end{vmatrix}, \quad \underline{H}_{2k+1} = \begin{vmatrix} m_1 & \cdots & m_{k+1} \\ \vdots & \ddots & \vdots \\ m_{k+1} & \cdots & m_{2k+1} \end{vmatrix},$$

$$\overline{H}_{2k} = \begin{vmatrix} m_1 - m_2 & \cdots & m_k - m_{k+1} \\ \vdots & \ddots & \vdots \\ m_k - m_{k+1} & \cdots & m_{2k-1} - m_{2k} \end{vmatrix},$$

$$\overline{H}_{2k+1} = \begin{vmatrix} m_0 - m_1 & \cdots & m_k - m_{k+1} \\ \vdots & \ddots & \vdots \\ m_k - m_{k+1} & \cdots & m_{2k} - m_{2k+1} \end{vmatrix}.$$

Кроме собственно канонических моментов, для решения проблемы Хаусдорфа часто применяются так называемые ассоциированные  $\zeta$ -характеристики случайной величины, определяемые как

$$\zeta_0 = 1, \quad \zeta_1 = p_1, \quad \zeta_k = (1 - p_{k-1})p_k, \quad k = 2, 3, \dots$$

Эти характеристики имеют более удобное, чем канонические моменты, представление в терминах определителей Ганкеля, а именно

$$\zeta_k = \frac{\underline{H}_k \underline{H}_{k-3}}{\underline{H}_{k-1} \underline{H}_{k-2}}.$$

Таким образом, коэффициенты разложения по ортогональным полиномам Якоби (4) могут быть представлены не только как функции обычных моментов (6), но и как функции канонических моментов и/или  $\zeta$ -характеристик. Решение этой задачи может быть получено на основе следующей известной теоремы, доказательство которой можно найти в [15].

**Теорема (Скибински).** Если  $\overline{m}_n = (m_1, \dots, m_n) \in \text{Int } M_n$ , то

$$m_n = S_{n,n} = \sum_{i=0}^{[n/2]} \left[ S_{i,n-i}^2 \prod_{j=1}^{n-2i} \zeta_j \right],$$

где  $S_{i,j} = S_{i,j-1} + \zeta_{j-i+1} S_{i-1,j}$ ,  $1 \leq i \leq j$  и  $S_{i,j} = 0$ , если  $0 \leq j < i$ ,  $S_{0,j} = 1$  при  $j \geq 0$ .

В теореме Скибински для пространства моментов  $M$  определена величина  $\text{Int } M = \{\overline{m} \in M : \overline{m}_n \in M_n \forall n \in N\}$ , а  $M_n$  является  $n$ -мерным подпространством пространства  $M$ , определяемым первыми  $n$  моментами случайной величины.

Возвращаясь к искомому разложению (3), следует отметить, что вычисление всех его членов не имеет практического смысла, поскольку обычно в распоряжении исследователя имеется информация, позволяющая более или менее точно оценить лишь первые несколько моментов рассматриваемой случайной величины. Известно также, что с увеличением порядка момента резко ухудшается качество его оценивания даже для тех методов, которые отличаются высокой степенью устойчивости [7].

С другой стороны, для исследователя важной составляющей является возможность качественной интерпретации получаемых результатов, что достигается вычислением

основных статистических характеристик, определяющих форму распределения. К ним относятся математическое ожидание, дисперсия, коэффициенты асимметрии и эксцесса. Следовательно, с этой точки зрения в разложении (3) имеет смысл рассматривать члены до четвертого порядка включительно.

Представим выражение (3) в виде

$$f_\xi(x) = w_{\alpha,\beta}(x) \sum_{k=0}^{\infty} r_k \tilde{P}_k^{(\alpha,\beta)}(x) = w_{\alpha,\beta}(x) \left( \tilde{\Psi}_n(x) + \tilde{R}_n(x) \right) = \Psi_n(x) + R_n(x), \quad (8)$$

где  $\tilde{\Psi}_n(x)$  —  $n$ -я частичная сумма ряда (3), а  $\tilde{R}_n(x)$  —  $n$ -й остаток ряда,  $\Psi_n(x) = w_{\alpha,\beta}(x) \tilde{\Psi}_n(x)$ ,  $R_n(x) = w_{\alpha,\beta}(x) \tilde{R}_n(x)$ . Тогда, пренебрегая остатком ряда, будем считать, что  $f_\xi(x) \approx \Psi_n(x)$ , причем далее будем рассматривать случаи, когда  $n \leq 4$ .

