

Д. Фантаццини

Моделирование многомерных распределений с использованием копула-функций. II¹

Статья содержит вторую часть консультации, посвященной копула-функциям и их использованию в моделировании многомерных распределений вероятностей. В ней описываются парные копула-функции (включая понятия канонической и D-ветвизации), различные характеристики зависимости анализируемых случайных величин (в том числе меры хвостовой зависимости, особенно актуальные в случае несимметричных распределений), а также параметрические, полупараметрические и непараметрические методы статистического оценивания распределений, представленных с помощью копула-функций.

Ключевые слова: парные копула-функции, ветвизация, меры зависимости, хвостовая зависимость, метод максимального правдоподобия (одношаговый, двухшаговый, канонический, трехшаговый), полупараметрические и непараметрические методы статистического оценивания.

JEL classification: C69, C49.

Содержание

4. Парные копула-функции

- 4.1. Введение
- 4.2. Каноническая ветвизация
- 4.3. D-ветвизация
- 4.4. Эмпирические приложения в статистическом пакете R: построение парных копула-функций

5. Меры зависимости

- 5.1. Коэффициент корреляции
- 5.2. Коэффициенты ранговой корреляции: ρ -Спирмена и τ -Кендалла
- 5.3. Хвостовая зависимость
- 5.4. Эмпирические приложения в статистическом пакете R: меры зависимости

6. Процедуры оценивания: параметрические методы

- 6.1. Метод максимального правдоподобия (ММП или одношаговый метод)
- 6.2. Двухшаговый метод максимального правдоподобия
- 6.3. Эмпирические приложения в статистическом пакете R: процедуры параметрического оценивания

7. Процедуры оценивания: полупараметрические и непараметрические методы

- 7.1. Канонический метод максимума правдоподобия (КММП)
- 7.2. Трехшаговый канонический метод максимального правдоподобия (КМЕ–СМЛ метод)
- 7.3. Методы непараметрического оценивания
- 7.4. Эмпирические приложения в статистическом пакете R: процедуры полупараметрического оценивания

¹ Продолжение. Начало консультации см. в (Фантаццини, 2011). Перевод на русский язык осуществлен А. В. Кудровым под научной редакцией С. А. Айвазяна. Сохранена сквозная нумерация разделов, рисунков, формул, определений и теорем, начатая в первой части. Окончание консультации, посвященное описанию подходов к экспериментальному подбору копула-функции и методов статистической проверки гипотез, связанных с моделями копула-функций, предполагается опубликовать в следующем номере нашего журнала.

4. ПАРНЫЕ КОПУЛА-ФУНКЦИИ

4.1. Введение

Детальное изучение парных копула-функций было начато в работе (Joe, 1997), а позднее продолжено в (Bedford, Cooke, 2001, 2002; Kurowicka, Cooke, 2006) (моделирование) и (Aas et al., 2009) (статистические выводы). Их использование дает возможность получить значительно более гибкую структуру зависимости, чем структура перестановочных или иерархически вложенных архимедовых копула-функций. При помощи парных копула-функций можно разложить многомерную плотность на произведение $n(n-1)/2$ двумерных копула-функций, из которых $n-1$ являются безусловными, а остальные — условными. Важно отметить, что используемые двумерные копула-функции не обязаны принадлежать одному классу.

Рассмотрим n случайных величин (X_1, \dots, X_n) с совместной функцией распределения $H(x_1, \dots, x_n)$, частными распределениями $F_i(x_i)$, $i = 1, \dots, n$, совместной плотностью $f(x_1, \dots, x_n)$ и плотностями частных распределений $f_i(x_i)$, $i = 1, \dots, n$. Тогда для точек, в которых все плотности непрерывны и положительны²:

$$f(x_1, \dots, x_n) = f(x_n | x_1, \dots, x_{n-1}) \cdot f(x_1, \dots, x_{n-1}) = \prod_{i=2}^n f(x_i | x_1, \dots, x_{i-1}) \cdot f_1(x_1). \quad (11)$$

Из теоремы Склара известно, что для абсолютно непрерывного n -мерного распределения $H(\cdot)$, имеющего строго возрастающие непрерывные частные распределения $F_1(\cdot), \dots, F_n(\cdot)$, имеем:

$$f(x_1, \dots, x_n) = c_{1, \dots, n}(F_1(x_1), \dots, F_n(x_n)) \cdot f_1(x_1) \cdot \dots \cdot f_n(x_n),$$

где $c_{1, \dots, n}(\cdot, \dots, \cdot)$ — n -мерная плотность копула-функции, описывающей анализируемое распределение $H(x_1, \dots, x_n)$, а $F_1(x_1), \dots, F_n(x_n)$ — значения частных функций распределения рассматриваемых случайных величин в точках x_1, \dots, x_n соответственно.

В двумерном случае

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

откуда следует, что:

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

Для трех случайных величин X_1, X_2 и X_3 имеем:

$$f(x_1 | x_2, x_3) = c_{12|3}(F(x_1 | x_3), F(x_2 | x_3)) f(x_1 | x_3),$$

где $c_{12|3}(\cdot)$ — копула-функция для условных функций распределений $((X_1 | X_3), (X_2 | X_3))$. Кроме того,

$$f(x_1 | x_2, x_3) = c_{13|2}(F(x_1 | x_2), F(x_3 | x_2)) f(x_1 | x_2).$$

² В правой части формулы (11) и некоторых других местах далее для того, чтобы избежать усложнения обозначений, функции условной плотности и условного распределения случайной величины X_i при условиях на переменные с номерами J_1, J_2, \dots, J_k не будут помечаться соответствующими индексами $i | J_1, J_2, \dots, J_k$.

Последнее соотношение может быть обобщено на n -мерный случай:

$$f(x | \mathbf{v}) = c_{xv_j | v_{-j}} \left(F(x | v_{-j}), F(v_j | v_{-j}) \right) \cdot f(x | v_{-j}), \quad (12)$$

где \mathbf{v} — вектор значений компонент, на которые наложены условия, v_j — произвольно выбранная компонента вектора \mathbf{v} , а через \mathbf{v}_{-j} обозначен вектор \mathbf{v} , из которого исключена j -ая компонента.

Теперь, если воспользоваться (12), то из (11) получим следующее представление для $f(x_t | x_1, \dots, x_{t-1})$:

$$f(x_t | x_1, \dots, x_{t-1}) = \frac{f(x_t, x_{t-1} | x_1, \dots, x_{t-2})}{f(x_{t-1} | x_1, \dots, x_{t-2})} = c_{t-1, t | 1, \dots, t-2} \cdot f(x_t | x_1, \dots, x_{t-2}), \quad (13)$$

где для различающихся индексов i, j, i_1, \dots, i_k таких, что $i < j$ и $i_1 < \dots < i_k$, можно выразить двумерную условную копула-функцию при значениях аргументов, равных соответствующим условным функциям распределения, а именно:

$$c_{i, j | i_1, \dots, i_k} = c_{i, j | i_1, \dots, i_k} \left(F(x_i | x_{i_1}, \dots, x_{i_k}), F(x_j | x_{i_1}, \dots, x_{i_k}) \right).$$

Используя (13), можно записать (11) в виде:

$$\begin{aligned} f(x_1, \dots, x_n) &= f(x_1) \cdot \prod_{t=2}^n \prod_{k=1}^{t-1} c_{t-k, t | 1, \dots, t-k-1} \cdot f(x_t) = \prod_{r=1}^n f(x_r) \cdot \prod_{t=2}^n \prod_{k=1}^{t-1} c_{t-k, t | 1, \dots, t-k-1} = \\ &= \prod_{r=1}^n f(x_r) \cdot \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j, j+1 | 1, \dots, j-1}, \quad \text{где } j = t - k, \quad j + i = t. \end{aligned} \quad (14)$$

Представление (14) называют разложением парных копула-функций. Для многомерных распределений существует много возможностей разложения на парные копула-функции. В (Bedford, Cooke, 2001, 2002) представлена графическая модель, называемая *моделью регулярной ветвизации*. Основной идеей этого метода является представление разложения на парные копула-функции типа (14) в виде последовательности вложенных деревьев с неориентированными ребрами, названными ветвями. В n -мерном случае ветвизация представлена $n - 1$ деревом так, что j -ое дерево имеет $n + 1 - j$ узлов и $n - j$ ребер. Каждое ребро отвечает плотности некоторой парной копула-функции, а ребра j -го дерева становятся узлами $(j + 1)$ -го дерева. Два узла $(j + 1)$ -го дерева связаны ребром в том случае, если соответствующие ребра j -го дерева имеют общий узел. Разложение на парные копула-функции определяется $n(n - 1) / 2$ ребрами, а также плотностями частных распределений.

Сосредоточим внимание на двух примерах регулярной ветвизации, которые на сегодняшний день привлекли большое внимание исследователей. Речь пойдет о *канонической ветвизации* и *D-ветвизации*. Для более детальной информации относительно общей теории регулярной ветвизации и ее графического представления рекомендуется работа (Kurowicka, Cooke, 2006).

Рассмотрение упомянутых выше примеров регулярной ветвизации начнем с двумерных условных распределений, которые являются ключевыми элементами этой процедуры.

Условные распределения для наиболее часто используемых двумерных копула-функций

В работе (Джозеф, 1997) показано, что:

$$F(x | \mathbf{v}) = \frac{\partial C_{x,v_j | \mathbf{v}_{-j}}(F(x | \mathbf{v}_{-j}), F(v_j | \mathbf{v}_{-j}))}{\partial F(v_j | \mathbf{v}_{-j})} \text{ для любого } j, \tag{15}$$

где $C_{x,v_j | \mathbf{v}_{-j}}$ — условная двумерная копула-функция (формальное определение см. выше). Если \mathbf{v} имеет размерность 1 (т. е. $\mathbf{v} = v$), то (15) принимает вид:

$$F(x | v) = \frac{\partial C_{x,v}(F_1(x), F_2(v))}{\partial F(v)}.$$

Пусть $F_1(x) = P(X \leq x) = x$ и $F_2(v) = P(V \leq v) = v$ (т. е. $f(x) = f(v) = 1$). Введем обозначение:

$$h(x, v; \Theta) = F(x | v) = \frac{\partial C_{x,v}(x, v; \Theta)}{\partial v}, \tag{16}$$

где через Θ обозначены параметры копула-функции совместного распределения $P(X \leq x, V \leq v)$. Согласно терминологии работы (Аас и др., 2009) функцию $h(x, v; \Theta)$ назовем h -функцией, а $h^{-1}(u, v; \Theta)$ — обратной h -функцией относительно первого аргумента u .

Для двумерной *нормальной копула-функции* h -функция имеет вид:

$$h(u_1, u_2; \rho_{12}) = \Phi \left(\frac{\Phi^{-1}(u_1) - \rho_{12} \Phi^{-1}(u_2)}{\sqrt{1 - \rho_{12}^2}} \right),$$

а обратная h -функция относительно первого аргумента равна:

$$h^{-1}(u_1, u_2; \rho_{12}) = \Phi \left(\Phi^{-1}(u_1) \sqrt{1 - \rho_{12}^2} + \rho_{12} \Phi^{-1}(u_2) \right),$$

где $\Phi(z)$ — значение функции распределения стандартного нормального закона в точке z , а ρ_{12} — коэффициент корреляции между анализируемыми случайными величинами.

Для двумерной *копула-функции Стьюдента* с ν_{12} числом степеней свободы h -функция равна:

$$h(u_1, u_2; \rho_{12}, \nu_{12}) = t_{\nu_{12}+1} \left(\left(t_{\nu_{12}}^{-1}(u_1) - \rho_{12} t_{\nu_{12}}^{-1}(u_2) \right) \left(\frac{\left(\nu_{12} + \left(t_{\nu_{12}}^{-1}(u_2) \right)^2 \right) (1 - \rho_{12}^2)}{\nu_{12} + 1} \right)^{-1/2} \right),$$

а обратная h -функция относительно первого аргумента равна

$$h^{-1}(u_1, u_2; \rho_{12}, \nu_{12}) = t_{\nu_{12}} \left(t_{\nu_{12}+1}^{-1} \left(\sqrt{\frac{(\nu_{12} + (t_{\nu_{12}}^{-1}(u_2))^2)(1 - \rho_{12}^2)}{\nu_{12} + 1}} \right) + \rho_{12} t_{\nu_{12}}^{-1}(u_2) \right).$$

В приведенных формулах используются следующие обозначения: $t_\nu(z)$ — значение функции распределения Стьюдента с ν степенями свободы в точке z , а $t_\nu^{-1}(u)$ — значение обратной к $t_\nu(z)$ функции распределения в точке u .

