ВОСПРОИЗВЕДЕНИЕ ДИСКРЕТНЫХ РАСПРЕДЕЛЕНИЙ С ЗАДАННОЙ КОВАРИАЦИОННОЙ СТРУКТУРОЙ

А.А.Новоселов
Сибирский федеральный университет
Красноярский государственный торгово-экономический институт

Аннотация

Воспроизведение многомерного нормального распределения с заданной корреляционной структурой является хорошо изученной задачей, и обычно осуществляется с использованием факторизации ковариационной матрицы. Для дискретных распределений такой метод неприменим. В настоящей работе описывается один метод построения двумерного дискретного распределения по заданным маргинальным распределениям и коэффициенту корреляции. Метод основан на смешивании некоторых базисных двумерных дискретных распределений на заданной сетке.

Введение

В задачах управления риском часто возникает проблема построения многомерного вероятностного распределения с заданными маргинальными распределениями и отдельными моментными характеристиками, а также воспроизведения значений случайного вектора, обладающего таким распределением. В случае непрерывных распределений для решения этой задачи имеются хорошо разработанные методы. Например, для случайного вектора $X = (X_1, \ldots, X_d)$, имеющего многомерное нормальное распределение с вектором средних $a = (a_1, \ldots, a_d)$ и ковариационной матрицей C размера $d \times d$ известны алгоритмы воспроизведения значений этого вектора, основанные на вращении и масштабировании стандартного нормального вектора [1]. Аналогичные методы имеются и для других непрерывных распределений.

В случае дискретных распределений такие методы оказываются совершенно неработоспособными, поскольку при вращении не сохраняется заданная сетка значений случайного вектора. В настоящей работе предлагается один метод построения двумерного дискретного распределения по заданным маргинальным распределениям и коэффициенту корреляции. Метод основан на смешивании некоторых базовых распределений и использует свойства смесей вероятностных распределений.

Смесь двумерных распределений

Обозначим $X = (X_1, X_2)$ двумерный случайный вектор, принимающий значения на прямоугольной сетке

$$\{x_{11}, \ldots, x_{1m}\} \times \{x_{21} \ldots, x_{2n}\} \tag{1}$$

размера $m \times n$. Вероятностное распределение этого вектора описывается матрицей R с элементами

$$r_{ij} = \mathbf{P}(X_1 = x_{1i};\ X_2 = x_{2j});\ i = 1, \dots, m;\ j = 1, \dots, n, \qquad (2)$$

а маргинальные распределения компонент – векторами p, q вида

$$p_i = \mathbf{P}(X_1 = x_{1i}) = \sum_{j=1}^{n} r_{ij},\ i = 1, \dots, m \qquad (3)$$

и

$$q_j = \mathbf{P}(X_2 = x_{2j}) = \sum_{i=1}^{m} r_{ij},\ j = 1, \dots, n, \qquad (4)$$

соответственно.

Если имеется несколько двумерных распределений, заданных на одной и той же сетке значений (1), и описываемых матрицами вероятностей $R^{(k)}$, $k = 1, \dots, N$, а также набор неотрицательных чисел $\lambda = (\lambda_, \dots, \lambda_N)$, в сумме дающих 1, то λ-смесью распределений называется двумерное распределение, заданное на сетке значений (1), и описываемое матрицей вероятностей

$$R = \sum_{k=1}^{N} \lambda_k R^{(k)}. \qquad (5)$$

Моментом распределения с матрицей R порядка s, t называется величина

$$\mu_{st} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{1i}^{s} x_{2j}^{t} r_{ij}. \qquad (6)$$

В частности, $a_1 = \mu_{10}$ и $a_2 = \mu_{01}$ являются средними значениями маргинальных распределений, а $\sigma_1^2 = \mu_{20} - \mu_{10}^2$ и $\sigma_2^2 = \mu_{02} - \mu_{01}^2$ - их дисперсиями. Прямым вычислением нетрудно убедиться в том, что произвольный момент μ_{st} λ-смеси двумерных распределений является соответствующей смесью моментов

$$\mu_{st}^{(k)} = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{1i}^{s} x_{2j}^{t} r_{ij}^{(k)},$$

а именно:

$$\mu_{st} = \sum_{k=1}^{N} \lambda_k \mu_{st}^{(k)}. \qquad (7)$$

Если же все смешиваемые распределения имеют одинаковые маргинальные распределения, то аналогичным свойством обладают и коэффициенты корреляции

$$\rho = \frac{\mu_{11} - a_1 a_2}{\sigma_1 \sigma_2},\ \ \rho^{(k)} = \frac{\mu_{11}^{(k)} - a_1 a_2}{\sigma_1 \sigma_2},\ k = 1, \dots, N, \qquad (8)$$

то есть, коэффициент корреляции λ-смеси является λ-смесью коэффициентов корреляции

$$\rho = \sum_{k=1}^{N} \lambda_k \rho^{(k)}. \tag{9}$$

На этом факте и основан метод получения совместного распределения с заданными маргинальными распределениями и коэффициентом корреляции, который описывается в следующем разделе. Для дальнейшего обозначим $\mathcal{F}(p,q)$ класс распределений на сетке (1), обладающих заданными маргинальными распределениями p, q, а $\mathcal{F}_\rho(p,q)$ - его подкласс, содержащий распределения с коэффициентом корреляции ρ.

