УДК 621.3911:519.28

И. К. ВАСИЛЬЕВА

Национальный аэрокосмический университет им. Н. Е. Жуковского "ХАИ", Украина

ВЕРОЯТНОСТНАЯ МОДЕЛЬ МНОГОКАНАЛЬНЫХ НЕГАУССОВЫХ ДАННЫХ НА БАЗЕ МНОГОМЕРНОГО ВАРИАНТА УСЕЧЕННОГО НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Предложена многомерная вероятностная модель для описания совместного распределения негауссовых данных на базе многомерного варианта усеченного нормального закона распределения. Описана методика оценки параметров модели и формирования уравнений границ, задающих подпространство усечения, при наличии корреляционных взаимосвязей компонент случайного вектора наблюдений. Приведены результаты аппроксимации с использованием предлагаемой модели двумерных случайных величин, маргинальные распределения которых существенно отличаются от нормального вида. Показано, что данная модель может быть использована для аналитического описания совместного распределения коррелированных случайных данных, в т.ч., если отдельные компоненты вектора наблюдений характеризуются различными законами распределения.

Ключевые слова: *аппроксимация, вероятностная модель, многомерный закон распределения, усеченное нормальное распределение, параметры распределения, преобразование корреляции, целевая функция.*

Введение

В системах автоматического обнаружения объектов, распознавания образов и диагностики состояний широко распространены статистические правила принятия решений, основанные на формировании отношения правдоподобия $L_{uv}(\vec{x}^*)$ и сравнении его (или его логарифма) с некоторым порогом в соответствии с принятым критерием качества [1, 2]:

$$L_{uv}(\vec{x}^*) = \frac{f_p(x_1^*, \ldots, x_p^* | a_u)}{f_p(x_1^*, \ldots, x_p^* | a_v)} \geq c,$$

где $\vec{x}^*$ – p-компонентный вектор наблюдений;

$f_p(\vec{x}^* | a_k)$ – значение условной p-мерной плотности распределения вероятности (ПРВ) выборки $\vec{x}^*$ при условии ее принадлежности к классу a_k.

Априорные сведения об исследуемых генеральных совокупностях относятся обычно лишь к виду или некоторым общим свойствам закона распределения (ЗР) исследуемого случайного вектора; их получают в результате предварительных исследований или из теоретических соображений о природе исследуемого объекта. В большинстве случаев для формирования описаний объектов используется статистическая, выборочная информация, т.е. в отношении правдоподобия подставляют статистические оценки ПРВ, получаемые в процессе обучения. При этом объем и достоверность информации об объектах наблюдения, а также эффективность использования этой информации в решающих правилах во многом определяют и достоверность принимаемых решений. Одним из способов снижения вероятностей ошибок является увеличение количества классификационных признаков. При этом выводы, получаемые в результате анализа и классификации множества статистически обследованных (по ряду свойств) объектов, должны опираться на совокупность этих взаимосвязанных свойств с обязательным учетом структуры и характера связей. Трудности при построении адекватных многомерных статистических моделей обусловлены тем, что методология обработки данных при наличии корреляционных связей основывается на предположении о нормальности рассматриваемых распределений. В то же время исследуемые признаки объектов обычно имеют различную физическую природу, как, например, амплитуда и фаза сигнала, следовательно, формы распределений их вероятных значений могут существенно отличаться друг от друга и, в частности, от нормального ЗР. Так, большинство радиолокационных характеристик объектов дистанционного зондирования [3 – 6] имеют негауссовы распределения в ограниченной области допустимых значений.

В работе [7] приведена методика получения оптимальных оценок параметров усеченных нормальных ПРВ для одномерных маргинальных распределений компонент вектора наблюдений.

Целью работы является разработка теоретико-вероятностной модели вида многомерного варианта усеченного нормального ЗР с заданными вектором математических ожиданий (МО) $\vec{m}$, корреляционной матрицей (КМ) $\mathbf{R}$ и параметрами усечения, а также проверка адекватности разработанной модели по результатам аппроксимации негауссовых распределений компонент случайного вектора.

1. Усеченное нормальное распределение совокупности негауссовых данных

Распределения реальных многоканальных данных зачастую имеют негауссов вид. Даже при допущении о нормализации ЗР наблюдений в силу центральной предельной теоремы нельзя исключить нелинейность, вносимую при приеме информации, что проявляется в изменении вида и параметров наблюдаемого распределения. Тем не менее, в практических задачах обработки многоканальной информации обычно используют модели многомерных нормальных совокупностей вида [1, 8]:

$$f(\vec{x})=(2\pi)^{-p/2}|\mathbf{R}|^{-1/2}\exp\left[-\frac{1}{2}(\vec{x}-\vec{m})^T\mathbf{R}^{-1}(\vec{x}-\vec{m})\right], \quad (1)$$

где $\vec{m}$ – вектор МО данных наблюдений в p-мерном пространстве **X**;

R – корреляционная матрица, элементы которой $\{r_{ij}\sigma_i\sigma_j\}$ – центральные моменты второго порядка, образованные компонентами случайного вектора $\vec{x}$, где σ_i – среднеквадратическое отклонение (СКО);

r_{ij} – коэффициент корреляции между i-й и j-й компонентами случайного вектора.

