

Обнаружение структурных сдвигов в Copula-моделях

Бродский Б.Е., Сафарян И.А.

(ЦЭМИ РАН, Москва; ИИАП АН Армении, Ереван)

План:

1. Введение: мотивация, обзор
2. Постановка задачи
3. Метод
4. Показатели эффективности
5. Основные результаты
6. Эксперименты
7. Выводы

1. Введение

Доклад посвящен проблеме обнаружения структурных сдвигов в Copula-моделях. Эти модели широко распространены в статистике и финансовой математике. В докладе рассматриваются нестационарные Copula-модели с дискретным временем, структурные параметры которых могут резко изменяться в неизвестные моменты.

- B.Remillard, O.Scaillet (2009) Testing for equality between two copulas, *Journal of Multivariate Analysis*, vol.100, is.3, pp.377-386.

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

- C.Genest, B.Remillard, F.Beaudoin (2007) Goodness of fit tests for copulas: a review and a power study.

2. Постановка задачи

Рассмотрим непрерывный случайный вектор $X = \{X_1, \dots, X_d\}$ с совместной кумулятивной функцией распределения V и одномерными ф.р. компонент вектора $F_1, \dots, F_d$. Copula представление для V запишется в виде $V(x_1, \dots, x_d) = G\{F_1(x_1), \dots, F_d(x_d)\}$, где G - единственная кумулятивная ф.р., имеющая одномерные маргинальные ф.р., равномерно распределенные на $[0, 1]$. Copula-модель возникает в том случае, когда G неизвестна, но принадлежит классу

$$\mathcal{G}_0 = \{G_\theta : \theta \in \mathcal{O}\},$$

где $\mathcal{O}$ - открытое множество в R^p .

Предположим, что получены независимые наблюдения $\{X_1, X_2, \dots, X_N\}$ R^d -значных векторов. В некоторый неизвестный момент $m = [\theta N]$, $0 \leq \theta < 1$, паттерн зависимости компонент $X_{i1}, \dots, X_{id}$ вектора X_i резко изменяется, т.е. совместная функция распределения

$$V(x_{i1}, \dots, x_{id}) = \begin{cases} G_1(F(x_{i1}), \dots, F(x_{id})), & 1 \leq i \leq m \\ G_2(F(x_{i1}), \dots, F(x_{id})), & i > m, \end{cases}$$

где $m = [\theta N]$, θ - неизвестный параметр "разладки".

Необходимо проверить гипотезу

$$H_0 : G_1 = G_2 \quad \text{против альтернативы} \quad H_1 : G_1 \neq G_2.$$

Иными словами, проверяется гипотеза о наличии структурного сдвига в паттерне зависимости между компонентами векторов наблюдений. При отклонении нулевой гипотезы H_0 необходимо построить оценку момента разладки m . Мы заинтересованы в построении метода, для которого вероятности ошибок 1-го рода ("ложная тревога") и 2-го рода ("ложное спокойствие") были бы достаточно малы (стремились к нулю с увеличением объема выборки N), а оценка параметра разладки $\hat{\theta}_N$, была бы состоятельной, т.е. стремилась бы к истинному параметру разладки θ при увеличении объема выборки.

3. Метод

Предлагаемый метод обнаружения основан на непараметрическом подходе. Рассмотрим эмпирические Copula-процессы: для каждого $l = 1, \dots, N-1$

$$D_l(u) = \frac{1}{l} \sum_{i=1}^l I(U_{i,l} \leq u) = \frac{1}{l} \sum_{i=1}^l \prod_{j=1}^d I(U_{ij,l} \leq u_j)$$

$$D_{N-l}(u) = \frac{1}{N-l} \sum_{i=l+1}^N I(U_{i,N-l} \leq u) = \frac{1}{N-l} \sum_{i=l+1}^N \prod_{j=1}^d I(U_{ij,N-l} \leq u_j),$$

где $U_{i,l} = (U_{i1,l}, \dots, U_{id,l})$ и для каждого $j \in [1, \dots, d]$,

$$U_{ij,l} = \frac{l}{l+1} F_{j,l}(X_{ij}) = \text{rank}(X_{ij})/(l+1), \quad 1 \leq i \leq l,$$

$$U_{ij,N-l} = \text{rank}(X_{ij})/(N-l+1), \quad l+1 \leq i \leq N$$

с

$$F_{j,l}(x_j) = \frac{1}{l} \sum_{i=1}^l I(X_{ij} \leq x_j).$$

Для последовательного обнаружения момента разладки m зафиксируем константу $0 < \beta < 1/2$ и рассмотрим следующую статистику (модификация статистики Колмогорова-Смирнова):

$$E_{l,N-l}(u) = (D_l(u) - D_{N-l}(u))\sqrt{l(N-l)}/N$$

и

$$T_N = \max_{[\beta N] \leq l \leq [(1-\beta)N]} \sup_u |E_{l,N-l}(u)|.$$

Оценка момента разладки строится следующим образом:

$$\hat{m}_N \in \arg \max_{[\beta N] \leq l \leq [(1-\beta)N]} \sup_u |E_{l,N-l}(u)|,$$

а оценка параметра разладки $\hat{\theta}_N = \hat{m}_N/N$.

