УДК 519.06

Иерархические типы структур статистических зависимостей

К. А. Коновалов, Е. Ю. Щетинин

Кафедра прикладной математики
МГТУ «Станкин»
Вадковский пер., 3а, Москва, Россия, 127055

В работе предложен подход к моделированию многомерных структур статистических зависимостей на основе введённой спецификации иерархий двумерных условных копул. Разработаны численные алгоритмы симуляции иерархических структур и оценивания параметров копул, входящих в эти иерархии.

Ключевые слова: условная копула, иерархическая структура, каскад, правильный каскад.

Моделирование многомерных структур статистических зависимостей является современной успешно развивающейся математической теорией [1]. Концепция копулы [2], отделяющая частные распределения случайного вектора от описания структуры зависимости, представляется эффективным и универсальным инструментом для описания сложных распределений большой размерности. Хорошо изученным в настоящее время является двумерный случай: предложены семейства копул, описывающие распределения экстремального типа, распределения с асимметрией и методы их оценивания. При обобщении семейств копул на случаи более высокой размерности существует проблема выражения копулы в явном виде, а оценивание представляет собой сложную численную задачу. Известные модели эллиптических (Гаусса, Стьюдента) и архимедовых копул обычно имеют один параметр для характеризации хвоста распределения вне зависимости от размерности, тогда как К. Аас показывает [3], что связь между парами рисковых факторов в портфеле может различаться значительно.

Одним из подходов к решению поставленной проблемы является моделирование связей случайного вектора через описание попарных связей, впервые предложенное Джоем [4] и получившее развитие в работах Бедфорда и Кука [5]. Принцип моделирования основывается на декомпозиции многомерной плотности вероятности в каскад двумерных копул основных величин, а также производных от них условных и безусловных функций распределения. В настоящей работе развивается этот подход.

Пусть имеется случайный вектор $X = (X_1, X_2, \ldots, X_n)$ с совместной плотностью вероятности $f(x_1, x_2, \ldots, x_n)$, которая может быть записана через условные распределения в виде

$$f(x_1, \ldots, x_n) = f(x_n) f(x_{n-1}|x_n) f(x_{n-2}|x_{n-1}, x_n) \ldots f(x_1|x_2, \ldots, x_n). \qquad (1)$$

Согласно теореме Скляра [2] каждая многомерная функция распределения может быть представлена через подходящую копулу C,

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)). \qquad (2)$$

Возвращаясь к совместной плотности вероятности всюду непрерывной случайной величины со строго возрастающими частными распределениями $F_1, F_2, \ldots, F_n$, перепишем её в виде

$$f(x_1, x_2, \ldots, x_n) = c_{12\ldots n}(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \cdot f_1(x_1) f_2(x_2) \ldots f_n(x_n) \qquad (3)$$

Статья поступила в редакцию 18 июня 2009 г.

для некоторой n-мерной плотности копулы $c_{12\ldots n}(u_1, u_2, \ldots, u_n)$, $u_i \in [0,1]$. В двумерном случае последняя формула (3) принимает вид

$$f(x_1, x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) f_2(x_2).$$

Условная плотность в этом случае выражается формулой

$$f(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) f_1(x_1).$$

В случае нескольких переменных в условии возможны альтернативные декомпозиции, а произвольное условное частное распределение может быть представлено в виде

$$f(x|\upsilon) = c_{x\upsilon_j|\upsilon_{-j}}(F(x|\upsilon_{-j}), F(\upsilon_j|\upsilon_{-j})) \cdot f(x|\upsilon_{-j}), \tag{4}$$

где υ — d-мерный вектор, часть вектора $(x_1, x_2, \ldots, x_n)$, υ_j обозначает произвольно выбранную переменную из υ, а υ_{-j} — вектор без этой компоненты. Таким образом, плотность вероятности многомерного распределения (1) может быть выражена через двумерные копулы, аргументы которых, в свою очередь, являются одномерными условными распределениями. В силу итеративности этой конструкции возможна множественность подобных декомпозиций.

Для организации иерархических типов структур статистических зависимостей введём несколько определений.