С одной стороны, такой подход упрощает оценку параметров ЗР и дает возможность учесть в модели наличие взаимных корреляционных связей между данными в различных каналах, но с другой – ухудшает качество аппроксимации эмпирических распределений, соответственно, при этом теряется информация о свойствах объекта наблюдения.

Если вид маргинальных ПРВ с допустимой степенью приближения нельзя признать нормальным, для аппроксимации распределений многомерных данных применяют многопараметрические ЗР, основанные на нелинейных преобразованиях нормальных случайных величин [4].

Как альтернатива применению нелинейных преобразований в работе [7] предложена модель на базе усеченного нормального распределения. Маргинальные одномерные ПРВ при этом имеют вид

$$f(x_j)=N'(x_j)=\begin{cases}N(x_j;m_j,\sigma_j)C_j, & x_j\in[a_j,b_j];\\ 0, & x_j\notin[a_j,b_j],\end{cases} \quad (2)$$

где $N(x_j;m_j,\sigma_j)$ – ПРВ нормального закона с параметрами m_j, σ_j;

C_j – константа, определяемая из условия нормировки ПРВ на интервале усечения [a, b]:

$$C_j=C(m_j,\sigma_j,a,b)=1\Big/\int_a^b N(x_j;m_j,\sigma_j)dx_j. \quad (3)$$

Значения МО и СКО усеченного нормального ЗР должны обеспечивать наилучшее совпадение кривых эмпирического распределения и аппроксимирующей модели (2); для получения оценок этих параметров используется оптимизационный подход, а целевая функция формируется по методу наименьших квадратов и с учетом ограничений на допустимые значения параметров ($\sigma_j>0$).

Модель (2) можно обобщить на многомерный случай, при этом p-мерный усеченный нормальный ЗР $N'_p(\vec{x})$ будет полностью описан следующими параметрами:

– вектором $\vec{m}$, составленным из МО усеченных нормальных ПРВ компонент $\{x_j\}$, j = 1…p;

– выборочной корреляционной матрицей **R**;

– векторами $\vec{a}=\{a_j\}$ и $\vec{b}=\{b_j\}$, определяющими координаты границ интервалов усечения для каждой компоненты $\{x_j\}$, j = 1…p.

Таким образом, многомерную модель на базе усеченных нормальных ЗР можно записать в виде

$$N'_p(\vec{x})=\begin{cases}N_p(\vec{x};\vec{m},\mathbf{R})C, & \vec{x}\in D;\\ 0, & \vec{x}\notin D,\end{cases} \quad (4)$$

где $N_p(\vec{x};\vec{m},\mathbf{R})$ – многомерное нормальное распределение с параметрами $\vec{m}$ и **R**;

C – обобщенная константа нормировки:

$$C=\prod_{j=1}^{p}C_j; \quad (5)$$

D – (p + 1)-мерное подпространство усечения, границами которого являются гиперплоскости, определяемые системами уравнений:

$$\begin{cases}x_j=a_j;\\ x_i=0, i\neq j;\end{cases}\quad \begin{cases}x_j=b_j;\\ x_i=0, i\neq j.\end{cases}$$

Влияние корреляционных связей проявляется в преобразовании системы координат: повороте и масштабировании координатных осей; при этом углы поворота определяются матрицей собственных векторов **Φ** КМ, а коэффициенты масштаба находят как $\sqrt{\lambda_j}$, где $\{\lambda_j\}$ – собственные значения КМ **R**.

Т.о., чтобы задать подпространство усечения D в новой системе координат необходимо:

– привести $\vec{a}$ и $\vec{b}$ к стандартному виду:

$$\vec{a}^{(0)}=\{(a_j-m_j)/\sigma_j\};\ \vec{b}^{(0)}=\{(b_j-m_j)/\sigma_j\};$$

– получить матрицу линейного преобразования корреляции [8]:

$$\mathbf{A}=\mathbf{\Phi}\cdot\mathrm{diag}\left(\sqrt{\lambda_j}\right);$$

– определить координаты границ областей усечения компонент в преобразованной системе:

$$\vec{g}=\mathbf{A}\cdot\vec{a}^{(0)}+\vec{m};\ \vec{q}=\mathbf{A}\cdot\vec{b}^{(0)}+\vec{m};$$

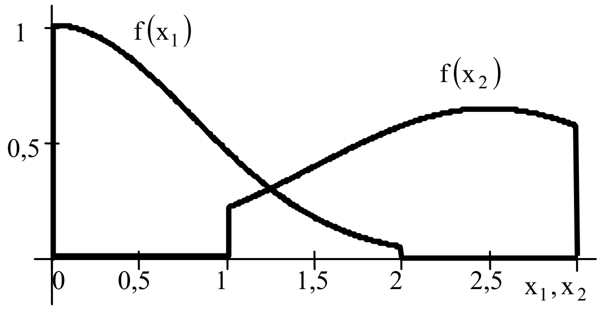
– сформировать уравнения границ D:

$$\{\psi_j(\vec{x})\}=\mathbf{\Phi}^T(\vec{x}-\vec{g})=0;\ \{\varphi_j(\vec{x})\}=\mathbf{\Phi}^T(\vec{x}-\vec{q})=0.$$