Для двумерной копула-функции Клейтона h -функция имеет вид:

$$h(u_1, u_2; \alpha_{12}) = u_2^{-\alpha_{12}-1} (u_1^{-\alpha_{12}} + u_2^{-\alpha_{12}} - 1)^{-1/\alpha_{12}},$$

а обратная h -функция относительно первого аргумента равна:

$$h^{-1}(u_1, u_2; \alpha_{12}) = \left((u_1 \cdot u_2^{\alpha_{12}+1})^{-\frac{\alpha_{12}}{\alpha_{12}+1}} + 1 - u_2^{-\alpha_{12}} \right)^{-1/\alpha_{12}}.$$

Для двумерной копула-функции Гумбеля h -функция принимает вид:

$$h(u_1, u_2; \alpha_{12}) = C_{12}(u_1, u_2) \cdot \frac{1}{u_2} \cdot (-\ln u_2)^{\alpha_{12}-1} \cdot \{(-\ln u_1)^{\alpha_{12}} + (-\ln u_2)^{\alpha_{12}}\}^{1/\alpha_{12}-1},$$

где $C_{12}(u_1, u_2)$ — двумерное распределение Гумбеля. К сожалению, в случае копула-функции Гумбеля обратная h -функция может быть вычислена лишь с использованием численных методов. В условиях большой размерности целесообразнее использовать копула-функцию Клейтона, см. (Joe, 1997) или (Cherubini et al., 2004).

4.2. Каноническая ветвизация

Регулярную ветвизацию, при которой каждое дерево имеет единственный узел, называют *канонической ветвизацией*. Разложение n -мерной плотности, отвечающее канонической ветвизации, представлено в формуле (14), т. е.

$$f(x_1, x_2, \dots, x_n) = \prod_{k=1}^n f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j, j+i|1, \dots, j-1}(F(x_j | x_1, \dots, x_{j-1}), F(x_{j+i} | x_1, \dots, x_{j-1})). \quad (17)$$

На рисунке 11 показаны канонические ветвизации в 5-мерном случае:

Представления вида (17) обладают преимуществом в ситуациях, когда можно выделить одну ключевую переменную, влияющую на другие переменные. Такая переменная может быть использована в качестве корня канонической ветвизации (переменная 1 на рис. 11).

Генерирование многомерных наблюдений

В работе (Aas et al., 2009) показано, что алгоритм генерирования n зависимых случайных величин является общим как для канонической, так и для D-ветвизации:

- смоделируем независимые равномерно распределенные на $[0, 1]$ случайные величины w_1, \dots, w_n ;

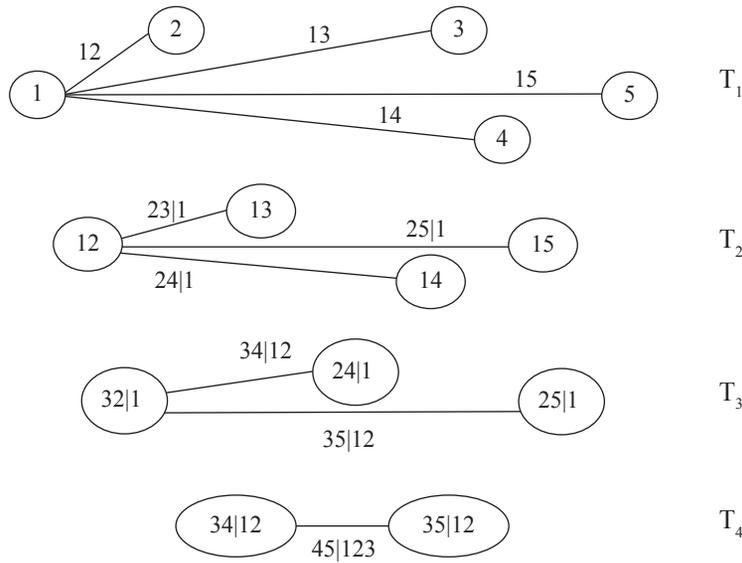


Рис. 11. Примеры канонических ветвизаций в пятимерном случае

- ВОЗЬМЕМ

$$x_1 = w_1; x_2 = F^{-1}(w_2 | x_1); x_3 = F^{-1}(w_3 | x_1, x_2); \dots; x_n = F^{-1}(w_n | x_1, \dots, x_{n-1}).$$

Для того чтобы при любом j определить распределение случайной величины $(X_j | X_1, X_2, \dots, X_{j-1})$, в работе (Aas et al., 2009) предлагается использовать h -функцию, заданную соотношением (16), и соотношение (15). Отметим, что отличие канонической ветвизации от D-ветвизации состоит в выборе переменной v_j для (15). Кроме того,

$$F(x_j | x_1, \dots, x_{j-1}) = \frac{\partial C_{j,j-1|1, \dots, j-2}(F(x_j | x_1, \dots, x_{j-2}), F(x_{j-1} | x_1, \dots, x_{j-2}))}{\partial F(x_{j-1} | x_1, \dots, x_{j-2})}.$$

На основании этих результатов в работе (Aas et al., 2009) предложена следующая процедура генерирования наблюдений, имеющих плотность в форме канонической ветвизации:

```

Sample  $w_1, \dots, w_n$  independent uniform on  $[0, 1]$ 
Set  $x_1 = v_{1,1} = w_1$ 
for  $i \leftarrow 2, \dots, n$ 
   $v_{i,1} = w_i$ 
  for  $k \leftarrow i-1, i-2, \dots, 1$ 
     $v_{i,k} = h^{-1}(v_{i,1}, v_{k,k}, \Theta_{k,i-k})$ 
  end for
   $x_i = v_{i,1}$ 
  if  $i == n$  then
    Stop
  end if
  for  $j \leftarrow 1, \dots, i-1$ 
     $v_{i,j+1} = h(v_{i,j}, v_{j,j}, \Theta_{j,i-j})$ 
  end for
end for
    
```

Внешний цикл указанной выше процедуры работает по порядковым номерам компонент моделируемых многомерных наблюдений. Он включает в себя два подцикла: первый из них позволяет вычислить смоделированное значение i -ой компоненты наблюдения, второй — вычислить условные распределения, необходимые для вычисления $(i + 1)$ -ой компоненты. Для вычисления этих условных распределений в работе (Aas et al., 2009) используется h -функция, определенная соотношением (16), для которой в качестве аргументов берутся рассчитанные $v_{i,j} = F(x_i | x_1, \dots, x_{j-1})$. Параметры $\Theta_{j,i}$ для h -функции — это параметры плотности соответствующей копула-функции $c_{j,j+i|1,\dots,j-1}(\cdot | \cdot)$.

Оценивание параметров

Копула-функции с канонической ветвизацией могут быть оценены с помощью метода максимального правдоподобия. Для простоты положим, как и в (Aas et al., 2009), что $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,T})$, $i = 1, \dots, n$ независимы во времени. Такое предположение не является ограничительным, поскольку при наличии зависимости можно на первом этапе оценить одномерные модели временных рядов, а затем для стандартизированных остатков использовать копула-функцию с канонической ветвизацией. Такой подход к оцениванию можно считать расширением метода максимального псевдоправдоподобия. Для копула-функций он впервые был предложен в работе (Oakes, 1994), а позднее в работах (Genest et al., 1995; Shih, Louis, 1995) были представлены такие асимптотические свойства получаемых оценок, как состоятельность и асимптотическая нормальность. В (Kim et al., 2007) при помощи методов симуляционного моделирования показано, что в условиях, когда частные распределения неизвестны (это довольно распространенная ситуация в прикладных задачах), метод максимального псевдоправдоподобия дает более точные оценки, чем метод максимального правдоподобия. Кроме того, в (Kim et al., 2008) показано, что в случае многомерных моделей с гетероскедастичностью этот метод дает оценки, являющиеся состоятельными и асимптотически нормальными.

Однако в случае, когда вместо многомерной копула-функции используется разложение на парные копула-функции, методология оценивания будет отличаться от представленной выше (Aas et al., 2009).

Для копула-функции с канонической ветвизацией логарифм правдоподобия имеет следующий вид:

$$\sum_{j=1}^{n-1} \sum_{i=1}^{n-j} \sum_{t=1}^T \ln [c_{j,j+i|1,\dots,j-1}(F(x_{j,t} | x_{1,t}, \dots, x_{j-1,t}), F(x_{j+i,t} | x_{1,t}, \dots, x_{j-1,t}))]. \tag{18}$$

В работе (Aas et al., 2009) предложен следующий алгоритм вычисления логарифмической функции правдоподобия (18)

```

log-likelihood = 0.
for i ← 1, ..., n
  v0,i = xi
end for
for j ← 1, ..., n-1
  for i ← 1, ..., n-j
    log-likelihood = log-likelihood + L(vj-1,1, vj-1,i+1; Θj,i)
  end for
end for
    
```

Моделирование многомерных распределений с использованием копула-функций. II

```

if j==n-1 then
Stop
end if
for i<-1,...,n-j
 $\mathbf{v}_{j,i} = h(\mathbf{v}_{j-1,i+1}, \mathbf{v}_{j-1,i}; \Theta_{j,i})$ 
end for
end for

```

Здесь $L(\mathbf{x}, \mathbf{v}; \Theta)$ — логарифмическая функция правдоподобия выбираемой двумерной копула-функции, параметры которой определены вектором Θ , а наблюдения \mathbf{x} и \mathbf{v} заданы. Кроме того, в упомянутой работе представлена процедура нахождения начальных значений оценок параметров, вычисляемых при помощи численной максимизации логарифмического правдоподобия. Эмпирический анализ показывает, что начальные значения оценок, вычисленных при помощи процедуры из (Aas et al., 2009), и итоговые значения оценок, полученных при одновременном оценивании всех параметров, весьма близки, в то время как значение функции правдоподобия увеличивается незначительно. Это может означать, что процедура вычисления начальных значений оценок параметров дает состоятельную оценку всех параметров, что значительно облегчает вычислительную процедуру. Однако этот вопрос требует более детального исследования.

4.3. D-ветвизация

Регулярная ветвизация, при которой у любого дерева T_j не существует узла, соединенного с более чем двумя ребрами, называется *D-ветвизацией*. Многомерная (n -мерная) плотность, отвечающая D-ветвизации, имеет вид:

$$\prod_{k=1}^n f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1,\dots,i+j-1}(F(x_i | x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j} | x_{i+1}, \dots, x_{i+j-1})).$$

На рисунке 12 показана 5-мерная D-ветвизация:

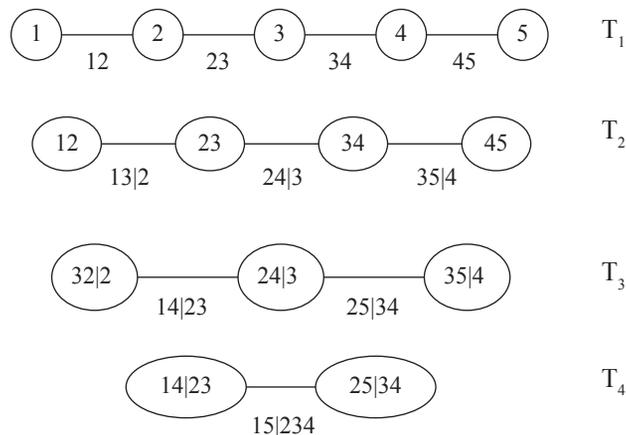


Рис. 12. D-ветвизация для пяти переменных

Генерирование многомерных наблюдений

В работе (Aas et al., 2009) предложена следующая процедура моделирования наблюдений, закон распределения которых представлен с помощью D-ветвизации:

```

Sample  $w_1, \dots, w_n$  independent uniform on  $[0, 1]$ 
Set  $x_1 = v_{1,1} = w_1$ 
Set  $x_2 = v_{2,1} = h^{-1}(w_2, v_{1,1}; \Theta_{1,1})$ 
 $v_{2,2} = h(v_{1,1}, v_{2,1}; \Theta_{1,1})$ 
for  $i \leftarrow 3, \dots, n$ 
   $v_{i,1} = w_i$ 
  for  $k \leftarrow i-1, i-2, \dots, 2$ 
     $v_{i,k} = h^{-1}(v_{i,1}, v_{i-1,2k-2}; \Theta_{k,i-k})$ 
  end for
   $v_{i,1} = h^{-1}(v_{i,1}, v_{i-1,1}; \Theta_{1,i-1})$ 
   $x_i = v_{i,1}$ 
  if  $i = n$  then
    Stop
  end if
   $v_{i,2} = h(v_{i-1,1}, v_{i,1}; \Theta_{1,i-1})$ 
   $v_{i,3} = h(v_{i,1}, v_{i-1,1}; \Theta_{1,i-1})$ 
  for  $i > 3$  then
    for  $j \leftarrow 2, \dots, i-2$ 
       $v_{i,2j} = h(v_{i-1,2j-2}, v_{i,2j-1}; \Theta_{j,i-j})$ 
       $v_{i,2j+1} = h(v_{i,2j-1}, v_{i-1,2j-2}; \Theta_{j,i-j})$ 
    end for
  end if
   $v_{i,2i-2} = h(v_{i-1,2i-4}, v_{i,2i-3}; \Theta_{i-1,1})$ 
end for

```

Как и в случае канонической ветвизации, процедура моделирования D-ветвизации состоит из одного внешнего цикла, включающего один подцикл для моделирования переменных, а также один подцикл для вычисления необходимых условных распределений. Тем не менее, с вычислительной точки зрения этот алгоритм является менее эффективным, чем алгоритм моделирования канонической ветвизации, поскольку число условных распределений, которые необходимо вычислить, для D-ветвизации равно $(n - 2)^2$, а для канонической ветвизации — $(n - 2)(n - 1) / 2$. Отметим, что в случае D-ветвизации параметры $\Theta_{j,i}$ в h -функции — это множество параметров соответствующей копула-функции $c_{i,i+j|i+1, \dots, i+j-1}(\cdot | \cdot)$.