Пользуясь теоремой Скибински, определим выражения для начальных моментов случайной величины  $\xi$ :

$$m_1 = \zeta_1, \quad m_2 = \zeta_1(\zeta_1 + \zeta_2), \quad m_3 = \zeta_1((\zeta_1 + \zeta_2)^2 + \zeta_2\zeta_3), \\ m_4 = \zeta_1((\zeta_1 + \zeta_2)^3 + \zeta_2\zeta_3(2(\zeta_1 + \zeta_2) + \zeta_3 + \zeta_4)).$$

Применив известные соотношения, связывающие начальные и центральные моменты [3], получаем

$$\mu_1 = 0, \quad \mu_2 = \zeta_1\zeta_2, \quad \mu_3 = -\zeta_1\zeta_2(\zeta_1 - \zeta_2 - \zeta_3), \\ \mu_4 = \zeta_1\zeta_2((\zeta_1 - \zeta_2 + \zeta_3)^2 + \zeta_1\zeta_2 - \zeta_3\zeta_4).$$

Необходимость в этих соотношениях определяется выбором разложения (6) или (7), при этом последнее применяется, если случайная величина центрирована.

Ортогональные полиномы Якоби вычисляются по рекуррентным соотношениям через  $\zeta$ -характеристики:

$$\tilde{P}_1^{(\alpha,\beta)}(x) = x - \zeta_1, \\ \tilde{P}_{k+1}^{(\alpha,\beta)}(x) = (x - \zeta_{2k} - \zeta_{2k+1})\tilde{P}_k^{(\alpha,\beta)}(x) - \zeta_{2k-1}\zeta_{2k}\tilde{P}_{k-1}^{(\alpha,\beta)}(x). \quad (9)$$

Для перехода от (9) к (5) определим коэффициенты  $a_{ki}$  в терминах  $\zeta$ -характеристик:

$$a_{10} = -\zeta_1, \quad a_{20} = \zeta_1\zeta_3, \quad a_{22} = 1, \\ a_{30} = -\zeta_1\zeta_3\zeta_5, \quad a_{32} = -\sum_{i=1}^5 \zeta_i, \quad a_{33} = 1, \\ a_{40} = \prod_{i=1}^4 \zeta_{2i-1}, \quad a_{42} = \sum_{i=1}^5 \left( \zeta_i \sum_{j=i+2}^7 \zeta_j \right), \quad a_{43} = -\sum_{i=1}^7 \zeta_i, \quad a_{44} = 1.$$

Подставляя значения этих коэффициентов в (7) и заменяя соответствующие моменты случайной величины их оценками, вычисленными по выборке, получаем оценку искомой функции плотности в виде  $n$ -й частичной суммы ряда (7):

$$\hat{f}_\xi^{(n)}(x) \approx w_{\alpha,\beta}(x) \sum_{k=0}^n r_k \tilde{P}_k^{(\alpha,\beta)}(x), \quad (10)$$

где  $r_k$  определяются из соотношений (7), а  $n$  может принимать значения 2, 3 или 4.

Следует отметить некоторые закономерности, возникающие в соотношениях, связывающих коэффициенты ортогональных полиномов (5) и канонические моменты случайной величины (и соответствующие определители Ганкеля).

**Утверждение 1.** При  $i \geq 1$  верны следующие соотношения:

- 1)  $a_{ii} = 1$ ;
- 2)  $a_{i0} = (-1)^i \prod_{k=1}^i \zeta_{2k-1} = (-1)^i \frac{H_{2i-1}}{H_{2i-2}}$ ;
- 3)  $a_{i,i-1} = - \sum_{k=1}^{2i-1} \zeta_k$ ;
- 4)  $a_{i,i-2} = \sum_{k=1}^{2i-3} \left( \zeta_k \sum_{l=k+2}^{2i-1} \zeta_l \right)$ .