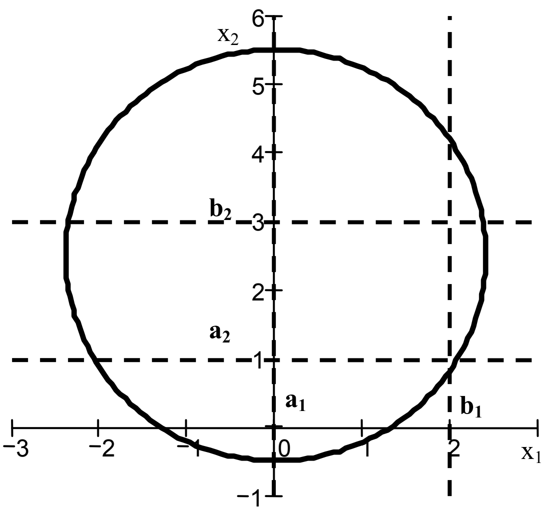
В результате, многомерная область усечения модели (4) определяется как непустое множество, удовлетворяющее системе неравенств-ограничений:

$$D=\left\{\vec{x}\in R^p\,\middle|\,\psi_j(\vec{x})\geq 0;\varphi_j(\vec{x})\leq 0;\ j=1\ldots p\right\}.$$

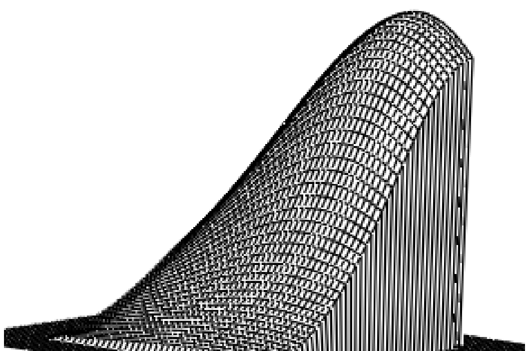
Для примера рассмотрены двумерные модели (4). ПРВ компонент с параметрами $m_1 = 0$, $\sigma_1 = 0{,}8$, $a_1 = 0$, $b_1 = 2$ и $m_2 = 2{,}5$, $\sigma_2 = 1$, $a_2 = 1$, $b_2 = 3$ показаны на рис. 1, а; корреляционные эллипсы, границы области D и совместные распределения (4) для различных значений коэффициента r_{12} – на рис. 1(б, в), 3, 4. Рис. 2 иллюстрирует трансформацию области усечения, обусловленную корреляционными связями компонент случайного вектора.



а



б



в

Рис. 1. Случайный вектор, компоненты которого описываются усеченными нормальными ПРВ: а – маргинальные ПРВ; б – корреляционный эллипс и границы области усечения при $r_{12} = 0$; в – вид результирующего двумерного ЗР $N'_2(x_1, x_2; m_1, m_2, \sigma_1, \sigma_2, r_{12}, a_1, a_2, b_1, b_2)$

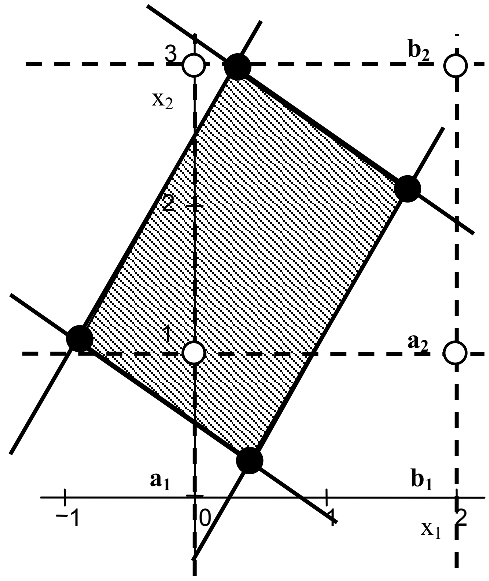
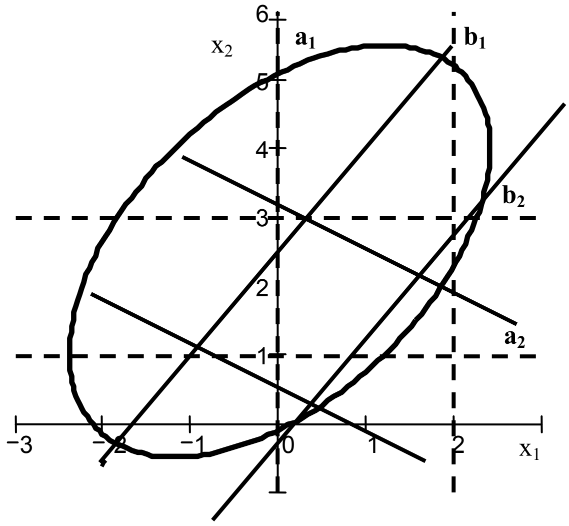
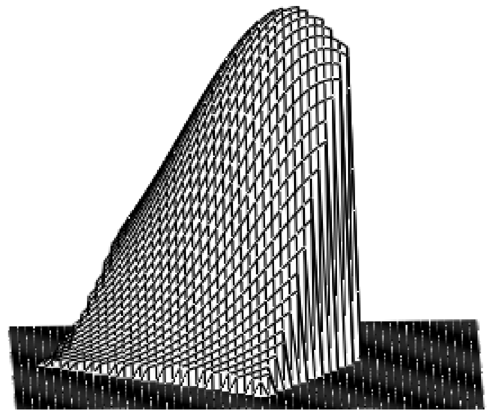