Оценивание параметров

В случае D-ветвизации логарифмическая функция правдоподобия имеет следующий вид:

$$\sum_{j=1}^{n-1} \sum_{i=1}^{n-j} \sum_{t=1}^T \ln \left[c_{i,i+j|i+1, \dots, i+j-1} \left(F(x_{i,t} | x_{i+1,t}, \dots, x_{i+j-1,t}), F(x_{i+j,t} | x_{i+1,t}, \dots, x_{i+j-1,t}) \right) \right].$$

В работе (Aas et al., 2009) предлагается следующий алгоритм вычисления логарифмической функции правдоподобия:

Моделирование многомерных распределений с использованием копула-функций. II

```

log-likelihood = 0.
for i=1,...,n
  v0,i = xi
end for
for i=1,...,n-1
log-likelihood = log-likelihood + l(v0,i, v0,i+1; Θ1,i)
end for
v1,1 = h(v0,1, v0,2; Θ1,1)
for k=1,...,n-3
  v1,2k = h(v0,k+2, v0,k+1; Θ1,k+1)
  v1,2k+1 = h(v0,k+1, v0,k+2; Θ1,k+1)
end for
v1,2n-4 = h(v0,n, v0,n-1; Θ1,n-1)
for j=2,...,n-1
for i=1,...,n-j
log-likelihood = log-likelihood + l(vj-1,2i-1, vj-1,2i; Θj,i)
end for
if j=n-1 then
Stop
end if
vj,1 = h(vj-1,1, vj-1,2; Θj,1)
if n>4 then
for i=1,...,n-j-2
  vj,2i = h(vj-1,2i+2, vj-1,2i+1; Θj,i+1)
  vj,2i+1 = h(vj-1,2i+1, vj-1,2i+2; Θj,i+1)
end for
end if
vj,2n-2j-2 = h(vj-1,2n-2j, vj-1,2n-2j-1; Θj,n-j)
end for

```

Здесь $l(\mathbf{x}, \mathbf{v}; \Theta)$ — логарифмическая функция правдоподобия для выбираемой копула-функции, параметры которой определяются вектором Θ , а векторы наблюдений \mathbf{x} и \mathbf{v} заданы. Отметим, что $\Theta_{j,i}$ — параметры плотности копула-функции $c_{i,j|i+1,\&i,i+j-1}(\cdot|\cdot)$.

4.4. Эмпирические приложения в статистическом пакете R: построение парных копула-функций

Если нужно смоделировать и оценить копула-функцию, используя ее разложение на парные копула-функции, необходимо воспользоваться модулем процедур `copulaGOF` из пакета R, разработанного Даниэлем Бергом. Так, например, D-ветвизация в четырехмерном случае может быть смоделирована и оценена следующим образом:

```

x = SimulateCopulae(n=1000,d=4,construction=list(type="dpcc",copula=c("clayton",
"gumbel","frank","gumbel","clayton","gumbel"),param=list(c(2,3,6,1.3,1,1.4),
rep(0,6)))
pairs(x)

```

```

dpcc.par = EstimateCopulaParameter(x, construction=list(type="dpcc", copula=
c("clayton", "gumbel", "frank", "gumbel", "clayton", "gumbel")))
dpcc.par
$t [1] 1.951387 2.991580 6.052075 1.343258 0.949556 1.410108
$nu [1] 0
$loglik [1] 1899.111
    
```

Биплоты (диаграммы рассеяния пар компонент) для наблюдений из четырехмерной копула-функции с D-ветвизацией представлены на рис. 13.

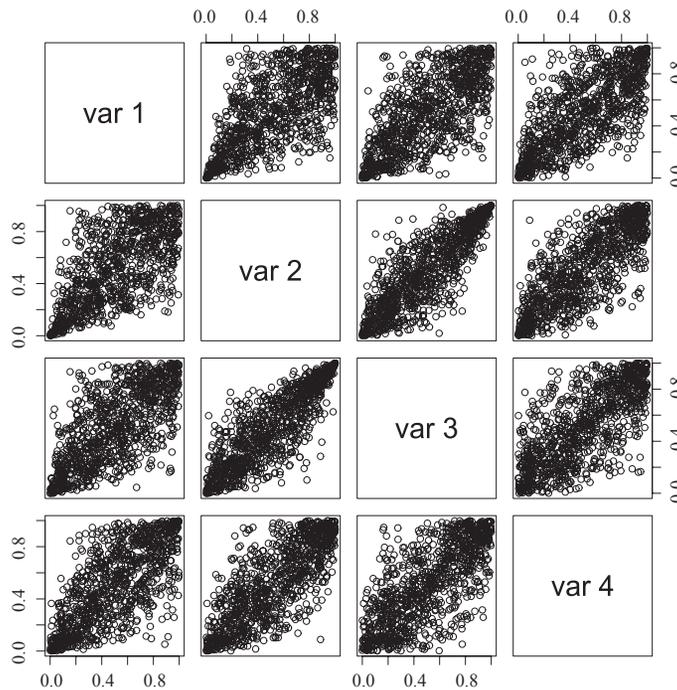


Рис. 13. Диаграмма рассеяния пар компонент для наблюдений из четырехмерной копула-функции с D-ветвизацией

5. Меры зависимости

Меры зависимости являются весьма полезными инструментами для описания структуры двумерной зависимости. В этой главе рассмотрим три возможные меры: коэффициент корреляции, коэффициент ранговой корреляции и коэффициент хвостовой зависимости.

Достаточно хорошей мерой зависимости в классе эллиптических распределений является коэффициент корреляции. Этот класс включает в себя, например, нормальное распределение, смеси нормальных распределений. Однако коэффициент корреляции как мера зависимости для многомерных распределений вне класса эллиптических распределений обладает рядом недостатков. В качестве альтернативы предлагаются две другие меры зависимости, которые в некоторых случаях оказываются более приемлемыми.

Моделирование многомерных распределений с использованием копула-функций. II

5.1. Коэффициент корреляции

Коэффициент корреляции между случайными величинами X и Y , как известно, определяется следующим образом (см., например, (Айвазян, 2010, п. 3.2.2)):

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

где $\text{Cov}(X, Y)$ — ковариация между X и Y , $\text{Var}(X)$ и $\text{Var}(Y)$ — дисперсии X и Y соответственно. Ниже представлены основные свойства коэффициента корреляции.

1. $|\rho(X, Y)| \leq 1$.
2. Если X и Y независимы, то $\rho(X, Y) = 0$.
3. $|\rho(X, Y)| = 1$ тогда и только тогда, когда $P(X = a + bY) = 1$ для некоторых a и $b \neq 0$.
4. $\rho(\alpha X + \beta, \gamma Y + \delta) = \text{sign}(\alpha\gamma)\rho(X, Y)$.
5. Пусть (X, Y) имеет совместное двумерное нормальное распределение с частными стандартными нормальными распределениями. Тогда коэффициент корреляции ρ между X и Y однозначно определяет совместное распределение (X, Y) .

Обобщение коэффициента корреляции на многомерный случай можно найти во многих учебниках по статистике, например, (Mardia et al., 1997).

Пусть $\hat{\rho}$ — выборочный коэффициент корреляции. Хорошо известно, что для проверки гипотезы $H_0 : \rho = 0$ против альтернативы $H_1 : \rho \neq 0$ может быть использована следующая тестовая статистика:

$$\hat{t} = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}},$$

которая при справедливости нулевой гипотезы имеет t_{n-2} -распределение — см., например, (Айвазян, 2010, с. 78).

Если воспользоваться z -преобразованием Фишера, то получим:

$$z = \frac{1}{2} \log\left(\frac{1-\hat{\rho}}{1+\hat{\rho}}\right), \quad \xi = \frac{1}{2} \log\left(\frac{1-\rho}{1+\rho}\right),$$

$$\sqrt{n-3}(z - \xi) \xrightarrow{d} N(0, 1),$$

где знаком \xrightarrow{d} обозначена асимптотическая сходимость по распределению при бесконечно растущем объеме выборки n ($n \rightarrow \infty$). Более того, хорошо известно, что если $\rho \neq 0$, то $\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, (1 - \rho^2)^2)$.

Если X и Y имеют распределения (не обязательно нормальные), для которых конечен четвертый момент и $\rho(X, Y) \neq 0$, то

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, \gamma^2)$$

при некотором γ^2 (см. (Fang et al., 1987; Rodgers, Nicewander, 1988; Embrechts et al., 2002) и ссылки в этих работах).

К сожалению, для многомерных распределений неэллиптического типа коэффициент корреляции обладает рядом недостатков.

1. Равенство коэффициента корреляции нулю эквивалентно независимости в случае многомерного нормального распределения. Однако уже для многомерного распределения Стьюдента это не так.

2. Коэффициент корреляции инвариантен относительно линейных преобразований, но, как правило, не инвариантен относительно более общих преобразований $T(\cdot)$:

$$\rho(T(X), T(Y)) \neq \rho(X, Y).$$

Например, две случайные величины, имеющие логнормальное распределение, имеют коэффициент корреляции, отличный от коэффициента корреляции между их лог-преобразованиями.

3. Частные распределения и корреляционная матрица однозначно определяют лишь только совместное распределение эллиптического типа, но это неверно в общем случае.

4. Дисперсии случайных величин X и Y должны быть конечными, иначе корреляция не определена. Это свойство указывает на то, что коэффициент корреляции — далеко не идеальная мера зависимости, которая может быть неопределенной для распределений с «тяжелыми хвостами».

Рассмотрим пример, иллюстрирующий один из вышеупомянутых случаев. Пусть X_1, X_2 — нормально распределенные случайные величины с нулевым средним и дисперсией $\sigma^2 > 0$, а коэффициент корреляции между ними равен ρ . Тогда коэффициент корреляции случайных величин $Y_i = \exp(X_i)$, $i = 1, 2$, имеющих логнормальное распределение, равен

$$\text{Corr}(Y_1, Y_2) = \frac{e^{\rho\sigma^2} - 1}{e^{\sigma^2} - 1}.$$

При $\rho = 1$ всегда получим, что $\text{Corr}(Y_1, Y_2) = 1$, но наименьшее возможное (при $-1 \leq \rho \leq 1$) значение $\text{Corr}(Y_1, Y_2)$ всегда будет больше -1 . Например, если $\sigma = 1$, то $\text{Corr}(Y_1, Y_2) \in [-0.368, 1]$. Кроме того, интервал достижимых значений $\text{Corr}(Y_1, Y_2)$ будет тем меньше, чем больше σ .

В качестве другого примера, иллюстрирующего недостатки коэффициента корреляции, рассмотрим величину X_1 со стандартным нормальным распределением $N(0, 1)$ и $X_2 = X_1^2$. Тогда

$$\text{Cov}(X_1, X_2) = E(X_1 \cdot (X_1^2 - 1)) = E(X_1^3) - E(X_1) = 0.$$

Таким образом, с одной стороны, имеем функциональную зависимость между X_1 и X_2 , а с другой стороны, нулевую корреляцию между ними.

5.2. Коэффициенты ранговой корреляции: ρ -Спирмена и τ -Кендалла

Для того чтобы преодолеть указанные выше недостатки, необходимо обратиться к идее, возникшей в теории непараметрических статистик, которая предлагает сосредоточить внимание не на самих данных, а на их рангах. В результате были предложены две важные меры зависимости: ρ -Спирмена и τ -Кендалла (определение и основные свойства этих мер пар-

ной зависимости см., например, в (Айвазян, 2010, с. 96–107)). Перед тем, как приступить к обсуждению этих мер зависимости, коротко представим понятия согласованности и несогласованности.