4. Показатели эффективности

В докладе предлагаются следующие показатели эффективности метода обнаружения:

1) Вероятность ошибки 1-го рода (“ложное решение”):

$$\alpha_N = P_0\{T_N > C\},$$

где $C > 0$ - некоторая граница принятия решения о наличии разладки;

2) Вероятность ошибки 2-го рода:

$$\delta_N = P_m\{T_N \leq C\}.$$

3) Вероятность ошибки оценивания: для любого $0 < \epsilon < 1/2$

$$\gamma_N = P_m\{|\hat{\theta}_N - \theta| > \epsilon\},$$

5. Основные результаты

В следующей теореме получена верхняя оценка для вероятности ошибки 1-го рода для предложенного метода.

Теорема 1.

$$\alpha_N \leq L_1 \exp(-L_2 CN),$$

где положительные константы L_1, L_2 не зависят от N .

Теорема 2.

Обозначим $\eta = \sup_u |G_1(u) - G_2(u)|$ и предположим, что $0 < C < \eta/4$. Пусть $d = \eta/4 - C$. Тогда

$$\delta_N \leq L_1 \exp(-L_2 dN),$$

где положительные константы L_1, L_2 не зависят от N .

Для вероятности ошибки оценивания справедлива экспоненциальная оценка сверху

$$\gamma_N \leq L_1 \exp(-L_2 N),$$

где положительные константы L_1, L_2 не зависят от N .

6. Экспериментальное тестирование

Предложенный метод тестировался в имитационных экспериментах с выборками двумерных зависимых наблюдений, описываемых Copula-моделями следующего вида:

Clayton copula : для всех $u, v \in (0, 1)$ и $\kappa > 0$

$$C_{\kappa}(u, v) = (u^{-\kappa} + v^{-\kappa} - 1)^{-1/\kappa}.$$

Gumbel copula: для всех $u, v \in (0, 1)$ и $\kappa > 0$

$$C_{\kappa}(u, v) = \exp[-\{(-\log u)^{1/\kappa} + (-\log v)^{1/\kappa}\}^{\kappa}].$$

Параметр κ в данных моделях может изменяться в некоторый неизвестный момент разладки $m = [\theta N]$.

В первой серии экспериментов оценивались критические пороги для предложенного метода для различных моделей наблюдений и объемов выборки. С этой целью использовались выборки с однородными наблюдениями (без разладки). Для каждого значения объема выборки N в 500 независимых повторениях эксперимента оценивались 95- и 99- процентные квантили выборки максимумов статистики T_N для предложенного метода. Значения 95-процентных квантилей далее использовались в качестве критических порогов. Полученные результаты приведены в таблицах 1 и 2.

Таблица 1 (Критические пороги, Clayton copula, однородная выборка $\kappa = 0.3$)

N	50	100	200	300	500	700	1000	1500	2000
0.95	0.1156	0.0850	0.0615	0.0492	0.0372	0.0314	0.0278	0.0213	0.0197
0.99	0.1343	0.0945	0.0674	0.0550	0.0426	0.0348	0.0323	0.0232	0.0214

Таблица 2 (Критические пороги, Gumbel copula, однородная выборка $\kappa = 0.3$)

N	50	100	200	300	500	700	1000	1500	2000
0.95	0.1033	0.0749	0.0508	0.0402	0.0313	0.0243	0.0206	0.0158	0.0146
0.99	0.1187	0.0836	0.0585	0.0461	0.0343	0.0292	0.0233	0.0168	0.0154

Из этих результатов видно, что критические пороги предложенного метода не очень чувствительны к конкретной Copula-модели, что позволяет провести робастную настройку его параметров для обнаружения и тестирования момента разладки. Полученные результаты приведены в таблицах 3 и 4.

Таблица 3 (Обнаружение и оценивание разладки, Clayton copula, разладка $\kappa_1 = 0.3$, $\kappa_2 = 1.0$, параметр разладки $\theta = 0.3$, C - критический порог; w_2 - ошибка 2-го рода)

$\theta = 0.3$	$\kappa_1 = 0.3, \kappa_2 = 1.0$			
N	500	700	1000	1500
C	0.037	0.031	0.027	0.020
w_2	0.56	0.43	0.15	0.02
$\hat{\theta}_N$	0.337	0.335	0.303	0.30

Таблица 4 (Обнаружение и оценивание разладки, Gumbel copula, разладка $\kappa_1 = 0.3$, $\kappa_2 = 0.7$, параметр разладки $\theta = 0.3$)

$\theta = 0.3$	$\kappa_1 = 0.3, \kappa_2 = 0.7$						
N	100	200	300	500	700	1000	1500
C	0.07	0.05	0.04	0.03	0.02	0.017	0.015
w_2	0.69	0.60	0.44	0.33	0.04	0.01	0
$\hat{\theta}_N$	0.45	0.40	0.35	0.33	0.31	0.305	0.30

7. Выводы

- 1) Предложен новый метод обнаружения структурных сдвигов в Copula-моделях
- 2) Свойства метода исследованы теоретически (экспоненциальные верхние оценки для вероятностей ошибки 1-го и 2-го рода, а также для вероятности ошибки оценивания)
- 3) Проведено экспериментальное тестирование метода на Copula-моделях Gumbel, Clayton
- 4) Для решения поставленной задачи обнаружения структурных сдвигов могут быть также использованы статистики интегрального типа

Литература

1. B.Remillard, O.Scaillet (2009) Testing for equality between two copulas, Journal of Multivariate Analysis, vol.100, is.3, pp.377-386.
2. J.-D.Fermanian, O.Scaillet (2003) Nonparametric estimation of copulas for time series, J.Risk, 5, pp.25-54.