Рис. 2. Трансформация исходной области усечения в результате преобразования корреляции, $r_{12} = 0{,}5$

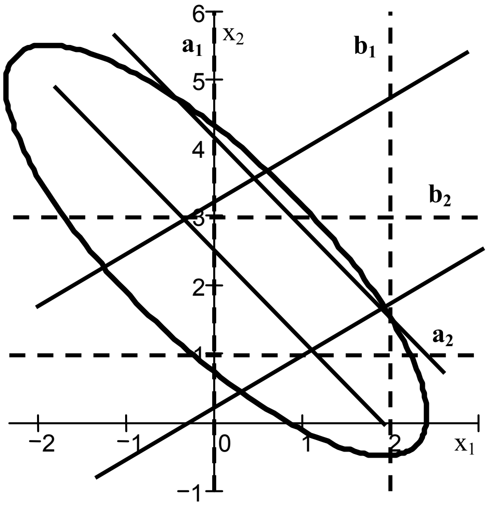


а

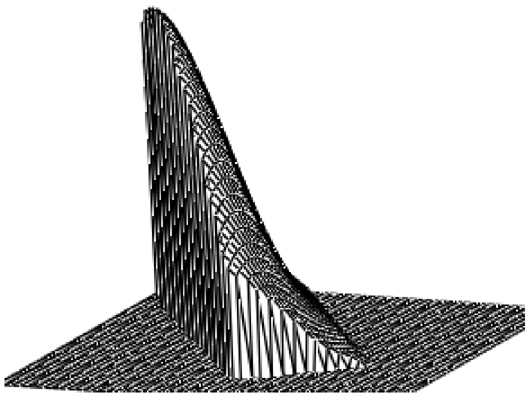


б

Рис. 3. Влияние корреляционной зависимости между компонентами случайного вектора (x_1, x_2) на вид совместного усеченного распределения (маргинальные ПРВ такие же, как на рис. 1, а): а – корреляционный эллипс и границы области усечения при $r_{12} = 0{,}5$; б – вид двумерного ЗР



а

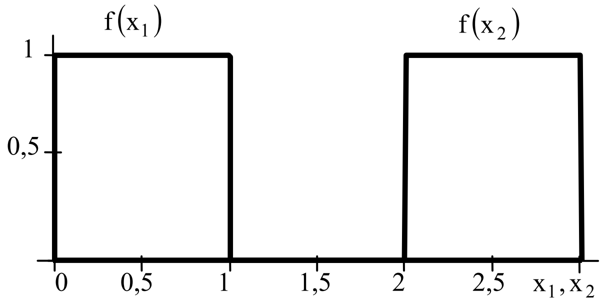


б

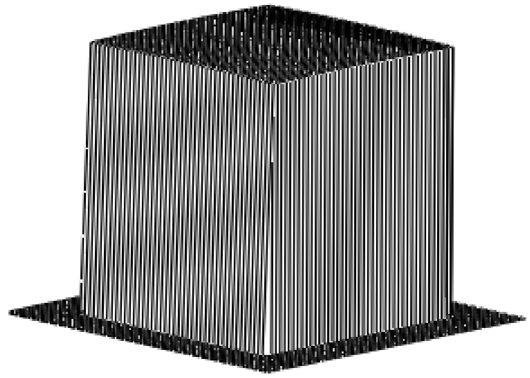
Рис. 4. Совместное усеченное распределение при наличии отрицательной корреляционной связи между компонентами случайного вектора (x_1, x_2) (маргинальные ПРВ такие же, как на рис. 1, а): а – корреляционный эллипс и границы области усечения при $r_{12} = -0{,}8$; б – вид двумерного ЗР

Модели (2), (4) позволяют путем выбора оптимальных значений параметров m_j и σ_j изменять форму кривой нормального ЗР с нулевыми асимметрией β_1 и эксцессом β_2 [8, 9], что дает возможность с достаточной степенью приближения описывать как асимметричные ЗР, так и распределения, характеризующиеся отличной от нормального закона крутизной подъема кривой ПРВ.