Определение 5. Наблюдения (x_i, y_i) и (x_j, y_j) называют согласованными, если $x_i < x_j$ и $y_i < y_j$, или если $x_i > x_j$ и $y_i > y_j$. Аналогично, (x_i, y_i) и (x_j, y_j) называют несогласованными наблюдениями, если $x_i < x_j$ и $y_i > y_j$ или если $x_i > x_j$ и $y_i < y_j$. Это эквивалентно тому, что (x_i, y_i) и (x_j, y_j) являются согласованными наблюдениями, если $(x_i - x_j)(y_i - y_j) > 0$, и несогласованными, если $(x_i - x_j)(y_i - y_j) < 0$.

Другими словами, согласованность наблюдений возникает в случае, если большие значения одного наблюдения соответствуют большим значениям другого наблюдения, а малые значения также соответствуют малым значениям. Если это не так, то наблюдения называют несогласованными.

Коэффициент ранговой корреляции Спирмена (ρ -Спирмена)

Коэффициент ρ -Спирмена — мера согласованности между двумя случайными величинами, основанная на использовании понятий согласованности и несогласованности.

Определение 6. Для двух непрерывных случайных величин X и Y , совместное распределение которых имеет копула-функцию $C(\cdot)$, ρ -Спирмена определяется как

$$\rho_s(X, Y) = 3Q(C, P) = 12 \int \int uv dC(u, v) - 3 = 12 \int \int C(u, v) du dv - 3, \tag{19}$$

где $u = F_X(x)$, $v = F_Y(y)$, а $Q(\cdot)$ — разность между вероятностями согласованности и несогласованности (для более детальной информации см. (Nelsen, 2006)).

Эквивалентное выражение для ρ -Спирмена может быть дано в форме

$$\rho_s(X, Y) = \text{Corr}(F(X), G(Y)),$$

где корреляция (Corr) понимается в классическом смысле, а F и G — функции частных распределений случайных величин X и Y . Этот результат вытекает из того факта, что ранги u и v являются наблюдениями из равномерно распределенных на $[0; 1]$ случайных величин $U = F(X)$ и $V = G(Y)$, совместное распределение которых определяется копула-функцией C .

Поскольку U и V имеют среднее $1/2$ и дисперсию $1/12$, выражение для $\rho_s(X, Y)$ в (18) может быть представлено в следующем виде:

$$\rho_s(X, Y) = 12 \int \int uv dC(u, v) - 3 = 12E(UV) - 3 = \frac{E(UV) - 1/4}{1/12} = \frac{E(UV) - E(U)E(V)}{\sqrt{\text{Var}(U) \text{Var}(V)}}.$$

Таким образом, ρ -Спирмена для пары непрерывных случайных величин X и Y равно корреляции между рангами X и Y .

В многомерном случае ρ -Спирмена (*матрица*) определяется как

$$\rho_s(\mathbf{X}) = \text{Corr}(F_1(X_1), \dots, F_n(X_n)),$$

где $\text{Corr}(\cdot)$ — корреляционная матрица; компоненты матрицы $\rho_S(\mathbf{X})$ определяются как

$$\rho_S(\mathbf{X})_{ij} = \text{Corr}(F_i(X_i), F_j(X_j)).$$

Перечислим основные свойства коэффициента ρ -Спирмена.

1. ρ_S симметричен.
2. $|\rho_S| \leq 1$. В частности, для $Y = T(X)$, где $T(\cdot)$ — строго возрастающая функция, $\rho_S(X, Y) = 1$; а для $Y = T(X)$, где $T(\cdot)$ — строго убывающая функция, $\rho_S(X, Y) = -1$.
3. $\rho_S(T_x(X), T_y(Y)) = \rho_S(X, Y)$ для любых строго возрастающих отображений $T_x(\cdot)$ и $T_y(\cdot)$.
4. Если X и Y независимы, то $\rho_S(X, Y) = 0$.
5. Если (X, Y) имеет нормальное распределение с коэффициентом корреляции ρ и стандартными нормальными частными распределениями, то

$$\rho = 2 \sin[(\pi / 6)\rho_S] \quad \text{или} \quad \rho_S = \frac{6}{\pi} \arcsin\left(\frac{1}{2}\rho\right).$$

Последний результат справедлив также в случае, когда совместное распределение случайных величин X и Y имеет нормальную копула-функцию, а частные распределения непрерывны. К сожалению, соотношение между коэффициентом ρ -Спирмена и коэффициентом корреляции не выполняется для всех распределений из эллиптического семейства — см., например, (Hult, Lindskog, 2002; Nelsen 2006).

Можно показать, что непараметрическая оценка для ρ_S , построенная по наблюдениям (X_i, Y_i) , $i = 1, \dots, T$, имеет вид

$$\hat{\rho}_S = \frac{6}{T(T-1)(T-2)} \sum_{1 \leq i < j < k \leq T} 3 \cdot \text{sign}(X_i - X_j) \cdot \text{sign}(Y_i - Y_k)$$

и является несмещенной.

Если X_i и Y_i независимы, $i = 1, \dots, T$, то распределение $\sqrt{T} \hat{\rho}_S$ асимптотически сходится к $N(0, 1)$. Более детальную информацию см. в (Айвазян, 2010; Hollander, Wolfe, 1973; Statistics with confidence, 1989; Conover, 1999; van de Wiel, Bucchianico, 2001).

Коэффициент ранговой корреляции Кендалла (τ -Кендалла)

Коэффициент τ -Кендалла — это также мера связи между двумя случайными величинами, основанная на использовании понятий согласованности и рассогласованности и определяемая следующим образом.

Определение 7 (заимствовано из (Kruskal, 1958), см. также (Айвазян, 2010, с. 102–107)). Пусть (X_1, Y_1) — случайный вектор с распределением $H(x, y)$, а (X_2, Y_2) — случайный вектор, такой, что X_2 и Y_2 независимы, распределение X_1 совпадает с распределением X_2 , распределение Y_1 совпадает с распределением Y_2 . Коэффициент τ -Кендалла для вектора (X, Y) , компоненты которого есть непрерывные случайные величины, определяется как разность между вероятностью согласованности и вероятностью рассогласованности, т. е.

$$\begin{aligned} \tau(X, Y) &= P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\} = E[\text{sign}(X_1 - X_2)\text{sign}(Y_1 - Y_2)] \\ &= P\{X_1 > X_2, Y_1 > Y_2\} + P\{X_1 < X_2, Y_1 < Y_2\} - P\{X_1 < X_2, Y_1 > Y_2\} - P\{X_1 > X_2, Y_1 < Y_2\}. \end{aligned}$$

Тогда для n -мерного случайного вектора \mathbf{X} и случайного вектора $\tilde{\mathbf{X}}$, компоненты которого независимы, а их частные распределения равны частным распределениям соответствующих компонент вектора \mathbf{X} , матрица τ -Кендалла $\rho_\tau(\mathbf{X})$ определяется следующим образом:

$$\rho_\tau(\mathbf{X}) = \text{Cov}[\text{sign}(\mathbf{X} - \tilde{\mathbf{X}})],$$

где (i, j) -ый элемент этой матрицы определяется формулой

$$\rho_\tau(\mathbf{X})_{ij} = \text{Cov}[\text{sign}(X_i - \tilde{X}_i), \text{sign}(X_j - \tilde{X}_j)].$$

В частности, в двумерном случае ($n = 2$) матрицей τ -Кендалла будет

$$\rho_\tau(\mathbf{X}) = \begin{pmatrix} 1 & \tau(X, Y) \\ \tau(X, Y) & 1 \end{pmatrix}.$$

Следует отметить, что существуют некоторые другие n -мерные обобщения τ -Кендалла. Более детальную информацию можно найти, например, в (Clemen, Jouini, 1996) и (Barbe et al., 1996). Основные свойства этой меры зависимости, а также соответствующие доказательства представлены в работах (Embrechts et al., 2002) и (Nelsen 2006).

Как правило, если τ -Кендалла положителен, то с большой вероятностью мы имеем положительную зависимость. Если этот коэффициент отрицателен, то следует ожидать отрицательную зависимость. Кроме того, τ -Кендалла может быть представлен в терминах копула-функции, что облегчает вычислительную работу.

Теорема 5. Коэффициент τ -Кендалла для двумерной случайной величины, характеризующейся копулой $C(u, v)$, определяется формулой

$$\tau(X, Y) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

Доказательство см. в (Nelsen, 1999, p. 127–129).

Основные свойства коэффициента τ -Кендалла:

1. $\tau(X, Y)$ симметричен.
2. $-1 \leq \tau(X, Y) \leq 1$.
3. Если X и Y независимы, то $\tau(X, Y) = 0$.
4. $\tau(T_x(X), T_y(Y)) = \tau(X, Y)$ для любых строго возрастающих отображений $T_x(\cdot)$ и $T_y(\cdot)$.

Оценка коэффициента τ -Кендалла требует вычисления двойного интеграла, что для распределений эллиптического типа представляет собой довольно непростую задачу. Однако можно показать (см. (Lindskog et al., 2003)), что τ -Кендалла для эллиптических распределений определяется следующим образом:

$$\tau(X, Y) = \frac{2}{\pi} \arcsin \rho,$$

где ρ — обычный парный коэффициент корреляции.

Для архимедовых копула-функций ситуация значительно проще, т. к. $\tau(X, Y)$ может быть оценен с использованием генератора копула-функции. А именно, в работе (Genest, MacKay,

1986) показано, что τ -Кендалла для архимедовых копула-функций определяется соотношением:

$$\tau(X, Y) = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt,$$

где $\varphi(\cdot)$ — генератор копула-функции. Например, можно показать, что для копула-функций Клейтона или Гумбеля мы имеем следующие результаты:

- для копула-функции Клейтона: $\tau(X, Y) = \frac{\alpha}{\alpha + 2}$;
- для копула-функции Гумбеля: $\tau(X, Y) = 1 - \frac{1}{\alpha}$,

где α — параметр копула-функции.

Можно показать, что непараметрическая оценка для τ -Кендалла, построенная по наблюдениям (X_i, Y_i) , $i = 1, \dots, T$ и имеющая вид

$$\hat{\tau} = \frac{2}{T(T-1)} \sum_{1 \leq i < j \leq T} \text{sign}(X_i - X_j) \text{sign}(Y_i - Y_j)$$

является несмещенной.

Если X_i и Y_i независимы, то распределение случайной величины $\sqrt{T} \hat{\tau}$ сходится к $N(0, 4/9)$. Более подробная информация содержится, например, в (Ferguson et al., 2000).

5.3. Хвостовая зависимость

Определим понятия верхней и нижней хвостовой зависимости (Джозе, 1997).

Определение 8. Пусть (X, Y) — случайный вектор, компоненты которого есть непрерывные случайные величины с частными распределениями F_X и F_Y . Тогда коэффициент λ^U верхней хвостовой зависимости X и Y определяется соотношением:

$$\lambda^U = \lim_{u \rightarrow 1} P[Y > F_Y^{-1}(u) | X > F_X^{-1}(u)] = \lim_{u \rightarrow 1} P[X > F_X^{-1}(u) | Y > F_Y^{-1}(u)] = \lim_{u \rightarrow 1} \frac{1 - 2u + C(u, u)}{1 - u}$$

при условии, что пределы существуют, где $C(u, v)$ — двумерная копула-функция случайного вектора (X, Y) .

При этом случайные величины X и Y называют асимптотически зависимыми на верхнем хвосте, если $\lambda^U \in (0, 1]$, X и Y называют асимптотически независимыми, если $\lambda^U = 0$.

Другими словами, верхняя хвостовая зависимость существует тогда, когда имеется положительная вероятность одновременного возникновения положительные выбросов. λ^U широко используется в теории экстремальных значений и представляет собой вероятность того, что одна переменная примет экстремальные значения при условии, что другая переменная также принимает экстремальные значения. Коэффициент λ^U можно рассматривать как меру зависимости квантилей — см., например, (Coles et al., 1999).

Аналогично можно определить нижнюю хвостовую зависимость.

Определение 9. Пусть (X, Y) — двумерный случайный вектор, компонентами которого являются непрерывные случайные величины, частные распределения которых равны F_X и F_Y . Тогда коэффициент λ^L нижней хвостовой зависимости между X и Y равен:

$$\lambda^L = \lim_{u \rightarrow 0} P[Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u)] = \lim_{u \rightarrow 0} P[X \leq F_X^{-1}(u) | Y \leq F_Y^{-1}(u)] = \lim_{u \rightarrow 0} \frac{C(u, u)}{u}$$

при условии, что пределы существуют.