Доказательство п. 1–4 утверждения 1 проводится непосредственными вычислениями коэффициентов ортогональных полиномов. Кроме того, второе равенство п. 2 получается применением свойств определителей Ганкеля [15].

Далее имеет смысл обсудить вопрос о корректном выборе параметров базовой функции бета-распределения  $w_{\alpha,\beta}(x)$ . В отличие от рассматриваемого в [3, 11] разложения Грама—Шарлье по симметричной нормальной плотности, в данном случае возникает проблема, связанная с тем, что при разных соотношениях параметров  $\alpha, \beta$  плотность распределения имеет существенно различные формы. При  $\alpha > \beta$  распределение обладает положительной асимметрией, а при  $\alpha < \beta$  — отрицательной. Чтобы учесть этот факт, необходимо при построении разложения (2) соотнести свойства плотности (1) и свойства искомой плотности  $f_{\xi}(x)$ , определяемые по выборке. Для этой цели авторами предлагается следующее

**Утверждение 2.** Центральные моменты двух случайных величин  $\eta_1$  и  $\eta_2$ , распределенных на отрезке  $[0, 1]$ , со взаимно симметричными функциями плотности  $f_1(x)$  и  $f_2(x)$  соответственно, удовлетворяют соотношению

$$\mu_k^{(2)} = (-1)^k \mu_k^{(1)}.$$

*Доказательство.* Условие взаимной симметричности функций  $f_1(x)$  и  $f_2(x)$  означает, что верно

$$f_2(x) = f_1(1-x). \quad (11)$$

Тогда центральные моменты случайной величины  $\eta_1$  вычисляются по формулам

$$\mu_k^{(1)} = E(\eta_1 - E\eta_1)^k = \int_0^1 (x - m_1^{(1)})^k f_1(x) dx,$$

где  $m_1^{(1)}$  — первый начальный момент случайной величины  $\eta_1$ . Для случайной величины  $\eta_2$  в силу (11) этот момент будет иметь вид

$$m_1^{(2)} = \int_0^1 x f_2(x) dx = \int_0^1 x f_1(1-x) dx = \int_0^1 (1-x) f_1(x) dx = 1 - m_1^{(1)}.$$

Пользуясь последним соотношением и (11), определим центральные моменты случайной величины  $\eta_2$ :

$$\mu_k^{(2)} = E(\eta_2 - E\eta_2)^k = \int_0^1 (x - m_1^{(2)})^k f_2(x) dx = \int_0^1 (x - 1 + m_1^{(1)})^k f_1(1 - x) dx.$$

Далее проводя замену переменных  $y = 1 - x$ , окончательно получим

$$\mu_k^{(2)} = \int_1^0 (m_1^{(2)} - y)^k f_2(y) d(1 - y) = (-1)^k \int_0^1 (y - m_1^{(2)})^k f_2(y) dy = (-1)^k \mu_k^{(1)},$$

что и требовалось доказать. □

Таким образом, для построения разложения (3) на основе бета-распределения достаточно ограничиться рассмотрением случаев положительной асимметрии, поскольку верно

$$w_{\alpha,\beta}(x) = w_{\beta,\alpha}(1 - x),$$

что удовлетворяет условию утверждения 2.

Далее заметим, что для бета-распределения с фиксированными параметрами  $\alpha$  и  $\beta$  оценки плотностей по формуле (10) могут существенно отличаться от искомым плотностей с точки зрения критериев согласия, таких как критерий максимального правдоподобия, хи-квадрат Пирсона, критерий Колмогорова и др. Чтобы получить наилучшие результаты оценивания, естественно проводить оптимизацию по параметрам базовой функции плотности  $w_{\alpha,\beta}(x)$ . С учетом этого замечания авторами предлагается следующий МЕС-алгоритм (Measure Estimation on Canonical moments) оценивания плотности распределения случайной величины, распределенной на финитном интервале.