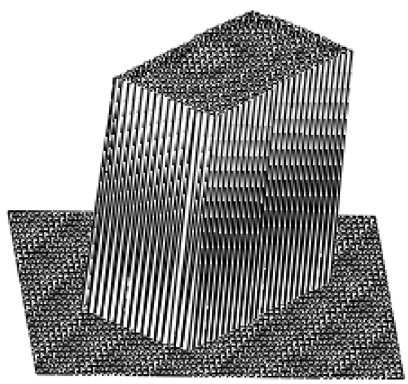
Так, для равномерного ЗР $\beta_1 = 0$, $\beta_2 = -1{,}2$. Вид маргинальных ПРВ (2), аппроксимирующих равновероятные распределения компонент $f(x_1)$ и $f(x_2)$, показан на рис. 5, а. Параметры модели (2), соответственно, – $m_1 = 0{,}5$, $\sigma_1 = 4$, $a_1 = 0$, $b_1 = 1$ и $m_2 = 2{,}5$, $\sigma_2 = 4{,}1$, $a_2 = 2$, $b_2 = 3$. Двумерные модели (4) для случаев $r_{12} = 0$ (отсутствие корреляции), $r_{12} > 0$ и $r_{12} < 0$ изображены на рис. 5, б – г.



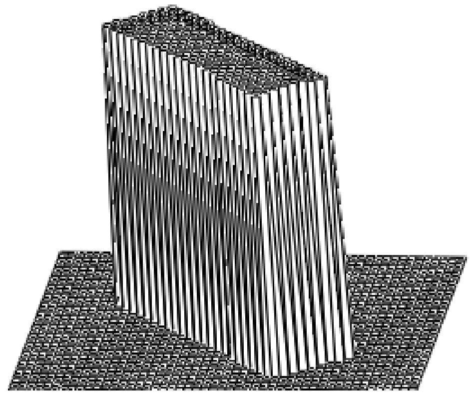
а



б



в



г

Рис. 5. Аппроксимация двумерного равновероятного распределения: а – маргинальные ПРВ, представляющие собой усеченные нормальные ПРВ; б – $r_{12} = 0$; в – $r_{12} = 0{,}5$; г – $r_{12} = -0{,}8$

Для экспоненциального ЗР $\beta_1 = 2$, $\beta_2 = 6$, т.е. мода распределения сдвинута влево, а кривая ПРВ имеет более высокую и острую вершину по сравнению с нормальной. На рис. 6, а показаны маргинальные ПРВ (2), аппроксимирующие экспоненциальные распределения компонент $f(x_1)$ и $f(x_2)$ случайного вектора. Значения параметров модели (2) при этом составили: $m_1 = -1{,}5$, $\sigma_1 = 1$, $a_1 = 0$, $b_1 = 2$ и $m_2 = -2$, $\sigma_2 = 1{,}5$, $a_2 = 0$, $b_2 = 2$.

На рис. 6, б изображено совместное распределение некоррелированных компонент x_1 и x_2, а на рис. 6, в – вид двумерного распределения с учетом корреляционных связей между компонентами.

Достоинством модели (4) является возможность адекватного формального описания совместного распределения коррелированных компонент с различными законами распределения.

В качестве примера предложен двумерный случай; когда одна составляющая вектора (x_1) имеет экспоненциальное, а вторая (x_2) – нормальное распределение (рис. 7, а). Параметры модели (2), аппроксимирующей маргинальные ПРВ, имели следующие значения: $m_1 = -1{,}5$, $\sigma_1 = 1$, $a_1 = 0$, $b_1 = 2$ и $m_2 = 1{,}5$, $\sigma_2 = 4$, $a_2 = 1$, $b_2 = 2$. Двумерная «разнокомпонентная» ПРВ для двух значений коэффициента корреляции r_{12} показана на рис. 7, б, в.