Случайные величины X и Y называются асимптотически зависимыми на нижнем хвосте, если $\lambda^L \in (0, 1]$ и асимптотически независимыми на нижнем хвосте, если $\lambda^L = 0$.

Таким образом, нижняя хвостовая зависимость существует тогда, когда существует положительная вероятность возникающих одновременно отрицательных выбросов.

Поскольку эллиптические распределения являются радиально симметричными, для них коэффициенты верхней и нижней хвостовой зависимости будут равны между собой. Например, для нормальной копула-функции коэффициенты верхней и нижней хвостовой зависимости определяются соотношением

$$\lambda^L = \lambda^U = 2 \lim_{x \rightarrow \infty} [P(Y > x | X = x)] = 2 \lim_{x \rightarrow \infty} \left[1 - \Phi \left(x \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) \right] = 0.$$

Таким образом, вне зависимости от величины парного коэффициента корреляции ρ ($\rho \neq 1$), если рассматривать экстремально большие положительные значения X и Y на правом хвосте, то для них получим асимптотическую независимость. Этот факт хорошо известен — см., например, (Sibuya, 1961; Resnick, 1987).

Для копула-функции Стьюдента коэффициенты верхней и нижней хвостовой зависимости также совпадают и равны

$$\lambda^L = \lambda^U = 2 \lim_{x \rightarrow \infty} [P(Y > x | X = x)] = 2 - 2t_{\nu+1} \left(\sqrt{\nu+1} \cdot \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right).$$

Очевидно, они возрастают по ρ и убывают по ν . В этой формуле через $t_{\nu+1}$, как и ранее, обозначена одномерная функция распределения Стьюдента с $\nu+1$ степенями свободы. Если число степеней свободы устремить к бесконечности, то λ^U будет стремиться к 0 при $\rho < 1$.

Рассмотрим коэффициенты λ^L, λ^U для архимедовых копула-функций. Так, для копула-функции Гумбеля лишь верхняя хвостовая зависимость положительна: $\lambda^U = 2 - 2^{1/\alpha}$, а $\lambda^L = 0$; тогда как для копула-функции Клейтона положительна лишь нижняя хвостовая зависимость: $\lambda^U = 0, \lambda^L = 2^{-1/\alpha}$.

Следует отметить, что хвостовая зависимость для архимедовых копула-функций может быть представлена в терминах генераторов, обратных к ним функций и первых производных, см. (Joe, 1997; Nelsen, 2006). В работах (Coles et al., 1999; Poon et al., 2004) представлены робастные непараметрические оценки для коэффициентов хвостовой зависимости.

5.4. Эмпирические приложения в статистическом пакете *R*: меры зависимости

В пакете *R* использование модулей *copula* и *fcopulae* дает возможность вычислять упомянутые выше меры зависимости для множества типов копула-функций. Ниже представлены некоторые примеры (заинтересованным читателям рекомендуется ознакомиться с руководством к пакету *R*).

```
# EXAMPLES WITH ARCHIMEDEAN COPULAS AND THE "copula" PACKAGE
gumbel.cop <- gumbelCopula(3)
kendallsTau(gumbel.cop)
[1] 0.6666667
spearmanRho(gumbel.cop)
[1] 0.848167
tailIndex(gumbel.cop)
lower upper
0.000000 0.740079
# let us compute the sample versions
x <- rcopula(gumbel.cop, 200)
cor(x, method = "kendall")

[,1] [,2]
[1,] 1.0000000 0.6592965
[2,] 0.6592965 1.0000000
cor(x, method = "spearman")

[,1] [,2]
[1,] 1.0000000 0.8483507
[2,] 0.8483507 1.0000000
# compare with the true parameter value 3
calibKendallsTau(gumbel.cop, cor(x, method="kendall")[1,2])
[1] 2.935103
calibSpearmanRho(gumbel.cop, cor(x, method="spearman")[1,2])
[1] 3.001893
calibSpearmanRho(gumbel.cop, cor(x, method="spearman")[1,2])
[1] 3.001893
# EXAMPLES WITH ARCHIMEDEAN COPULAS AND THE "fcopulae" PACKAGE
# ellipticalTau Computes Kendall's tau for elliptical copulae
# ellipticalRho computes Spearman's rho for elliptical copulae
ellipticalTau(rho = -0.5)

Tau
-0.3333333 attr(,"control")

rho
-0.5
```

```
ellipticalRho(rho = 0.75, type = "t", subdivisions = 100)

Rho
0.71
attr(,"control")

rho   type   tau

"0.75"  "t"   "0.5399"
# ellipticalTailCoeff
# Student-t Tail Coefficient:
ellipticalTailCoeff(rho = 0.25, param = 3, type = "t")

lambda
0.1962612
attr(,"control")

rho   type   param.nu

"0.25"  "t"   "3"
```

6. Процедуры оценивания: параметрические методы

Пусть x_{1t}, \dots, x_{nt} , $t = 1, \dots, T$ — многомерные наблюдения, где n — размерность наблюдаемой случайной величины, а T — число имеющихся наблюдений.

6.1. Метод максимального правдоподобия (ММП или одношаговый метод)

Пусть $f(\cdot)$ — плотность совместного распределения случайного вектора (X_1, X_2, \dots, X_n) .

$$f(\tilde{x}_1, \dots, \tilde{x}_n; \alpha_1, \dots, \alpha_n, \gamma) = c(F_1(\tilde{x}_1; \alpha_1), \dots, F_1(\tilde{x}_n; \alpha_n); \gamma) \cdot \prod_{i=1}^n f_i(\tilde{x}_i; \alpha_i),$$

где f_i — плотность одномерного частного распределения X_i , $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ — некоторые заданные (текущие) значения случайных величин X_1, X_2, \dots, X_n соответственно, а $c(\cdot)$ — плотность копула-функции, определяемая следующим соотношением:

$$c(u_1, \dots, u_n; \gamma) = \frac{\partial^n C(u_1, u_2, \dots, u_n; \gamma)}{\partial u_1 \partial u_2 \dots \partial u_n}.$$

Пусть $\theta = (\alpha_1, \dots, \alpha_n; \gamma)$ — вектор оцениваемых параметров, где $\alpha_i, i = 1, \dots, n$ — параметры частных распределений F_i , γ — вектор параметров копула-функции. Логарифмическая функция правдоподобия имеет следующий вид:

$$l(\theta) = \sum_{t=1}^T \ln(c(F_1(x_{1,t}; \alpha_1), \dots, F_n(x_{n,t}; \alpha_n); \gamma)) + \sum_{t=1}^T \sum_{i=1}^n \ln f_i(x_{i,t}; \alpha_{i,t}).$$

Оценка максимального правдоподобия $\hat{\theta}_{ML}$ для параметров определяется как

$$\hat{\theta}_{ML} = \arg \max_{\theta} l(\theta).$$

Ниже приводится результат, описывающий асимптотические свойства оценки максимального правдоподобия.

Теорема 6 (состоятельность и асимптотическое распределение ММП-оценок). Пусть $\hat{\theta}_{ML}$ — ММП-оценка. Тогда при выполнении некоторых стандартных условий регулярности, см. (White, 1994), справедливо следующее: $\hat{\theta}_{ML} \xrightarrow{P} \theta_0$, где $\theta_0 = (\alpha_1, \dots, \alpha_n, \gamma)'$. Кроме того,

$$\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0)),$$

где $I(\theta_0)$ — информационная матрица Фишера.

Доказательство см. в (White, 1994) или (Gourieroux, Monfort, 1995).

6.2. Двухшаговый метод максимального правдоподобия³

В соответствии с этим методом параметры частных распределений оцениваются независимо от параметров копула-функции. Другими словами, процедура оценивания разделена на следующие два шага:

Шаг 1: Используя ММП, оцениваются параметры $\alpha_i, i = 1, \dots, n$, частных одномерных распределений $F_i(\cdot)$:

$$\hat{\alpha}_i = \arg \max_{\alpha} l^i(\alpha) = \arg \max_{\alpha} \sum_{t=1}^T \ln f_i(x_{i,t}; \alpha),$$

где l^i — логарифмическая функция правдоподобия для частного распределения $F_i(\cdot)$.

Шаг 2: Оцениваются параметры γ копула-функции при заданных значениях оценок, полученные на шаге 1:

$$\hat{\gamma} = \arg \max_{\gamma} l^c(\gamma) = \arg \max_{\gamma} \sum_{t=1}^T \ln(c(F_1(x_{1,t}; \hat{\alpha}_1), \dots, F_n(x_{n,t}; \hat{\alpha}_n); \gamma)),$$

где l^c — логарифмическая функция правдоподобия для копула-функции.

³ В англоязычном оригинале этот метод назван The Inference Functions for Margins method (IFM-method).

Ниже (см. теоремы 7–9) описаны асимптотические свойства оценок, получаемых IFM-методом. Эти результаты представляют собой простые обобщения результатов, полученных для ММП-процедуры, см. (Newey, McFadden, 1994; White, 1994; Patton, 2006a, b).

Представленные ниже результаты получены в предположении некоторых стандартных условий регулярности, которые мы не будем конкретизировать, заинтересованный читатель может ознакомиться с ними в работе (White, 1994). Из соображений простоты рассматривается двумерный случай. Отметим, что обобщения на многомерный случай получаются достаточно просто.

Теорема 7 (состоятельность $\hat{\alpha}_1$ и $\hat{\alpha}_2$). Если $n \rightarrow \infty$, то $\hat{\alpha}_1 \xrightarrow{P} \alpha_1$ и $\hat{\alpha}_2 \xrightarrow{P} \alpha_2$. Доказательство см. в (White, 1994, Theorem 3.13).

Теорема 8 (состоятельность параметра γ). Пусть функция

$$n^{-1} \sum_{t=1}^T \ln c(F_1(x_{1,t}; \hat{\alpha}_1), F_2(x_{2,t}; \hat{\alpha}_2); \gamma)$$

имеет единственный максимум в точке $\gamma_0 \in \text{int}(\Gamma)$. Тогда $\hat{\gamma} \xrightarrow{P} \gamma_0$ при $n \rightarrow \infty$, где Γ — область допустимых значений параметра γ , а $\text{int}(\Gamma)$ — внутренность этого множества.

Доказательство см. в (White, 1994, Theorem 3.10).

Перед формулировкой теоремы 9 введем ряд обозначений. Определим штрафы для оценки $\hat{\theta}_{IFM} = (\hat{\alpha}_1, \hat{\alpha}_2, \hat{\gamma})'$:

$$s_{1t} = \frac{\partial}{\partial \alpha_1} \ln f_1(x_{1,t}; \alpha_1), \quad s_{2t} = \frac{\partial}{\partial \alpha_2} \ln f_2(x_{2,t}; \alpha_2), \quad s_{3t} = \frac{\partial}{\partial \gamma} \ln c(F_1(x_{1,t}; \alpha_1), F_2(x_{2,t}; \alpha_2); \gamma).$$

Матрица Гесса (гессиян) $Hess(\theta)$ для штрафов равна:

$$\begin{pmatrix} T^{-1} \sum_{t=1}^T \left(\frac{\partial^2 \ln f_1(x_{1,t}; \alpha_1)}{\partial \alpha_1^2} \right) & 0 & 0 \\ 0 & T^{-1} \sum_{t=1}^T \left(\frac{\partial^2 \ln f_2(x_{2,t}; \alpha_2)}{\partial \alpha_2^2} \right) & 0 \\ T^{-1} \sum_{t=1}^T \left(\frac{\partial^2 \ln c(F_1(\cdot), F_2(\cdot); \gamma)}{\partial \alpha_1 \partial \gamma} \right) & T^{-1} \sum_{t=1}^T \left(\frac{\partial^2 \ln c(F_1(\cdot), F_2(\cdot); \gamma)}{\partial \alpha_2^2} \right) & T^{-1} \sum_{t=1}^T \left(\frac{\partial^2 \ln c(F_1(\cdot), F_2(\cdot); \gamma)}{\partial \gamma^2} \right) \end{pmatrix}.$$

Матрица OPG определяется как

$$\begin{pmatrix} T^{-1} \sum_{t=1}^T s_{1t} \cdot s'_{1t} & T^{-1} \sum_{t=1}^T s_{1t} \cdot s'_{2t} & T^{-1} \sum_{t=1}^T s_{1t} \cdot s'_{3t} \\ T^{-1} \sum_{t=1}^T s_{2t} \cdot s'_{1t} & T^{-1} \sum_{t=1}^T s_{2t} \cdot s'_{2t} & T^{-1} \sum_{t=1}^T s_{2t} \cdot s'_{3t} \\ T^{-1} \sum_{t=1}^T s_{3t} \cdot s'_{1t} & T^{-1} \sum_{t=1}^T s_{3t} \cdot s'_{2t} & T^{-1} \sum_{t=1}^T s_{3t} \cdot s'_{3t} \end{pmatrix}.$$

Теорема 9 (асимптотика оценок, получаемых IFM-методом). Для IFM-оценок выполняется свойство асимптотической нормальности, а именно

$$\sqrt{T}(\hat{\theta}_{IFM} - \theta_0) \xrightarrow{d} N\left(0, H_0^{-1}B_0(H_0^{-1})'\right),$$

где $H_0 = \mathbf{E}[Hess(\theta_0)]$, $B_0 = \mathbf{E}[OPG(\theta_0)]$, т. е. H_0 и B_0 — средние значения для гессиана штрафов и матрицы OPG. Матрица $H_0^{-1}B_0(H_0^{-1})'$ известна в теории копула-функций как информационная матрица Годамби (Joe, Xu, 1996; Joe, 1997).