*Шаг 0.* Задать  $i = 0$ , выбрать начальные значения параметров базовой плотности  $(\alpha_0, \beta_0)$ .

*Шаг 1.* Вычислить начальные  $m_k$  и центральные  $\mu_k$  моменты случайной величины  $\xi$ .

*Шаг 2.* Вычислить  $\zeta$ -характеристики в соответствии с Q-D-алгоритмом.

*Шаг 3.* Вычислить коэффициенты ортогональных полиномов в (5).

*Шаг 4.* Вычислить оценку искомой функции плотности с помощью разложения (10).

*Шаг 5.* Оценить качество разложения (10) на основе выбранного критерия согласия. Если получен удовлетворительный результат, то аппроксимация (10) выбирается в качестве оценки искомой функции плотности. В противном случае производится выбор новых значений параметров  $\alpha_{i+1}, \beta_{i+1} = (\alpha_i + \Delta\alpha, \beta_i + \Delta\beta)$ , где  $(\Delta\alpha, \Delta\beta)$  определяются на основе выбранного метода оптимизации и заданной точности вычислений. Далее полагаем  $i := i + 1$  и переходим на шаг 4.

Отметим, что качество оценок МЕС-алгоритма определяется, с одной стороны, свойствами оценок  $\zeta$ -характеристик, получаемых на основе Q-D-алгоритма (шаг 2), а с другой — выбираемым критерием согласия. Как показано в [5, 14], оценки  $\zeta$ -характеристик Q-D-алгоритма обладают свойствами состоятельности и несмещенности. Если в качестве критерия согласия используется максимум функции правдоподобия или минимум статистики хи-квадрат, то согласно [3] получаемые оценки обладают свойствами асимптотической нормальности и эффективности.



### 3. Результаты моделирования

Исследование качества аппроксимации (10) проводилось посредством вычислительных экспериментов, в которых оценивались плотности распределения, соответствующие различным случайным величинам, распределенным на отрезке  $[0, 1]$ . Эти плотности оценивались при помощи МЕС-алгоритма, причем в качестве базовой плотности выбрана плотность бета-распределения.

Качество аппроксимации искомой функции плотности распределения определялось по критерию хи-квадрат Пирсона [3]. Сравнение получаемых результатов проводилось с точки зрения представления (8) при  $n = 2$ ,  $n = 3$  и  $n = 4$ , что соответствовало двум, трем и четырем слагаемым в разложении (10).

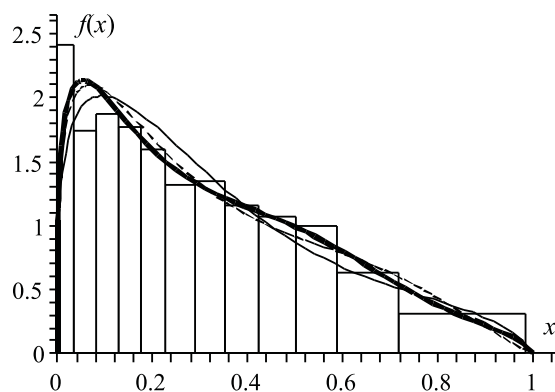
**Модель 1.** В качестве искомой плотности (1) выберем треугольное распределение с положительной асимметрией. На отрезке  $[0, 1]$  соответствующая плотность имеет вид

$$f_{\xi}(x) = 2 - 2x.$$

Смоделируем выборку из этого распределения объемом  $N = 1000$  и воспользуемся МЕС-алгоритмом для оценивания  $f_{\xi}(x)$ . Графики полученных оценок функции плотности  $\hat{f}_{\xi}^{(2)}(x)$ ,  $\hat{f}_{\xi}^{(3)}(x)$  и  $\hat{f}_{\xi}^{(4)}(x)$  представлены на рисунке (соответственно тонкой, штриховой и жирной линиями).