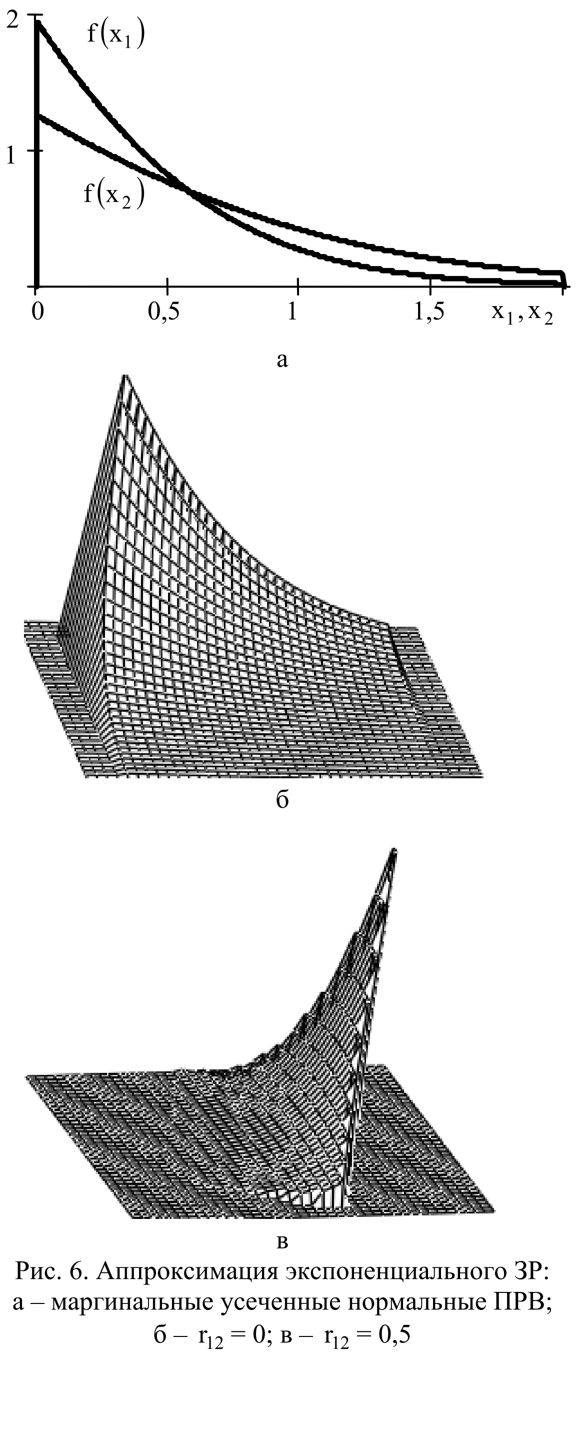


Рис. 6. Аппроксимация экспоненциального ЗР: а – маргинальные усеченные нормальные ПРВ; б – $r_{12} = 0$; в – $r_{12} = 0{,}5$

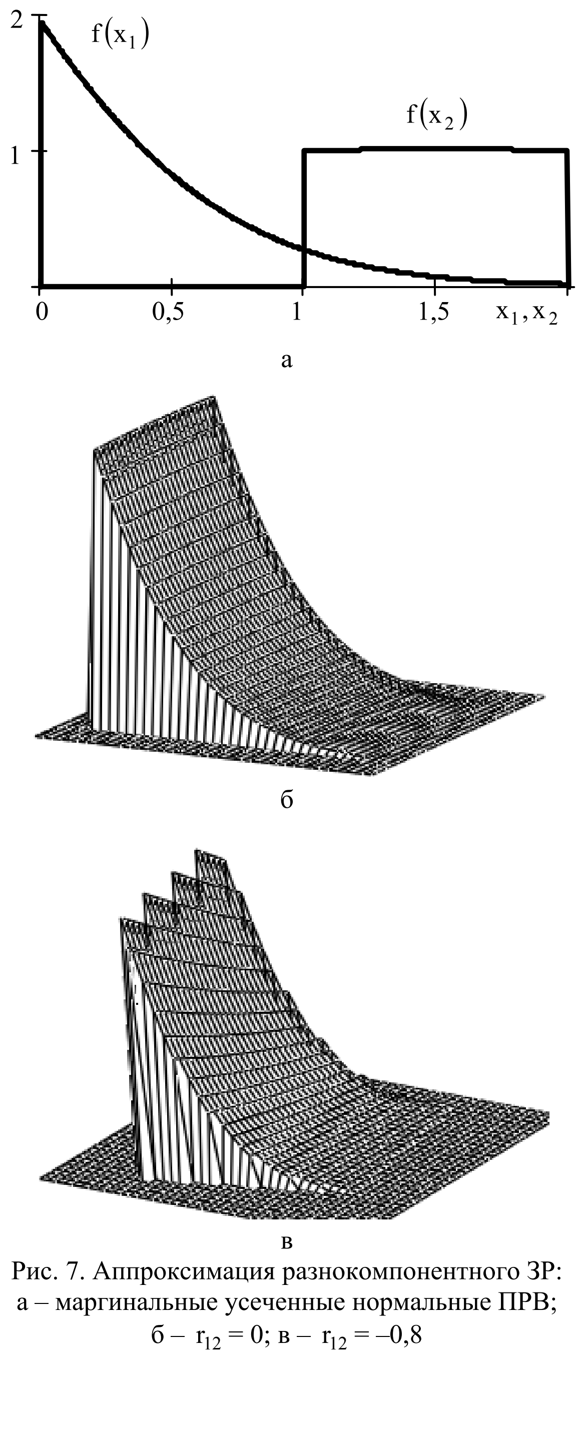


Рис. 7. Аппроксимация разнокомпонентного ЗР: а – маргинальные усеченные нормальные ПРВ; б – $r_{12} = 0$; в – $r_{12} = -0{,}8$