Из теорем 7–9 вытекает следующий весьма интересный факт.

Равная эффективность IFM-метода и ММП. Пусть x_{1t} и x_{2t} независимы (т. е. совместное распределение (x_{1t}, x_{2t}) имеет копула-функцию $C(u_1, u_2; \gamma) = u_1 \cdot u_2$, плотность которой $c(u_1, u_2; \gamma) = 1$), и для плотностей частных распределений выполнено информационное равенство. Тогда IFM-метод и ММП дают оценки, имеющие одинаковые асимптотические распределения.

6.3. Эмпирические приложения в статистическом пакете R: процедуры параметрического оценивания

Модуль `copula` в статистическом пакете R позволяет получать как ММП-, так и IFM-оценки. Ниже представлен пример, взятый из работы (Yan, 2007). Вначале генерируется выборка из двумерного распределения, частные распределения которого — гамма-распределения с параметрами $\theta_\gamma = (2, 1, 3, 2)$, а копула-функция — нормальная копула-функция с параметром $\alpha = 0.5$; затем по сгенерированной выборке строятся ММП- и IFM-оценки параметров.

```
# One step EML example from Yan (2007)
myMvd <- mvdc(copula=ellipCopula(family="normal", param=0.5),

margins=c("gamma", "gamma"),

paramMargins=list(list(shape=2, scale=1),

list(shape=3, scale=2)))
n <- 200
dat <- rmvdc(myMvd, n)
loglikMvdc(c(2,1,3,2,0.5), dat, myMvd)
[1] -778.3114
mm <- apply(dat, 2, mean)
vv <- apply(dat, 2, var)
b1.0 <- c(mm[1]^2 / vv[1], vv[1] / mm[1])
b2.0 <- c(mm[2]^2 / vv[2], vv[2] / mm[2])
a.0 <- sin(cor(dat[,1], dat[,2], method = "kendall") * pi / 2)
```

```
start <- c(b1.0, b2.0, a.0)
fit.eml <- fitMvdc(dat, myMvd, start=start, optim.control=list(trace=TRUE,
maxit=2000))
fit.eml
The Maximum Likelihood estimation is based on 200 observations.
Margin 1 :

Estimate Std. Error
m1.shape 2.1279360 0.19842701
m1.scale 0.9246803 0.09716286
Margin 2 :

Estimate Std. Error
m2.shape 3.213331 0.3061706
m2.scale 1.778994 0.1835066
Copula:

Estimate Std. Error
rho.1 0.4671964 0.0552821
The maximized loglikelihood is -777.1381
The convergence code is 0 see ?optim.
# IFM method
loglik.marg <- function(b, x) sum(dgamma(x, shape=b[1], scale=b[2],
log=TRUE))
ctrl <- list(fnscale = -1)
b1hat <- optim(b1.0, fn=loglik.marg, x=dat[,1], control=ctrl)$par
b2hat <- optim(b2.0, fn=loglik.marg, x=dat[,2], control=ctrl)$par
udat <- cbind(pgamma(dat[,1], shape=b1hat[1], scale=b1hat[2]),
pgamma(dat[,2], shape=b2hat[1], scale=b2hat[2]))
fit.ifm <- fitCopula(myMvd, udat, method="ml",start=a.0)
c(b1hat, b2hat, fit.ifl@estimate) fit.ifl
[1] 2.3494821 0.9395788 3.1645483 2.0540858 0.4782428
fit.ifm
The estimation method is Maximum Likelihood based on 200 observations.

Estimate Std. Error z value Pr(>|z|)
rho.1 0.4782428 0.04920972 9.718461 0
The maximized loglikelihood is 25.95508
The convergence code is 0
```

7. Процедуры оценивания: полупараметрические и непараметрические методы

7.1. Канонический метод максимума правдоподобия (КММП)⁴

КММП отличается от описанных выше методов тем, что в нем не делается никаких предположений относительно параметрической формы частных распределений. Процесс оценивания в КММП состоит из следующих двух шагов.

1. Преобразуем наблюдения $(x_{1t}, x_{2t}, \dots, x_{nt})$, $t = 1, \dots, T$ следующим образом:

$$\hat{u}_{it} = F_{i,T}(x_{it}),$$

где $F_{i,T}(\cdot) = \frac{1}{T+1} \sum_{i=1}^T \mathbf{1}_{\{x_{it} \leq \cdot\}}$, $i = 1, \dots, n$, а $\mathbf{1}_{\{\cdot\}}$ — индикаторная функция. Другими словами,

$F_{i,T}(x)$ — это непараметрическая оценка частной функции распределения переменной X_i .

2. Оценим параметры копула-функции, максимизируя логарифмическую функцию правдоподобия⁵:

$$\hat{\gamma}_{CML} = \arg \max_{\gamma} \sum_{t=1}^T \ln c(F_{1,T}(x_{1t}), \dots, F_{n,T}(x_{nt}); \gamma).$$

Для простоты при описании асимптотических свойств КММП-оценки снова будем рассматривать только двумерный случай. Пусть $l(u_1, u_2, \gamma) = \ln c(u_1, u_2; \gamma)$. Далее индексами 1, 2, γ будем обозначать частные производные функции $l(u_1, u_2, \gamma)$ относительно u_1, u_2, γ соответственно. Кроме того, положим

$$S_T = \frac{1}{T} \sum_{t=1}^T l_{\gamma} (F_{1,T}(x_{1t}), F_{2,T}(x_{2t}); \gamma), \quad H_T = \frac{1}{T} \sum_{t=1}^T l_{\gamma\gamma} (F_{1,T}(x_{1t}), F_{2,T}(x_{2t}); \gamma).$$

Поскольку $\hat{\gamma}_{CML}$ является решением следующего дифференциального уравнения:

$$S_T = \frac{1}{T} \sum_{t=1}^T l_{\gamma} (F_{1,T}(x_{1t}), F_{2,T}(x_{2t}); \gamma) = 0,$$

то, используя разложение функции S_T в ряд Тейлора, получим:

$$0 = \frac{1}{T} \sum_{t=1}^T l_{\gamma} (F_{1,T}(x_{1t}), F_{2,T}(x_{2t}); \gamma) \Big|_{\gamma=\hat{\gamma}_{CML}} \approx S_T + H_T (\hat{\gamma}_{CML} - \gamma),$$

и, следовательно,

$$\sqrt{T} (\hat{\gamma}_{CML} - \gamma) \approx \sqrt{T} \frac{S_T}{-H_T}.$$

⁴ В англоязычном оригинале *The Canonical Maximum Likelihood (CML)*.

⁵ В дальнейшем предполагается, что анализируемое семейство копул зависит от единственного параметра g .

Из работ (Ruymgaart et al., 1972; Genest et al., 1995; Ruud, 2000) отметим, в частности, следующие результаты, справедливые при соблюдении некоторых условий регулярности:

а) $\hat{\gamma}_{CML} \rightarrow \gamma_0$ (п. н.) при $T \rightarrow \infty$;

б) $\sqrt{T}(\hat{\gamma}_{CML} - \gamma_0) \xrightarrow{d} N(0, \sigma^2 h^{-2})$, где

$$\sigma^2 = \text{Var} \left[l_{\gamma} (F_1(x_{1t}), F_2(x_{2t}); \gamma) + W_1(x_{1t}) + W_2(x_{2t}) \right], \quad h = E \left[l_{\gamma} (F_1(x_{1t}), F_2(x_{2t}); \gamma) \right];$$

выражение для σ^2 дано здесь для *двумерного* случая ($n = 2$ и, соответственно, $i = 1, 2$), а $W_i(x_{it})$ определяются соотношениями:

$$W_i(x_{it}) = - \int \mathbf{1}_{\{F_i(x_{it}) \leq u_i\}} l_{\gamma i} (u_{1t}, u_{2t}; \gamma) c(u_{1t}, u_{2t}; \gamma) du_1 du_2;$$

в) полупараметрическая оценка $\hat{\gamma}_{CML}$ менее эффективна, чем оценка, полученная с помощью одношагового метода максимального правдоподобия $\hat{\gamma}_{ML}$ (за исключением случая, когда x_{1t} и x_{2t} статистически независимы).

7.2. Трехшаговый канонический метод максимального правдоподобия (КМЕ–СМЛ метод)⁶

Начиная с момента выхода работы (Genest et al., 1995), использование полупараметрических методов для копула-функций эллиптического типа стало обычной практикой, см., например, (Cherubini et al., 2004; McNeil et al., 2005). В частности, для копула-функций Стьюдента, после того, как на первом шаге получены непараметрические оценки частных функций распределения, можно получить оценку корреляционной матрицы, вычисленную с использованием метода моментов и оценок ранговых коэффициентов корреляции τ -Кендалла. При этом степени свободы оцениваются с использованием метода максимального правдоподобия. Отметим, что оценка частных распределений на первом шаге не является необходимым условием для оценки корреляционной матрицы, поскольку оценка τ -Кендалла основана на количестве согласованных наблюдений, которое не зависит от монотонных преобразований. Однако оценивание частных распределений является необходимым при оценке числа степеней свободы.

Ниже дается описание трехшагового КМЕ–СМЛ метода (см. также (Bouyé et al., 2000; McNeil et al., 2005; Fantazzini, 2010)).

Трехшаговый КМЕ–СМЛ метод оценивания параметров копула-функций

1. Преобразуем $(x_{1t}, x_{2t}, \dots, x_{nt})$ к нормированным рангам $(F_{1T}(x_{1t}), F_{2T}(x_{2t}), \dots, F_{nT}(x_{nt}))$, используя эмпирическую функцию распределения для каждой компоненты вектора наблюдений (здесь и далее n — размерность анализируемой случайной величины, T — общее число ее наблюдений).

2. Для всех пар (j, k) компонент вектора наблюдений оценим τ -Кендалла:

$$\hat{R}_{jk}^{\tau} = \tau \left[F_{jT}(X_j), F_{kT}(X_k) \right] = (C_T^2)^{-1} \sum_{1 \leq t < s \leq T} \text{sign}((x_{1t} - x_{1s})(x_{2t} - x_{2s})).$$

⁶ В англоязычном оригинале метод называется *Kendall-t Moment — Estimator of Canonical Maximum Likelihood Method*.

Возьмем оценку корреляции для (j, k) компонент вектора наблюдений $\hat{\Sigma}_{jk} = \sin(\frac{\pi}{2} \hat{R}_{jk}^{\tau})$. Поскольку покомпонентное преобразование, осуществленное на первом шаге, не гарантирует положительной определенности матрицы $\hat{\Sigma} = (\hat{\Sigma}_{jk})$, то для ее обеспечения обычно вносятся некоторые поправки (Rousseeuw, Molenberghs, 1993).

3. Находим оценку степеней свободы $\hat{\nu}_{CML}$ копула-функций Стьюдента с использованием метода максимального правдоподобия:

$$\hat{\nu}_{CML} = \arg \max_{\nu} \sum_{t=1}^T \ln c_{T-копула} (F_{1T}(x_{1t}), \dots, F_{nT}(x_{nt}); \hat{\Sigma}, \nu).$$

Второй шаг описанной процедуры реализует метод моментов, основанный на оценках $q = n(n-1)/2$ моментов и такого же количества коэффициентов τ -Кендалла, оцененных с использованием эмпирических распределений (эта оценка известна как моментная оценка τ -Кендалла). Следует еще раз отметить, что нет необходимости использовать оценку τ -Кендалла, полученную с использованием именно эмпирических распределений, т. к. τ -Кендалла сохраняет одно и то же значение при всех монотонных преобразованиях. Таким образом, можно взять q моментов от некоторой векторнозначной функции ψ , зависящей от корреляций $\theta_0 = (\rho_1, \dots, \rho_q)'$, а именно:

$$\psi(F_1, \dots, F_n; \theta_0) = \begin{pmatrix} E[\psi_1(F_1, F_2); \rho_1(\theta)] \\ \vdots \\ E[\psi_q(F_{n-1}, F_n); \rho_q(\theta)] \end{pmatrix}.$$

Тогда теоретические моментные тождества могут быть представлены в виде

$$\psi(F_1, \dots, F_n; \theta_0) = (0, 0, \dots, 0)', \tag{20}$$

где θ_0 — истинное значение оцениваемого векторного параметра. Соответственно, оценка $\hat{\theta}$ определится из (20) с заменой теоретических моментов их выборочными аналогами.

