

Д. Фантаццини

Моделирование многомерных распределений с использованием копула-функций. III¹

Заключительная часть консультации посвящена описанию подходов к эмпирическому подбору подходящей копула-функции и методов статистической проверки гипотез, связанных с моделями копула-функций.

Ключевые слова: критерий согласия, критическая статистика, копула-функция, байесовский выбор, информационный критерий качества подгонки, эмпирический анализ.

JEL classification: C69, C49.

Содержание

8. ВЫБОР КОПУЛА-ФУНКЦИИ

- 8.1. Информационный критерий Акаике (AIC)
- 8.2. Тесты отношения правдоподобия
- 8.3. Байесовский выбор копула-функции
- 8.4. Эмпирические приложения: пример выбора копула-функции

9. КРИТЕРИИ СОГЛАСИЯ ДЛЯ КОПУЛА-ФУНКЦИЙ

- 9.1. Тесты, основанные на преобразовании Розенблатта
- 9.2. Тесты, основанные на использовании эмпирических копула-функций
- 9.3. Другие подходы к построению критериев согласия
- 9.4. Тесты, основанные на усреднении критических статистик
- 9.5. Эмпирические приложения с пакетом R: пример реализации подхода A_2

8. Выбор копула-функции

Из нескольких моделей копула-функций (определение копула-функции см. в (Фантаццини, 2011a)) следует выбрать ту, которая наилучшим образом описывает рассматриваемые данные. В разделах 8.1–8.3 описываются различные критерии, руководствуясь которыми исследователи обычно осуществляют такой выбор.

8.1. Информационный критерий Акаике (AIC)

Этот метод выбора копула-функций является достаточно простым, см., например, (Breymann et al., 2003; Dias, Embrechts, 2004).

¹ Окончание. Первые две части консультации см. в журнале «Прикладная эконометрика», №№ 22 (3) и 23 (2), 2011. Перевод с английского А. В. Кудрова под научной редакцией С. А. Айвазяна. Сохранена сквозная нумерация разделов, рисунков, формул, определений и теорем, начатая в первой части.

AIC (Akaike Information Criterion) — это критерий, позволяющий выбрать копула-функцию, которая отвечает наименьшему значению из следующих:

$$AIC(C_k, \hat{\alpha}_k) = -2 \sum_{i=1}^T \log c_k(u_{1i}, \dots, u_{ni}; \hat{\alpha}_k) + 2q_k, \quad k = 1, \dots, K, \quad (23)$$

где C_k — k -ая модель копула-функции, $c_k(\cdot)$ — плотность для C_k ; $\hat{\alpha}_k$ — вектор параметров копула-функции C_k ; q_k — число параметров, от которых зависит функция C_k . Функция $AIC(C_k, \hat{\alpha}_k)$ в определенном смысле штрафует модели с большим числом параметров.

AIC предполагает, что наиболее адекватная модель находится среди рассматриваемых моделей. Если сравнивать «невложенные» модели, то слагаемое $2q_k$ является не вполне корректным штрафом, и указанная выше формула становится неприменимой. Для того чтобы разрешить эту проблему, Такеучи предложил в (Takeuchi, 1976) информационный критерий ТИС. Однако ТИС редко используется на практике, поскольку для него требуется наличие весьма больших (по размеру) выборок наблюдений.

В применении AIC некоторые вопросы остаются открытыми. Например, до сих пор неизвестно, что происходит с AIC в случае, если в качестве частных одномерных распределений использовать эмпирическую функцию распределения. Использование эмпирической функции распределения как оценки соответствующей частной функции распределения на практике является стандартной процедурой, но последствия такого использования для AIC до сих пор не изучены.

8.2. Тесты отношения псевдоправдоподобия

В работе (Chen, Fan, 2006) предлагаются так называемые тесты отношения псевдоправдоподобия, которые учитывают случайность, возникающую в AIC при использовании нормированных рангов. Эти тесты позволяют удостовериться в том, что среди рассматриваемых моделей нет такой, которая была бы значительно лучше выбранной (проверяемой) модели. При этом не требуется, чтобы рассматриваемый класс моделей содержал истинную модель. Таким образом можно сравнивать невложенные модели.

Основной идеей тестов является сравнение каждой из рассматриваемых моделей-кандидатов с остальными моделями с точки зрения меры их правдоподобия, по результатам сравнения выбирается та модель, которая оказывается наиболее правдоподобной. А именно, пусть $C_1(u_1, \dots, u_n; \alpha_1)$ — выбранная из числа рассматриваемых *модель-образец* копула-функции, $C_i(u_1, \dots, u_n; \alpha_i)$, $i = 2, \dots, M$ — остальные модели-кандидаты, $C^0(u_1, \dots, u_n)$ — истинная (и *ненаблюдаемая*) копула-функция, в действительности описывающая зависимости между компонентами многомерных наблюдений. В работе (Chen, Fan, 2006) предлагается, используя расстояние Кульбака–Лейблера, тестировать гипотезу о том, является ли модель-образец наиболее правдоподобной среди всех рассматриваемых моделей. Таким образом, нулевой гипотезой является:

$$H_0^M : \max_{i=2, \dots, M} E^0 \left[\ln \frac{c_i(F_1^0(\varepsilon_{1i}), \dots, F_n^0(\varepsilon_{ni}); \alpha_i^*)}{c_1(F_1^0(\varepsilon_{1i}), \dots, F_n^0(\varepsilon_{ni}); \alpha_1^*)} \right] \leq 0,$$