Из рисунка видно, что уже самая простая аппроксимация  $\hat{f}_{\xi}^{(2)}(x)$  достаточно хорошо описывает распределение исходных данных. В то же время график наглядно показывает, что увеличение числа слагаемых в разложении (10) для рассматриваемой модели приводит к повышению точности оценивания плотности распределения — оценки функции плотности  $\hat{f}_{\xi}^{(3)}(x)$  и  $\hat{f}_{\xi}^{(4)}(x)$  позволяют улучшить качество аппроксимации. Этот эффект скорее всего объясняется тем, что специфика конкретного распределения в разложении (10) учитывается за счет введения поправок к базовой функции, обеспечиваемых соответствующими ортогональными полиномами.

Кроме представленных результатов, заслуживает внимания исследование статистических свойств оценок, получаемых МЕС-алгоритмом. Для решения этого вопроса в условиях рассматриваемой модели было проведено ( $N_p = 100$ ) вычислительных экспериментов. В качестве характеристик, отражающих свойства МЕС-оценок, были выбраны среднее арифметическое, медиана, среднеквадратическое отклонение и коэффициент вариации. Результаты усреднения по  $N_p$  экспериментам представлены в табл. 1.



Результаты оценивания плотности МЕС-алгоритмом

Т а б л и ц а 1. Результаты оценивания плотности распределения при  $n = 2$ ,  $n = 3$  и  $n = 4$

Статистическая характеристика	Оценки		$\hat{\chi}_{(2)}^2$	$\hat{\chi}_{(3)}^2$	$\hat{\chi}_{(4)}^2$
	$\hat{\alpha}$	$\hat{\beta}$			
Среднее	0.85	0.34	19.91	11.46	8.93
Медиана	0.85	0.34	18.95	11.48	8.72
Стандартная ошибка	0.10	0.05	7.33	4.69	4.04
Коэффициент вариации	71.13	7.80	36.83	40.81	45.27

В табл. 1 величины  $\hat{\chi}_{(2)}^2$ ,  $\hat{\chi}_{(3)}^2$  и  $\hat{\chi}_{(4)}^2$  представляют собой значения статистик хи-квадрат Пирсона при оценивании искомой плотности функциями  $\hat{f}_{\xi}^{(2)}(x)$ ,  $\hat{f}_{\xi}^{(3)}(x)$  и  $\hat{f}_{\xi}^{(4)}(x)$  соответственно.

Следует отметить, что близость значений среднего и медианы оценок параметров  $\alpha$  и  $\beta$  свидетельствует о симметричности соответствующих (эмпирических) распределений этих оценок. Кроме того, стандартная ошибка оценки параметра  $\alpha$  практически в два раза больше, чем для  $\beta$ , что объясняется положительной асимметрией рассматриваемого распределения. Для распределения с отрицательной асимметрией картина обратная.

**Модель 2.** В качестве искомой плотности  $f_{\xi}(x)$  рассмотрим различные варианты бета-распределения (табл. 2). Для каждого из этих распределений смоделирована выборка объемом  $N = 1000$  и произведено оценивание функции плотности  $f_{\xi}(x)$  на основе метода максимального правдоподобия (*ML*-оценка) и МЕС-алгоритма. Полученные оценки и соответствующие им значения статистик хи-квадрат Пирсона также представлены в табл. 2.

Анализ данных табл. 2 позволяют сделать следующие выводы. Значения критерия хи-квадрат Пирсона для оценок, получаемых МЕС-алгоритмом, при симметричных распределениях типа бета оказываются существенно выше, чем соответствующие значения для оценок метода максимального правдоподобия. При этом с увеличением значений параметров точность оценивания по обоим методам выравнивается. С другой стороны, при отклонении распределений от симметричных явно просматривается улучшение качества оценок, получаемых МЕС-алгоритмом, причем, чем большей асимметрией обладает моделируемое распределение, тем ощутимее становится преимущество МЕС-алгоритма.