Для оценки, полученной с помощью метода КМЕ–CML, справедливы следующие теоремы (все доказательства представлены в приложении А работы (Fantazzini, 2010)).

Теорема 10 (состоятельность $\hat{\theta}$). *Предположим, что (x_{1t}, \dots, x_{nt}) — независимые (по t) наблюдения n -мерной случайной величины (X_1, \dots, X_n) со структурой зависимости, заданной плотностью копула-функции $c(u_{1,t}, \dots, u_{n,t}; \Sigma_0, \nu_0)$. Предположим, что:*

- а) пространство параметров Θ является компактным подмножеством \mathbb{R}^q ;
 - б) q -мерный вектор моментов $\psi(F_1(X_1), \dots, F_n(X_n); \theta_0)$ непрерывен относительно θ_0 для любых X_j ;
 - в) $\psi(F_1(X_1), \dots, F_n(X_n); \theta)$ измеримы относительно X_j для всех $\theta \in \Theta$;
 - г) $\psi(F_1(X_1), \dots, F_n(X_n); \theta) \neq 0$ для всех $\theta \neq \theta_0, \theta \in \Theta$;
 - д) $\sup_{\theta \in \Theta} \|\psi(F_1(X_1), \dots, F_n(X_n); \theta)\| < \infty$, где $\|A\|$ обозначает евклидову норму матрицы A .
- Тогда $\hat{\theta} \xrightarrow{P} \theta_0$ при $T \rightarrow \infty$.

Моделирование многомерных распределений с использованием копула-функций. II

При условиях теоремы 10, дополненных некоторыми условиями регулярности (Genest et al., 1995), доказана состоятельность оценки $\hat{\nu}_{CML}$ (т. е. $\hat{\nu}_{CML} \xrightarrow{P} \nu_0$ при $T \rightarrow \infty$).

Асимптотическая нормальность оценок, полученных методом КМЕ–СМЛ, не очевидна, т. к. используется трехшаговая процедура, при которой на втором и третьем шагах используются разные методы оценивания. Возможным решением является представление оценки на третьем шаге КММП, как оценки специального вида, полученной методом моментов. Отметим, что КММП-оценка определяется путем приравнивания к нулю производной (относительно параметра ν) логарифма функции правдоподобия:

$$\frac{\partial l(\cdot; \nu)}{\partial \nu} = \sum_{i=1}^T l_{\nu} (F_{1T}(x_{1i}), \dots, F_{nT}(x_{ni}); \hat{\Sigma}, \hat{\nu}) = 0.$$

Разделив обе части на T , получим определение оценки по методу моментов:

$$\frac{1}{T} \sum_{i=1}^T l_{\nu} (F_{1T}(x_{1i}), \dots, F_{nT}(x_{ni}); \hat{\Sigma}, \hat{\nu}) = \frac{1}{T} \sum_{i=1}^T \psi_{\nu}(F_{1T}(x_{1i}), \dots, F_{nT}(x_{ni}); \hat{\Sigma}, \hat{\nu}) = 0.$$

Таким образом, КММП-оценка может быть представлена как оценка, получаемая методом моментов (ММ).

Оценка методом моментов — это то значение θ , которое обеспечивает равенство между выборочными моментами (выборочные средние) и их теоретическими аналогами. Поскольку число параметров равно q , то требуются q моментных уравнений с неизвестными $\hat{\theta}$ (более подробно см. (Greene, 2002)). Таким образом, можно использовать известные асимптотические результаты, справедливые для метода моментов.

Определим выборочные моменты вектора $\Psi_{KME-CML}$, зависящего от параметров $\hat{\Xi} = (\hat{\rho}_1, \dots, \hat{\rho}_q, \hat{\nu})'$:

$$\Psi_{KME-CML} (F_{1T}(x_{1i}), \dots, F_{nT}(x_{ni}); \hat{\Xi}) = \begin{pmatrix} \frac{1}{T} \sum_{i=1}^T \psi_1 (F_{1T}(x_{1i}), F_{2T}(x_{i2}); \hat{\rho}_1) \\ \dots \\ \frac{1}{T} \sum_{i=1}^T \psi_q (F_{(n-1)T}(x_{(n-1)i}), F_{nT}(x_{ni}); \hat{\rho}_q) \\ \frac{1}{T} \sum_{i=1}^T \psi_{\nu} (F_{1T}(x_{1i}), \dots, F_{nT}(x_{ni}); \hat{\Sigma}, \hat{\nu}) \end{pmatrix} = 0. \quad (21)$$

Для определения ковариационной матрицы предельного распределения оценки $\hat{\Xi}$ необходимо подкорректировать (21) с учетом непараметрических оценок частных распределений и их дисперсий, см. (Genest et al., 1995, §4), а именно

$$\Delta_0 = \begin{pmatrix} \psi_1 (F_1(X_1), F_2(X_2); \rho_1) \\ \vdots \\ \psi_q (F_{n-1}(X_{n-1}), F_n(X_n); \rho_q) \\ \psi_{\nu} (F_1(X_1), \dots, F_n(X_n); \Sigma_0, \nu_0) + \sum_{j=1}^n W_{j,\nu} (X_j) \end{pmatrix} = 0,$$

$$Y_0 \equiv \text{Var}[\Delta_0] = \mathbf{E}[\Delta_{KME-CML} \Delta'_{KME-CML}],$$

где

$$W_{j,v}(X_j) = \int 1_{\{F_j(X_j) \leq u_j\}} \frac{\partial^2}{\partial v \partial u_j} \ln c(u_1, \dots, u_n) dC(u_1, \dots, u_n).$$

С учетом вышеизложенного справедлива следующая теорема.

Теорема 11 (асимптотическое распределение КМЕ–СМЛ оценок). Пусть выполнены предположения предыдущей теоремы и дополнительные условия из (Genest et al., 1995).

Кроме того, предположим, что матрица $\frac{\partial \Psi_{\text{КМЕ-СМЛ}}(\cdot; \Xi)}{\partial \Xi'}$ имеет ограниченные (по T) элементы и является отрицательно определенной, тогда как Y_0 также имеет ограниченные (по T) элементы, но положительно определена.

Тогда при $T \rightarrow \infty$:

$$\sqrt{T}(\hat{\Xi} - \Xi_0) \xrightarrow{d} N \left\{ 0, \mathbf{E} \left[\frac{\partial \Psi_{\text{КМЕ-СМЛ}}}{\partial \Xi'} \right]^{-1} Y_0 \left(\mathbf{E} \left[\frac{\partial \Psi_{\text{КМЕ-СМЛ}}}{\partial \Xi'} \right]^{-1} \right)' \right\}. \quad (22)$$

Отметим, что последнее асимптотическое свойство выполняется и для многомерных моделей с гетероскедастичностью. Для таких моделей на первом этапе получают состоятельные оценки параметров одномерных частных распределений, а затем рассчитывают соответствующие оценки остатков (так называемые «невязки») которые используются для оценки совместного распределения многомерного вектора остатков при помощи копула-функций. Этот результат является прямым следствием теорем 1 и 2 из работы (Kim et al., 2008), которая, в свою очередь, использует работы (Koul, Ling, 2006; Koul, 2002). Отметим, что похожие результаты представлены также и в (Chen, Fan, 2006), но без доказательств. Соответствующий результат формулируется в теореме 12.

Теорема 12 (асимптотика КМЕ–СМЛ оценок для многомерных моделей с гетероскедастичностью). Предположим, что выполнены условия а)–д) теоремы 10, а также А.1–А.9 из (Kim et al., 2008). Тогда для КМЕ–СМЛ оценки имеет место асимптотическая сходимость (22).

Условия (А.1)–(А.4) справедливы для моделей достаточно общего вида. В частности, условия (А.1)–(А.2) требуют, чтобы плотность копула-функций имела непрерывные частные производные до третьего порядка включительно. Кроме того, они должны быть конечными вместе с их вторым моментом. Условие (А.3) — это техническое условие на частные производные, а условие (А.4) требует, чтобы были выполнены условия (А.1) из (Genest et al., 1995). Условия (А.5)–(А.8) также являются техническими и взяты из работ (Koul, 2002) и (Koul, Ling, 2006). Условие (А.9) не является слишком ограничительным. Например, если условное среднее является функцией исторических значений строго стационарных и эргодических временных рядов, то каждое слагаемое суммы является строго стационарным и эргодическим процессом с нулевым средним, а значит, условие (А.9) выполняется (см. (Taniguchi, Kakizawa; 2000, теоремы 1.3.3–1.3.5), а также (Kim et al., 2008)).

Свойства и вычислительные аспекты в условиях малых выборок

Поскольку основные свойства предлагаемого полупараметрического метода являются асимптотическими, то на примерах смоделированных выборок в работе (Fantazzini, 2010) было проведено исследование, показавшее, что в условиях малых выборок и больших зна-

чений ν численная максимизация логарифмической функции правдоподобия не сходилась значительно чаще при использовании КМЕ–СМЛ оценок, чем при использовании ММП оценок. В то время как ММП-оценка параметра попадала в 95% доверительный интервал примерно для 95% случаев, КМЕ–СМЛ оценка попадала в соответствующий доверительный интервал лишь в 30% случаев. Однако такое снижение доли случаев со сходимостью более значительно для двумерных копула-функций Стьюдента по сравнению с копула-функциями Стьюдента большей размерности, которые типичны для финансовых портфелей. Кроме того, как ММП, так и КМЕ–СМЛ метод показали весьма существенные значения средних и медиан смещений для оцененных корреляций, когда фактические значения последних принимают значения, близкие к нулю.

Наконец, в работе (Fantazzini, 2010) показано, что метод собственных значений, представленный в (Rousseeuw, Molenberghs, 1993), должен быть использован для получения положительно определенной корреляционной матрицы не только в условиях малых выборок ($T < 100$), но и в условиях, когда описываемому процессу отвечает наименьшее собственное значение, близкое к нулю. Эта поправка влечет положительное смещение оценки параметра ν , но ее влияние на сходимость при максимизации для получения параметра ν весьма ограничено.

Таким образом, предыдущие результаты показывают, что КМЕ–СМЛ метод может быть использован в случае малых выборок и при относительно небольших значениях числа степеней свободы, однако ММП является более привлекательной альтернативой. Возможной стратегией при оценке параметров является следующее: на первом этапе использовать КМЕ–СМЛ метод; если оценка числа степеней свободы больше 20, то следует использовать ММП-оценку, если последняя обеспечивает сходимость при максимизации соответствующей функции правдоподобия. Иначе, в качестве альтернативного решения следует использовать нормальную копула-функцию, к которой стремится копула-функция Стьюдента при $\nu \rightarrow \infty$ (нормальная и стьюдентовская копула-функции достаточно близки уже при $\nu > 20$).

7.3. Методы непараметрического оценивания

В работе (Fermanian, Scaillet, 2003) предлагается непараметрическая оценка для копула-функций многомерных стационарных процессов, удовлетворяющих условиям сильного перемешивания.

Перед тем, как обратиться к асимптотическим свойствам этой оценки, определим понятие ядерной функции.

Определение 10. *Ядерная функция — это действительная функция $K(u)$, $-\infty < u < +\infty$, которая удовлетворяет следующим условиям:*

- 1) симметричность: $K(u) = K(-u)$;
- 2) в точке $u = 0$ достигается максимум;
- 3) $\int_{-\infty}^{+\infty} K(u) du = 1$;
- 4) неотрицательность: $K(u) \geq 0$;
- 5) $K(u)$ ограничена;
- 6) $\int_{-\infty}^{\infty} u^2 K(u) du < \infty$.

Свойства 5 и 6 необходимы для вывода асимптотических свойств оценок, получаемых на основе ядерных функций. Возьмем n -мерное ядро и его кумулятивную функцию:

$$k(\mathbf{x}) = \prod_{j=1}^n k_j(x_j), \quad K(\mathbf{X}) = \prod_{j=1}^n \int_{-\infty}^{x_j} k_j(x) dx = \prod_{j=1}^n K_j(x_j),$$

где для простоты взяты произведения одномерных ядерных функций. Кроме того, возьмем

$$k(\mathbf{x}; \mathbf{h}) = \prod_{j=1}^n k_j\left(\frac{x_j}{h_j(T)}\right), \quad K(\mathbf{X}; \mathbf{h}) = \prod_{j=1}^n K_j\left(\frac{x_j}{h_j(T)}\right),$$

где $\mathbf{h}(T)$ — диагональная матрица с элементами $h_j(T)$, $j=1, \dots, n$ на главной диагонали и детерминантом $|\mathbf{h}(T)|$. При этом $h_j(T)$ положительна и $h_j \rightarrow 0$ при $T \rightarrow \infty$. $\mathbf{h}(T)$ обычно называют «шириной окна» ядерной функции.