2. Проверка адекватности модели

Численное интегрирование усеченных двумерных нормальных ПРВ (4) подтвердило выполнение условия их нормировки на области усечения D:

$$\iint_D N_2'(x_1, x_2)dx_1dx_2 = 1 .$$

Форма корреляционных эллипсов (см. рис. 3, 4) соответствует заданным корреляционным связям между компонентами случайных векторов. Для проверки адекватности описания одномерных маргинальных распределений $f(x_j)$ усеченными нормальными ПРВ был проведен расчет значений критерия J_e – среднеквадратичной ошибки аппроксимации ПРВ $f(x_j)$ в точках x_j, $j = 0 \ldots M$ ($M = 15$):

$$J_e = \sum_{j=0}^{M}\left[N'(x_j) - f(x_j)\right]^2 .$$

В качестве аппроксимируемых негауссовых распределений были выбраны экспоненциальный ($\lambda = 1$) и равномерный ($a = 0$, $b = 1$) ЗР. Для сравнения приведены результаты аппроксимации этих ЗР нормальными ПРВ $N(x)$ с соответствующими значениями параметров m и σ:

– экспоненциальный ЗР

$$m = \sigma = 1/\lambda ;$$

– равновероятный ЗР

$$m = (a+b)/2 , \ \sigma = (b-a)/\sqrt{12} .$$

Оценки параметров моделей вида усеченных нормальных ПРВ $N'(x)$ были найдены как решение задачи минимизации функции критерия J_e с учетом ограничений на значения параметров ($\sigma > 0$) [10].

Вид моделей $N(x)$ и $N'(x)$ показан на рис. 8.

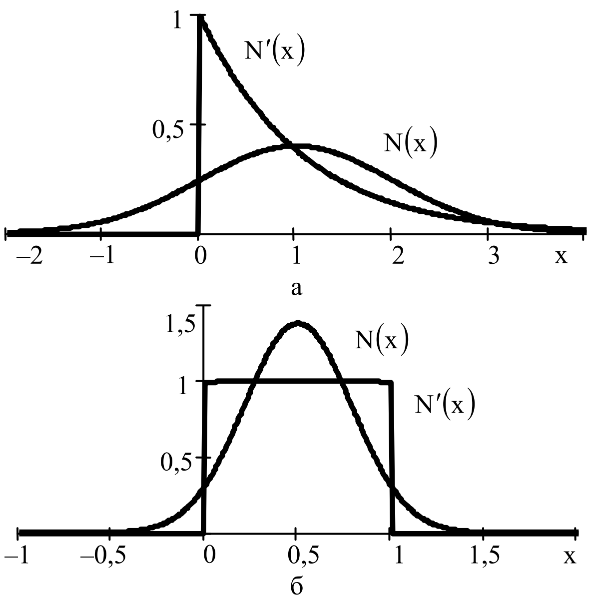


Рис. 8. Аппроксимация негауссовых распределений нормальными и усеченными нормальными ПРВ:
а – экспоненциальный ЗР; б – равномерный ЗР

В табл. 1 приведены оценки параметров модели $N'(x)$ (запись $N'(x|C)$ означает наличие дополнительного ограничения – $C < 1000$), параметры нормальных ПРВ $N(x)$, а также значения критерия J_e.

Полученные результаты дают основание считать модель $N_p'(x)$ приемлемой для описания совместных распределений многоканальных данных.

Таблица 1

Результаты проверки адекватности моделей

Модель	a	b	m	σ	C	J_e
Экспоненциальный ЗР, $\lambda = 1$						
$N'(x)$	0	4	–25,04	5,15	$1{,}75\cdot10^{6}$	$7{,}35\cdot10^{-4}$
$N'(x\|C)$	0	4	–10,15	3,37	785,75	$1{,}77\cdot10^{-3}$
$N(x)$	–	–	1	1	–	0,904
Равномерный ЗР, $a = 0$, $b = 1$						
$N'(x)$	0	1	0,5	4	10,05	$1{,}12\cdot10^{-4}$
$N(x)$	–	–	0,5	0,289	–	2,45

Заключение

Распространенная на практике модель многомерного нормального распределения, несмотря на очевидные достоинства, не является адекватной, если реальное распределение описываемых данных хотя бы в одном из каналов наблюдения имеет негауссов вид. Усеченные нормальные ЗР более точно аппроксимируют распределения, существенно отклоняющиеся от нормального вида как по характеристикам асимметрии, так и эксцесса. Большая степень универсальности модели на базе усеченных нормальных ПРВ достигается за счет большего количества параметров. В многомерном случае параметрами предлагаемой модели являются: вектор математических ожиданий компонент $\vec{m}$, корреляционная матрица **R** = $\{r_{ij}\sigma_i\sigma_j\}_{p\times p}$, векторы координат границ интервалов усечения $\vec{g}$ и $\vec{q}$, константа нормировки C. Оценки параметров $\{m_j\}$, $\{\sigma_j\}$ находят путем оптимизации целевой функции (суммы квадратов отклонений аппроксимируемого распределения и модели) с ограничением на допустимую область поиска. Константу C рассчитывают методом численного интегрирования. Векторы $\vec{g}$ и $\vec{q}$, а также уравнения, задающие область усечения D, определяют с использованием линейного матричного преобразования корреляции.

Т.о., усеченный вариант многомерного нормального ЗР позволяет учесть в модели наличие взаимных корреляционных связей между данными в различных каналах; точность аппроксимации маргинальных одномерных распределений при этом на 3 – 4 порядка выше, чем в модели без усечения.