Т а б л и ц а 2. Результаты оценивания плотности случайных величин с различными законами распределения

Искомое распределение $f_{\xi}(x)$	Оценки параметров бета-распределения				$\hat{\chi}_{ML}^2$	$\hat{\chi}_{(2)}^2$	$\hat{\chi}_{(3)}^2$	$\hat{\chi}_{(4)}^2$
	$\hat{\alpha}_{ML}$	$\hat{\beta}_{ML}$	$\hat{\alpha}$	$\hat{\beta}$				
Арсинус	-0.48	-0.50	-0.04	-0.62	3.86	69.21	33.80	17.27
Ray[0,1]	-0.01	-0.07	-0.49	-0.17	2.93	42.11	10.47	7.95
Бета(1,1)	0.97	0.97	1.65	0.97	3.68	8.45	5.04	6.04
Бета(1,3)	0.87	2.80	1.07	2.69	12.44	10.42	13.53	12.19
Бета(1,6)	1.19	6.40	1.22	6.36	2.45	2.36	2.36	2.39
Бета(6,6)	6.04	6.03	6.07	5.98	11.77	11.54	12.10	12.05
Бета(1,10)	1.00	10.43	0.98	10.13	3.54	3.11	3.11	3.12
Бета(10,10)	10.70	10.75	10.88	10.95	3.78	3.65	3.65	3.65

Анализ значений оценок параметров, получаемых МЕС-алгоритмом, позволяет говорить об их существенном отличии как от значений параметров моделируемых случайных величин, так и от оценок максимального правдоподобия. Этот факт не должен вводить в заблуждение исследователя, поскольку оценки МЕС-алгоритма на самом деле являются оценками базовой функции плотности, а не искомой, которая в общем случае, естественно, может и не принадлежать семейству бета-распределений. Ситуация такого типа рассмотрена ниже в модели 3.

**Модель 3.** В качестве искомой плотности  $f_\xi(x)$  рассмотрим смесь двух плотностей бета-распределения вида

$$f_\xi(x) = \lambda w_{\alpha_1, \beta_1}(x) + (1 - \lambda) w_{\alpha_2, \beta_2}(x),$$

где  $\lambda \in [0, 1]$  — параметр смеси. В качестве значений параметров смешиваемых плотностей  $(\alpha_1, \beta_1)$  и  $(\alpha_2, \beta_2)$  выбирались такие, которые обеспечивали существенно различные свойства функций  $w_{\alpha_1, \beta_1}(x)$  и  $w_{\alpha_2, \beta_2}(x)$  соответственно. Такой выбор обусловливается желанием исследовать качество оценок, получаемых с помощью МЕС-алгоритма, для плотностей, отличных от плотности бета-распределения как по форме, так и по свойствам. Оценки, рассчитанные на основе МЕС-алгоритма и метода максимального правдоподобия, для выборки объемом  $N = 1000$  и соответствующие им значения статистик хи-квадрат Пирсона представлены в табл. 3 и 4.

Данные табл. 3 и 4 позволяют предположить, что применение метода максимального правдоподобия в большинстве случаев оценивания смесей бета-распределений дает менее точные оценки по критерию хи-квадрат. С другой стороны, в дополнение к результатам модели 2 можно сказать, что при оценивании смешиваемых в разных

Т а б л и ц а 3. Результаты оценивания плотности случайных величин с различными законами распределения МЕС-алгоритмом

Параметр смеси $\lambda$	$(\alpha_1, \beta_1)$	$(\alpha_2, \beta_2)$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\chi}_{(2)}^2$	$\hat{\chi}_{(3)}^2$	$\hat{\chi}_{(4)}^2$
1/3	(1, 12)	(1, 1)	0.32	0.41	39.11	4.65	5.86
2/3	(1, 12)	(1, 1)	-0.06	1.17	48.34	56.10	67.05
1/3	(1, 12)	(3, 3)	-0.58	0.86	70.24	38.09	30.96
2/3	(1, 12)	(3, 3)	0.08	1.48	40.24	38.58	45.81
3/4	(2, 3)	(0, 0)	1.33	1.33	28.30	47.32	32.71
7/8	(2, 3)	(0, 0)	1.85	1.95	30.35	46.47	35.53

Т а б л и ц а 4. Результаты оценивания плотности случайных величин с различными законами распределения методом максимального правдоподобия