Определение 1. K — каскад на n элементах, если

1. $K = (T_1, T_2, \ldots, T_m)$,
2. T_1 — дерево с вершинами $N_1 = \{1, 2, \ldots, n\}$ и множеством рёбер E_1,
3. Для $i = 2, \ldots, m$, T_i — дерево с вершинами $N_i \subset N_1 \cup E_1 \cup E_2 \cup \ldots \cup E_{i-1}$ и множеством рёбер E_i.

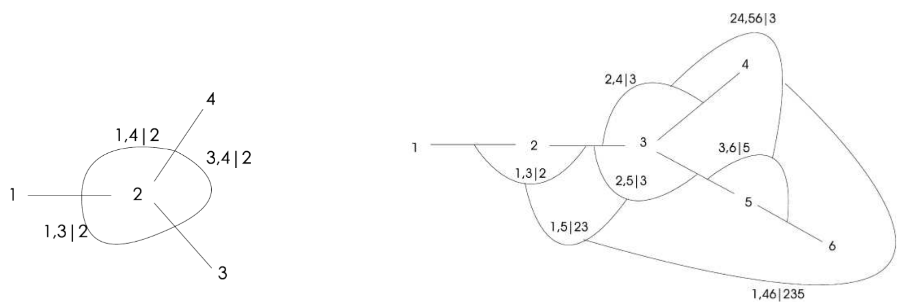

Рис. 1. **Примеры неправильных иерархий копул**

Определение 2. Каскад называется правильным, если

1. $m = n$,
2. T_i – связанное дерево с множеством рёбер E_i и вершин $N_i = E_{i-1}$, и $card(N_i) = n - (i-1)$ для $i = 1, 2, \ldots, n$,
3. Выполняется условие соседства: для $i = 2, 3, \ldots, n-1$, если $a = \{a_1, a_2\}$ и $b = \{b_1, b_2\}$ — вершины в N_i, соединённые ребром (т. е. $a_1, a_2, b_1, b_2 \in N_{i-1}$), то $card(a \cap b) = 1$.

Определение 3. (F, K, B) — спецификация иерархии копул, если

1. $F = (F_1, F_2, \ldots, F_n)$ – вектор одномерных непрерывных и обратимых функций распределения,
2. K — правильный каскад на n элементах,
3. $B = \{B_e | e \in E_i, i = 1, 2, \ldots, n\}$, где B_e — подмножество некоторого семейства копул.

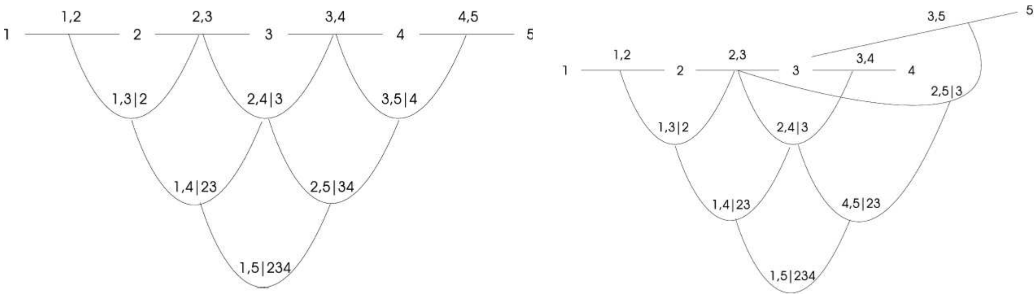

Рис. 2. **Примеры правильных иерархий копул**

Таким образом, совместное распределение F на R^n реализует спецификацию правильного каскада (F, K, B), если частные распределения F описываются F_i, $i = 1, 2, \ldots, n$, и $e \in E_i$ — двумерная копула — элемент из B_e.