Построение распределения с заданной корреляцией компонент

Сначала опишем понятия комонотонных и антикомонотонных распределений и их связь с корреляцией компонент случайного вектора. Компоненты случайного вектора $X = (X_1, X_2)$ называются комонотонными [2], если найдется случайная величина Z и неубывающие функции f_1, f_2 такие, что $X_1 = f_1(Z)$ и $X_2 = f_2(Z)$. Распределение R такого вектора также называется комонотонным. Компоненты случайного вектора $X = (X_1, X_2)$ называются антикомонотонными [2], если компоненты случайного вектора $(X_1, -X_2)$ комонотонны. Распределение R такого случайного вектора также называется антикомонотонным.

Оказывается [3], что среди всех распределений $R \in \mathcal{F}(p,q)$ наименьшей корреляцией $\rho_{\min} = \rho_{\min}(p,q)$ обладает антикомонотонное распределение, а наибольшей корреляцией $\rho_{\max} = \rho_{\max}(p,q)$ - комонотонное распределение. Обозначим $R^+(p,q),\ R^-(p,q) \in \mathcal{F}(p,q)$ комонотонное и антикомонотонное распределения, соответствующие маргинальным распределениям p, q. Обозначим еще $R^\diamond(p,q)$ независимое распределение из $\mathcal{F}(p,q)$.

Ясно, что

$$\mathcal{F}(p,q) = \bigcup_{\rho \in [\rho_{\min}(p,q), \rho_{\max}(p,q)]} \mathcal{F}_\rho(p,q), \tag{10}$$

причем каждый подкласс $\mathcal{F}_\rho(p,q)$ содержит, вообще говоря, много различных распределений при $\rho \in [\rho_{\min}(p,q), \rho_{\max}(p,q)]$, и является пустым при $\rho \notin [\rho_{\min}(p,q), \rho_{\max}(p,q)]$. Отметим также, что всегда

$$0 \in [\rho_{\min}(p,q), \rho_{\max}(p,q)].$$

Для фиксированного $\rho \in [\rho_{\min}(p,q), \rho_{\max}(p,q)]$ вычислим

$$\lambda_\rho = \frac{\rho - \rho_{\min}}{\rho_{\max} - \rho_{\min}} \tag{11}$$

и зададим распределение

$$R_\rho^{-+}(p,q) = \lambda_\rho R^+(p,q) + (1 - \lambda_\rho) R^-(p,q). \tag{12}$$

Используя (9), вычислим корреляцию этой смеси распределений:

$$\lambda_\rho \rho_{\max} + (1-\lambda_\rho)\rho_{\min} = \rho,$$
то есть, распределение $R_\rho^{-+}(p,q)$ имеет в точности требуемую корреляцию.

Другой способ построения смеси определяется по-разному для значений требуемой корреляции разного знака. Если ρ неотрицательно, то смешиваются независимое и комонотонное распределения. Коэффициент смеси вычисляется по формуле:
$$\lambda_\rho^+ = \frac{\rho}{\rho_{\max}}, \quad \rho \geq 0, \tag{13}$$
а искомое распределение задается выражением
$$R_\rho^+(p,q) = \lambda_\rho^+ R^+(p,q) + (1-\lambda_\rho^+) R^\diamond(p,q). \tag{14}$$
В случае неположительного ρ коэффициент смеси равен
$$\lambda_\rho^- = \frac{\rho}{\rho_{\min}}, \quad \rho \leq 0, \tag{15}$$
а распределение вычисляется по формуле
$$R_\rho^-(p,q) = \lambda_\rho^- R^-(p,q) + (1-\lambda_\rho^-) R^\diamond(p,q). \tag{16}$$
Как и ранее, легко проверяется, что корреляция распределений (14) и (16) равна в точности ρ.

Пример

Пусть сетка значений (1) имеет вид $\{0,1,2\} \times \{0,1\}$, а маргинальные распределения заданы посредством
$$p = (1/3, 1/2, 1/6), \quad q = (2/5, 3/5). \tag{17}$$
Произвольное распределение R с такими маргинальными распределениями, с учетом четырех независимых ограничений (3), (4), может быть представлено в двухпараметрическом виде, приведенном в табл. 1.