Параметр смеси $\lambda$	$(\alpha_1, \beta_1)$	$(\alpha_2, \beta_2)$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\chi}_{ML}^2$
1/3	(1, 12)	(1, 1)	-0.06	0.61	9.89
2/3	(1, 12)	(1, 1)	-0.28	1.20	76.43
1/3	(1, 12)	(3, 3)	0.29	1.16	30.40
2/3	(1, 12)	(3, 3)	-0.13	1.52	52.32
3/4	(2, 3)	(0, 0)	0.86	1.44	33.28
7/8	(2, 3)	(0, 0)	1.46	2.1	40.18

пропорциях симметричных и несимметричных распределений точность, достигаемая МЕС-алгоритмом, оказывается выше даже при наличии относительно небольшой доли одного из распределений смеси. Этот факт позволяет предположить, что при решении задач оценивания зашумленных распределений (с разным уровнем шума) применение МЕС-алгоритма предпочтительнее общепотребительных оценок максимального правдоподобия.

## Заключение

В работе предложен алгоритм оценивания функции плотности случайной величины, распределенной на конечном интервале, на основе теории канонических моментов. Отмечено, что получаемые оценки параметров функции плотности обладают рядом оптимальных статистических свойств. Проведенные вычислительные эксперименты показали, что полученное ортогональное разложение функции плотности в ряд является достаточно гибким инструментом, учитывающим и вид выбранной базовой функции плотности, и структуру заданной выборки. Этот факт дает возможность рекомендовать данный подход не только для решения задач идентификации распределений случайных величин, но и для построения на их основе алгоритмов оценивания параметров регрессионных зависимостей.

## Список литературы

- [1] КЕНДАЛЛ М., СТЬЮАРТ А. Теория распределений. М.: Наука, 1966. 587 с.
- [2] КРАМЕР Г. Математические методы статистики. М.: Мир, 1975. 648 с.
- [3] ФИШЕР Р.А. Статистические методы для исследователей. М.: Госстатиздат, 1958. 268 с.
- [4] ROUSSEUW P.J. Tutorial to robust statistics // J. of Chemometrics. 1991. Vol. 5. P. 1–20.
- [5] DETTE H., STUDDEN W.J. Theory of Canonical Moments and its Applications in Statistics, Probability and Analysis. N.Y.: John Wiley & Sons Inc., 1997. 330 p.
- [6] ROUSSEUW P.J., LEROY A.M. A robust scale estimator based on the shortest half // Statistica Neerlandica. 1988. Vol. 42, N 2. P. 103–116.
- [7] WELLING M. Robust higher order statistics // Proc. of the Tenth Intern. Workshop on Artificial Intelligence and Statistics (AISTATS 2005), Barbados, 2005. P. 405–412.
- [8] SKIBINSKI M. Extreme n-th moments for distributions on [0,1] and the inverse of a moment space map // J. App. Probab. 1968. Vol. 5. P. 693–701.
- [9] CHARLIER C.V.L. Researches into the Theory of Probability. Lund: Hakon Ohlsson, 1906.
- [10] ШОНАТ J.A., ТАМАРКИН J.D. The problems of moments. Providence: American Mathematical Society, 1963.
- [11] ПУГАЧЕВ В.С. Теория вероятностей и математическая статистика. М.: Наука, 1979. 496 с.
- [12] SZEGO G. Orthogonal Polynomials. N.Y.: American Mathematical Society, 1959. 500 p.
- [13] СУЕТИН П.К. Классические ортогональные многочлены. М.: Физматгиз, 1976. 328 с.
- [14] ГРИГОРЬЕВ Ю.Д., ЩЕКОЛДИН В.Ю. Канонические моменты вероятностных мер // Сб. научных трудов НГТУ. 2000. № 3(27). С. 17–20.
- [15] ИОХВИДОВ И.С. Ганкелевы и теплицевы матрицы и формы. М.: Наука, 1974. 264 с.

*Поступила в редакцию 8 октября 2008 г.,  
в переработанном виде — 14 мая 2009 г.*