Определение 11. Ядерная оценка $\hat{f}_j(\cdot)$ плотности $f_j(\cdot)$ распределения случайной величины Y_{jt} в точке y_j определяется как

$$\hat{f}_j(y_j) = \frac{1}{Th_j} \sum_{i=1}^T k_j\left(\frac{y_j - Y_{jt}}{h_j(T)}\right),$$

а ядерная оценка плотности случайного вектора \mathbf{Y}_t в точке $\mathbf{y} = (y_1, \dots, y_n)'$ равна

$$\hat{f}(\mathbf{y}) = \frac{1}{T|\mathbf{h}(T)|} \sum_{i=1}^T k(\mathbf{y} - \mathbf{Y}_i; \mathbf{h}(T)).$$

Аналогичным образом, ядерная оценка кумулятивной функции распределения случайной величины Y_{jt} в точке y_j равна

$$\hat{F}_j(y_j) = \int_{-\infty}^{y_j} \hat{f}_j(x) dx,$$

а ядерная оценка кумулятивной функции распределения случайного вектора \mathbf{Y}_t в точке \mathbf{y} равна:

$$\hat{F}(\mathbf{y}) = \int_{-\infty}^{y_1} \dots \int_{-\infty}^{y_n} \hat{f}(\mathbf{x}) d\mathbf{x}.$$

Например, для гауссовской ядерной функции $k_j(x) = \varphi(x)$ получаются следующие ядерные оценки функций распределения:

$$\hat{F}_j(y_j) = \frac{1}{T} \sum_{i=1}^T \Phi\left(\frac{y_j - Y_{jt}}{h_j}\right), \quad \hat{F}(\mathbf{y}) = \frac{1}{T} \sum_{i=1}^T \prod_{j=1}^n \Phi\left(\frac{y_j - Y_{jt}}{h_j}\right),$$

где через $\varphi(\cdot)$ и $\Phi(\cdot)$ обозначены соответственно плотность и функция распределения стандартного нормального распределения.

Определение 12 (ядерная оценка копула-функции). В силу следствия из теоремы Склара $C(u) = C(u_1, \dots, u_n) = H(F_1^{(-1)}(u_1), \dots, F_n^{(-1)}(u_n))$, где $H(\cdot)$ — многомерная функция распределения. Тогда ядерная оценка $\hat{C}(\mathbf{u})$ копула-функции $C(\mathbf{u})$ определяется следующим образом:

$$\hat{C}(\mathbf{u}) = \hat{H}(\hat{\xi}) = \hat{H}(\hat{\xi}_1, \dots, \hat{\xi}_n),$$

где $\hat{\xi}_j = \inf_{y \in \mathbb{R}} \{y: \hat{F}_j(y) \geq u_j\}$ соответствует ядерной оценке квантиля уровня u_j для функции распределения случайной величины Y_{ji} .

В работах (Fermanian, Scaillet, 2003; van der Vaart, Wellner, 1996) доказывается состоятельность и асимптотическая нормальность ядерных оценок $\hat{C}(\mathbf{u})$.

7.4. Эмпирические приложения в статистическом пакете R: процедуры полупараметрического оценивания

Продолжим рассмотрение примера, начатое в п. 6.3 (см. также (Фантаццини, 2011)). Можно оценить параметр зависимости α при помощи канонического метода максимума правдоподобия (КММП) без спецификации частных распределений.

```
# CML method
eu <- cbind((rank(dat[,1]) - 0.5)/n, (rank(dat[,2]) - 0.5)/n)
fit.cml <- fitCopula(myMvd, eu, method="mpl", start=a.0)
fit.cml
The estimation method is Maximum Pseudo-Likelihood based on 200
observations.
Estimate Std. Error z value Pr(>|z|)
rho.1 0.4806303 0.05066396 9.48663 0
The maximized loglikelihood is 26.02639
The convergence code is 0
```

Ниже представлены другие примеры, взятые из copula пакета.

```
# Example: Gumbel copula
gumbel.cop <- gumbelCopula(3, dim=2)
n <- 200 x <- rcopula(gumbel.cop, n) ## true observations
u <- apply(x, 2, rank) / (n + 1) ## pseudo-observations
## inverting Kendall's tau
fit.tau <- fitCopula(gumbel.cop, u, method="itau")
fit.tau
## inverting Spearman's rho
fit.rho <- fitCopula(gumbel.cop, u, method="irho")
fit.rho
```

```

## maximum pseudo-likelihood
fit.mpl <- fitCopula(gumbel.cop, u, method="mpl")
fit.mpl
## maximum likelihood
fit.ml <- fitCopula(gumbel.cop, x, method="ml")
fit.ml
## A multiparameter example
normal.cop <- normalCopula(c(0.6,0.36, 0.6),dim=3,dispstr="un")
x <- rcopula(normal.cop, n) ## true observations
u <- apply(x, 2, rank) / (n + 1) ## pseudo-observations
## inverting Kendall's tau
fit.tau <- fitCopula(normal.cop, u, method="itau")
fit.tau
## inverting Spearman's rho
fit.rho <- fitCopula(normal.cop, u, method="irho")
fit.rho
## maximum pseudo-likelihood
fit.mpl <- fitCopula(normal.cop, u, method="mpl")
fit.mpl
## maximum likelihood
fit.ml <- fitCopula(normal.cop, x, method="ml")
fit.ml

```

Список литературы

- Айвазян С. А. (2010). *Методы эконометрики*. М.: Магистр.
- Фантаццини Д. (2011). Моделирование многомерных распределений с использованием копула-функций. I. *Прикладная эконометрика*, 2 (22), 98–134.
- Aas K., Czado C., Frigessi A., Bakken H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44 (2), 182–198.
- Barbe P., Genest C., Ghoudi K., Rremillard B. (1996). On Kendall's process. *Journal of Multivariate Analysis*, 58, 197–229.
- Bedford T., Cooke R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32, 245–268.
- Bedford T., Cooke R. M. (2002). Vines — a new graphical model for dependent random variables. *Annals of Statistics*, 30, 1031–1068.
- Bouyé E., Durrleman V., Nikeghbali A., Riboulet G., Roncalli T. (2000). Copulas for finance: A reading guide and some applications. Groupe de Recherche Operationnelle, Credit Lyonnais, *Working Paper*.
- Chen X., Fan Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135, 125–154.
- Cherubini U., Vecchiato W., Luciano E. (2004). *Copula methods in finance*. Wiley.
- Clemen M. N., Jouini R. T. (1996). Copula models for aggregating expert opinions. *Operations Research*, 44, 444–457.

Coles S., Heffernan J., Tawn J. (1999). Dependence measures for extreme value analyses. *Extremes*, 2 (4), 339–365.

Conover W. J. (1999). *Practical nonparametric statistics*. 3rd ed. Wiley.

Embrechts P., McNeil A., Straumann D. (2002). Correlation and dependence in risk management: properties and pitfalls. In: M. A. H. Dempster (ed.), *Risk Management: Value at Risk and Beyond*. Cambridge: Cambridge University Press.

Fang K., Kotz S., Hg K. (1987). *Symmetric multivariate and related distributions*. London: Chapman Hall.

Fantazzini D. (2010). Three-stage semi-parametric estimation of t-copulas: Asymptotics, finite-sample properties and computational aspects. *Computational Statistics and Data Analysis*, forthcoming.

Ferguson S. T., Genest C., Hallin M. (2000). Kendall's tau for serial dependence. *Canadian Journal of Statistics*, 28, 587–604.

Fermanian J., Scaillet O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, 5, 25–54.

Genest C., MacKay J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *American Statistics*, 40, 280–285.

Genest C., Ghoudi K., Rivest L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82, 543–552.

Gourieroux C., Monfort A.. (1995). *Statistics and econometric models*. Cambridge: Cambridge University Press.

Greene W. (2002). *Econometric analysis*. Prentice Hall.

Hollander M., Wolfe D. A. (1973). *Nonparametric statistical inference*. New York: Wiley.

Hult H., Lindskog F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34, 587–608.

Joe H., Xu J. (1996). The estimation method of inference functions for margins for multivariate models. *Technical Report No. 166*. Department of Statistics, University of British Columbia.

Joe H. (1997). *Multivariate models and dependence concepts*. London: Chapman Hall.

Kim G., Silvapulle M. J., Silvapulle P. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Computational Statistics and Data Analysis*, 51 (6), 2836–2850.

Kim G., Silvapulle M. J., Silvapulle P. (2008). Estimating the error distribution in multivariate heteroscedastic time series models. *Journal of Statistical Planning and Inference*, 138 (5), 1442–1458.

Koul H. L. (2002). *Weighted empirical processes in dynamic nonlinear models*. Lecture Notes in Statistics, V. 166. Springer.

Koul H. L., Ling S. (2006). Fitting an error distribution in some heteroscedastic time series model. *Annals of Statistics*, 34, 994–1012.

Kruskal W. (1958). Ordinal measures of association. *Journal of the American Statistical Association*, 53, 814–861.

Kurowicka D., Cooke R. M. (2006). *Uncertainty analysis with high dimensional dependence modeling*. New York: Wiley.

Lindskog F., McNeil A., Schmock U. (2003) Kendall's tau for elliptical distributions. In: G. Bol, G. Nakhaeizaden, S. Rachev, T. Ridder, K.-H. Vollmer (eds.), *Credit Risk. Measurement, Evaluation and Management*, Physica-Verlag, A Springer-Verlag Company, Heidelberg, 149–156.

- Mardia K. V., Kent J., Bibby J. M. (1997). *Multivariate analysis*. San Diego: Academic Press.
- McNeil A., Frey R., Embrechts P. (2005). *Quantitative risk management: Concepts, techniques and tools*. Princeton Series in Finance, New Jersey.
- Nelsen R. B. (1999). *An introduction to copulas*. Lecture Notes in Statistics. New York: Springer.
- Nelsen R. B. (2006). *An introduction to copulas*. Lecture Notes in Statistics, 2nd Edition. New York: Springer.
- Newey W. K., McFadden D. (1994) Large sample estimation and hypothesis testing. *Handbook of Econometrics*, V. 4, 2111–2245, Amsterdam: Elsevier Science B. V.
- Oakes D. (1994). Multivariate survival distributions. *Journal of Nonparametric Statistics*, 3, 343–354.
- Patton A. (2006a). Estimation of copula models for time series of possibly different lengths. *Journal of Applied Econometrics*, 21, 147–173.
- Patton A. (2006b). Modelling asymmetric exchange rate dependence. *International Economic Review*, 47 (2), 527–556.
- Poon S., Rockinger M., Tawn J. (2004). Extreme value dependence in financial markets: Diagnostics, models, and financial implications. *Review of Financial Studies*, 17, 581–610.
- Resnick S. I. (1987). *Extreme values, regular variation and point processes*. New York: Springer.
- Rodgers J. L., Nicewander W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42, 59–66.
- Rousseeuw P., Molenberghs G. (1993). Transformation of non positive semidefinite correlation matrices. *Communications in Statistics — Theory and Methods*, 22, 965–984.
- Ruud P. (2000). *An introduction to classical econometric theory*. Oxford University Press.
- Ruymgaart F. H., Shorack G. R., Van Zwet W. R. (1972). Asymptotic normality of nonparametric tests for independence. *The Annals of Statistics*, 43, 1122–1135.
- Shih J. H., Louis T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51, 1384–1399.
- Sibuya M. (1961). Bivariate extreme statistics. *Annals of Mathematical Statistics*, 11, 195–210.
- Statistics with confidence*. (1989). Eds.: Gardner M. J., Altman D. G. London: BMJ.
- Taniguchi M., Kakizawa Y. (2000). *Asymptotic theory of statistical inference for time series*. New York: Springer.
- Van der Vaart A., Wellner J. (1996). *Weak convergence and empirical processes*. New York: Springer.
- Van De Wiel M. A., Bucchianico A. (2001). Fast computation of the exact null distribution of Spearman's rho and Page's L statistic for samples with and without ties. *Journal of Statistical Planning and Inference*, 92, 133–145.
- White H. (1994). *Estimation, inference and specification analysis*. Cambridge University Press.
- Yan J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21 (4), 1–21.