Таблица 1. Совместное распределение $R(p,q)$

x_2	p	R			
0	2/5	u	v	$2/5-u-v$	
1	3/5	$1/3-u$	$1/2-v$	$u+v-7/30$	
		0	1	2	x_1
		1/3	1/2	1/6	q

Область допустимых значений параметров (u,v) задается неравенствами
$$0 \leq u \leq \frac{1}{3}, \;\; 0 \leq v \leq \frac{1}{2}, \quad \frac{7}{30} \leq u+v \leq \frac{2}{5}$$
и изображена на рис. 1. Комонотонное, антикомонотонное и независимое распределения задаются парами значений параметров $\left(\frac{1}{3}, \frac{1}{15}\right)$, $\left(0, \frac{7}{30}\right)$, $\left(\frac{2}{15}, \frac{1}{5}\right)$,

соответственно, имеют вид, представленный в табл. 2 – 4, и отмечены на рис. 1.

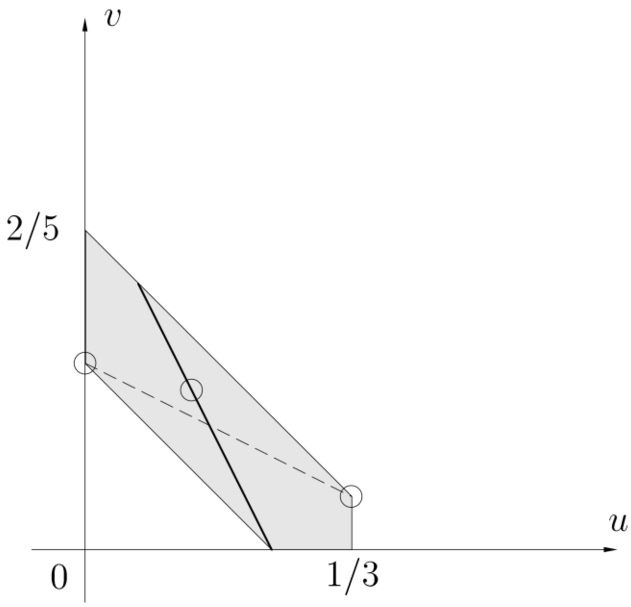

Рис. 1. Область допустимых значений параметров (u, v); окружностями отмечены комонотонное $R^{+}(p, q)$ (слева), антикомонотонное $R^{-}(p, q)$ (справа) и независимое $R^{\diamond}(p, q)$ (в середине) распределения; на жирной линии располагаются некоррелированные распределения, а на штриховой линии – всевозможные смеси экстремальных распределений $R^{-}(p, q)$ и $R^{+}(p, q)$; пересечение этих линий соответствует распределению $R_0^{-+}(p, q)$, приведенному в табл. 5

Таблица 2. Комонотонное распределение $R^{+}(p, q)$

x_2	p		R		
0	2/5	1/3	1/15	0	
1	3/5	0	13/30	1/6	
		0	1	2	x_1
		1/3	1/2	1/6	q

Непосредственный расчет дает значения корреляции экстремальных распределений $\rho_{\min} = -7/\sqrt{102}$, $\rho_{\max} = 8/\sqrt{102}$.

Зададим $\rho = 0$ и вычислим распределение по формулам (11), (12). Имеем $\lambda = 7/15$, а распределение $R_0^{-+}(p, q)$ приведено в табл. 5. Отметим, что это распределение с некоррелированными компонентами отличается от распределения $R^{\diamond}(p, q)$ табл. 4 с независимыми компонентами.

Таблица 3. Антикомонотонное распределение $R^{-}(p,q)$

x_2	p	R			
0	2/5	0	7/30	1/6	
1	3/5	1/3	4/15	0	
		0	1	2	x_1
		1/3	1/2	1/6	q

Таблица 4. Независимое распределение $R^{\diamond}(p,q)$

x_2	p	R			
0	2/5	2/15	1/5	1/15	
1	3/5	1/5	3/10	1/10	
		0	1	2	x_1
		1/3	1/2	1/6	q

Таблица 5. Смесь (12) для $\rho = 0$ ($\lambda = 7/15$)

x_2	p	R			
0	2/5	7/45	7/45	4/45	
1	3/5	8/45	31/90	7/90	
		0	1	2	x_1
		1/3	1/2	1/6	q

Заключение

В работе исследованы свойства моментов смеси многомерных распределений и рассмотрены два метода построения и воспроизведения двумерного дискретного распределения с заданной корреляцией. Методы основаны на смеси экстремальных и независимого совместных распределений, легко реализуются алгоритмически и проиллюстрированы в работе примерами.

Литература

1. P. Glasserman, Monte Carlo Methods in Financial Engineering. Springer, 2003, 602 p.
2. S.S. Wang, Aggregation of Correlated Risk Portfolios: Models and Algorithms. Proceedings of the Casualty Actuarial Society, 1998, p. 848-939
3. R.B. Nelsen, An Introduction to Copulas. Springer, 1998, 236 p.