Литература

1. *Фомин, Я. А. Статистическая теория распознавания образов [Текст] / Я. А. Фомин, Г. Р. Тарловский. – М. : Радио и связь, 1986. – 264 с.*
2. *Комлевая, Н. О. Сравнительный анализ двух подходов при решении задачи классификации [Текст] / Н. О. Комлевая, А. Н. Комлевой, Б. И. Тимченко // Радіоелектронні і комп'ютерні системи. – 2014. – № 6 (70). – С. 115 – 119.*
3. *Попов, А. В. О поляризационной модуляции радиолокационных сигналов, отраженных морской поверхностью [Текст] / А. В. Попов, М. В. Борцова // Системи управління, навігації та зв'язку. – 2011. – № 1 (17). – С. 46 – 55.*
4. *Попов, А. В. Критерий различимости объектов дистанционного зондирования при негауссовском распределении признаков [Текст] / А. В. Попов // Радіоелектронні і комп'ютерні системи. – 2011. – № 3 (51). – С. 30 – 36.*
5. *Lee, J.-S. Polarimetric Radar Imaging: From Basics to Applications [Text] / J.-S. Lee, E. Pottier. – CRC press., 2009. – 398 p.*
6. *Алгоритм обнаружения сигнала в неопределенной помеховой обстановке для многоканальных систем [Текст] / А. Д. Абрамов, А. М. Ветошко, А. В. Одокиенко, К. А. Щербина // Радіоелектронні і комп'ютерні системи. – 2013. – № 3 (62). – С. 5 – 11.*
7. *Васильева, И. К. Способ описания негауссовых данных вероятностными моделями на базе усеченных нормальных распределений [Текст] / И. К. Васильева // Радіоелектронні і комп'ютерні системи. – 2011. – № 4 (52). – С. 35 – 40.*
8. *Бендат, Дж. Прикладной анализ случайных данных [Текст] : пер. с англ. / Дж. Бендат, А. Пирсол. – М. : Мир, 1989. – 540 с.*
9. *Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход [Текст] : монография / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. – Новосибирск : Изд-во НГТУ, 2011. – 888 с.*
10. *Хан, Г. Статистические модели в инженерных задачах [Текст] : пер. с англ. / Г. Хан, С. Шапиро. – М. : Мир, 1969. – 369 с.*

Поступила в редакцию 6.10.2014, рассмотрена на редколлегии 18.11.2014

Рецензент: д-р техн. наук, проф., с.н.с отдела Дистанционного зондирования Земли Р. Э. Пащенко, Институт радиофизики и электроники им. А. Я. Усикова ИРЭ НАН Украины, Харьков.

ІМОВІРНІСНА МОДЕЛЬ БАГАТОКАНАЛЬНИХ НЕГАУСОВИХ ДАНИХ НА БАЗІ БАГАТОВИМІРНОГО ВАРІАНТУ ЗРІЗАНОГО НОРМАЛЬНОГО РОЗПОДІЛУ

І. К. Васильєва

Запропоновано багатовимірну імовірнісну модель для опису сумісного розподілу негаусових даних на базі багатовимірного варіанту зрізаного нормального закону розподілу. Описано методику оцінювання параметрів моделі та формування рівнянь границь, що задають підпростір зрізання за наявності кореляційних взаємозв'язків компонент випадкового вектора спостережень. Наведено результати апроксимації із використанням запропонованої моделі двовимірних випадкових величин, маргінальні розподіли яких суттєво відрізняються від нормального виду. Показано, що надана модель може використовуватись для аналітичного опису сумісного розподілу корельованих випадкових даних, у т.ч., якщо окремі компоненти вектора спостережень характеризуються різними законами розподілу.

Ключові слова: апроксимація, імовірнісна модель, багатовимірний закон розподілу, зрізаний нормальний розподіл, параметри розподілу, кореляційне перетворення, цільова функція.

MULTICHANNEL NONGAUSSIAN DATA PROBABILISTIC MODEL ON THE BASIS OF TRUNCATED MULTINORMAL DISTRIBUTION

I. K. Vasilyeva

The multivariate probabilistic model for description of nongaussian data simultaneous distribution on the basis of truncated multinormal distribution is proposed. Techniques of the model parameters estimation as well as boundary equations formation specifying truncation subspace involving correlations between components of a random vector of observations are described. The results of the approximation using proposed model of two-dimensional random variates, which marginal distributions differ substantially from the normal one, are presented. It is shown that the proposed model can be used in analytic description of correlated statistics simultaneous distribution including the case then separated components of the vector of observations are characterized by different distributions.

Key words: approximation, probabilistic model, multivariate distribution, truncated normal distribution, distribution parameters, correlating transform, criterion function.

Васильева Ирина Карловна – канд. техн. наук, доцент, доцент кафедры производства радиоэлектронных систем летательных аппаратов, Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Харьков, Украина, e-mail: i.vasilyeva@mail.ru.