Среди спецификаций иерархий копул на правильных каскадах можно выделить семейство, в котором предполагается связь между ведущей, предикативной случайной величиной и остальными компонентами случайного вектора, канонические иерархии копул. На основе этой математической модели построен алгоритм симуляции случайных величин, связанных известной зависимостью, и предложен метод оценивания параметров копул, входящих в представление, для многомерных выборок. Введём функцию h, связанную с формулой (4), где x и v соответствуют одномерным и равномерным на отрезке $[0, 1]$ случайным величинам, т. е. $F(x) = x$, $F(v) = v$,

$$h(x, v, \Theta) = F(x|v) = \partial C_{xv}(x, v, \Theta)/\partial v,$$

и предположим, что существует обратная функция $h^{-1}(u, v, \Theta)$ относительно первой переменной $h(x, v, \Theta) = u$. Для симуляции n зависимых равномерных на отрезке $[0, 1]$ переменных необходимо выполнить следующие шаги.

Шаг 1. Получить w_i, $i = 1, 2, \ldots, n$, независимые, равномерные на $[0, 1]$.

Шаг 2. $x_1 = w_1$, $x_2 = F^{-1}_{2|1}(w_2|x_1) = h^{-1}(w_2, x_1, \Theta_{11})$,

$$x_3 = F^{-1}_{3|1,2}(w_3|x_1, x_2) = h^{-1}\left(h^{-1}(w_3, h(x_2, x_1, \Theta_{11}), \Theta_{21}), x_1, \Theta_{12}\right)$$

$$\ldots = \ldots$$

$$x_n = F^{-1}_{n|1,2,\ldots,n-1}(w_n|x_1, x_2, \ldots, x_{n-1}).$$

Шаг 3. Повторить первый и второй шаги до необходимого объёма выборки.

Задача оценивания может быть решена методом максимального правдоподобия. Будем предполагать, что имеется выборка n-мерного случайного вектора в T временных точках, $x_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,T})$, $i = 1, 2, \ldots, n$. Для иерархии копул с каноническим каскадом функция максимального правдоподобия имеет вид

$$\sum_{j=1}^{n-1}\sum_{i=1}^{n-j}\sum_{t=1}^{T} \lg\left(c_{j,j+i|1,2,\ldots,j-1}\left(F(x_{j,t}|x_{1,t}, \ldots, x_{j-1,t}), F(x_{j+i,t}|x_{1,t}, \ldots, x_{j-1,t})\right)\right) \qquad (5)$$

Каждая из копул, входящих в (5), имеет, по крайней мере, один параметр, относительно которого функция (5) должна быть максимизирована. Алгоритм оценивания приведён ниже.

Шаг 1. Оценить параметры копул на первом уровне каскада T_1.

Шаг 2. Вычислить условные выборки для второго уровня каскада T_2 с использованием параметров копул, полученных на первом шаге, и функции h.

Шаг 3. Если не достигнут последний уровень каскада, перейти к первому шагу.

Отдельной задачей оценивания является выбор семейств копул, который может быть осуществлён визуально по диаграмме рассеяния или χ^2-диаграмме, или более строго при помощи процедуры Goodness-of-fit.

Литература

1. *Щетинин Е. Ю.* Теория математических структур статистической зависимости. — ИЦ ГОУ ВПО МГТУ СТАНКИН, 2005.
2. *Nelsen R. B.* Copulas, Characterization, Correlation and Counterexamples // Math. Mag. — 1995. — Vol. 68.
3. *Aas K.* The Generalized Hyperbolic Skew Student‘s t-Distribution // Journal of Financial Econometrics Advance Access. — 2006.
4. *Joe H.* Multivariate Models and Dependence Concepts. — London: Chapman & Hall, 1997.
5. *Bedford T., Cooke R. M.* Vines — a New Graphical Model for Dependent Random Variables // Annals of Statistics. — 2002. — Vol. 30 (4). — Pp. 1031–1068.

UDC 519.06

Hierarchical Types of Statistical Dependences

K. A. Konovalov, Eu. Yu. Schetinin

Applied Mathematics Department
MGTU “Stankin”
3a, Vadkovsky lane, Moscow, Russia, 127055

We offer an approach to modeling multi-dimensional structures of statistical dependencies based on the specification of hierarchies of conditional two-dimensional copulae. Numerical algorithms for the simulation of hierarchical structures and evaluating copula parameters of those hierarchies were developed.

Key words and phrases: conditional copula, hierarchical structure, vine, regular vine.