где α_i^* , $i=1, \dots, M$ — параметры рассматриваемых моделей копула-функций; $F_j^0(\cdot)$ — истинные (неизвестные) частные распределения для ε_{jt} , $j=1, \dots, n^2$. Указанная нулевая гипотеза означает, что среди всех рассматриваемых моделей модель-образец оказывается, в определенном смысле, ближе к истинной модели, чем все остальные из рассматриваемых. Альтернативная гипотеза

$$H_1^M: \max_{i=2, \dots, M} E^0 \left[\ln \frac{c_i(F_1^0(\varepsilon_{1t}), \dots, F_n^0(\varepsilon_{nt}); \alpha_i^*)}{c_1(F_1^0(\varepsilon_{1t}), \dots, F_n^0(\varepsilon_{nt}); \alpha_1^*)} \right] > 0$$

означает, что среди рассматриваемых моделей существует модель-кандидат, которая ближе к истинной модели, чем модель-образец. Тест из работы (Chen, Fan, 2006) основан на следующей статистике отношения псевдоправдоподобия ($i=2, \dots, M$):

$$LR_n(F_{1T}, \dots, F_{nT}; \tilde{\alpha}_i, \tilde{\alpha}_1) = \frac{1}{T} \sum_{t=1}^T \left[\ln \frac{c_i(F_{1T}(\varepsilon_{1t}), \dots, F_{nT}(\varepsilon_{nt}); \tilde{\alpha}_i)}{c_1(F_{1T}(\varepsilon_{1t}), \dots, F_{nT}(\varepsilon_{nt}); \tilde{\alpha}_1)} \right],$$

где F_{jT} , $j=1, \dots, n$ — нормированные ранги; знак $\sim$ над буквой означает, что соответствующий параметр α оценен при помощи квази метода максимального правдоподобия (КММП). Возьмем $\Omega = (\sigma_{ik})_{i,k=2}^M$, где

$$\sigma_{ik} = \text{Cov}^0 \left[\ln \frac{c_i(U_{1t}, \dots, U_{nt}; \alpha_i^*)}{c_1(U_{1t}, \dots, U_{nt}; \alpha_1^*)} + \sum_{j=1}^n \{Q_{i,j}(U_{jt}; \alpha_i^*) - Q_{1,j}(U_{jt}; \alpha_1^*)\}, \right. \\ \left. \ln \frac{c_k(U_{1t}, \dots, U_{nt}; \alpha_k^*)}{c_1(U_{1t}, \dots, U_{nt}; \alpha_1^*)} + \sum_{j=1}^n \{Q_{k,j}(U_{jt}; \alpha_k^*) - Q_{1,j}(U_{jt}; \alpha_1^*)\} \right]$$

при $U_{jt} = F_j^0(\varepsilon_{jt})$ и $Q_{i,j}(U_{jt}; \alpha_i^*) \equiv E^0 \{I_{ij}(U_s; \alpha_i^*) [1_{\{U_{jt} \leq U_{js}\}} - U_{js}] | U_{jt}\}$ для $i=1, \dots, M$, $j=1, \dots, d$, добавочные слагаемые $Q_{i,j}(\cdot)$ введены в связи с необходимостью оценивать частные распределения $F_j^0(\cdot)$, $j=1, \dots, n$, а $1_{\{x < x_0\}}$, как и прежде, индикаторная функция множества $\{x < x_0\}$. Более подробно см. (Chen, Fan, 2006). Отметим, что если частные распределения известны, то эти слагаемые исчезают. Как показано в работе (Chen, Fan, 2006), при некоторых условиях регулярности:

$$n^{1/2} \left\{ LR_n(F_{1T}, \dots, F_{nT}; \tilde{\alpha}_i, \tilde{\alpha}_1) - E^0 \left[\ln \frac{c_i(U_{1t}, \dots, U_{nt}; \alpha_i^*)}{c_1(U_{1t}, \dots, U_{nt}; \alpha_1^*)} \right] \right\}_{i=2, \dots, M} \xrightarrow{d} (Z_2, \dots, Z_M)',$$

где вектор $(Z_2, \dots, Z_M)'$ имеет нормальное распределение $N(0, \Omega)$, так что

$$\max_{i=2, \dots, M} n^{1/2} \left\{ LR_n(F_{1T}, \dots, F_{nT}; \tilde{\alpha}_i, \tilde{\alpha}_1) - E^0 \left[\ln \frac{c_i(U_t; \alpha_i^*)}{c_1(U_t; \alpha_1^*)} \right] \right\} \xrightarrow{d} \max_{i=2, \dots, M} Z_i. \quad (24)$$

² Верхний индекс ⁰ здесь и в дальнейшем означает, что соответствующая характеристика вычислена в условиях гипотезы H_0^M .

В (Chen, Fan, 2006) показано, что при так называемой «наименее благоприятной» конфигурации

$$T_T^M = \max_{i=2, \dots, M} [n^{1/2} LR_n(F_{1T}, \dots, F_{nT}; \tilde{\alpha}_i, \tilde{\alpha}_1, \tilde{\theta})] \xrightarrow{d} \max_{i=2, \dots, M} Z_i, \quad (25)$$

т. к. второе слагаемое в (24) обнуляется: $E^0[\cdot] = 0$. Поскольку асимптотическая ковариационная матрица Ω для $Z_2, \dots, Z_M$ зависит от $\alpha_1^*, \dots, \alpha_M^*$, распределение $\max_i Z_i$ неизвестно. Но, несмотря на это, можно использовать p -значения, получаемые либо методом Монте-Карло, либо бутстрап методом, см. (White, 2000; Chen, Fan, 2006).

Состоятельная оценка для $\Omega = (\sigma_{ik})_{i,k=2}^M$ определяется формулой:

$$\hat{\sigma}_{ik} = \frac{1}{T} \sum_{i=1}^T \left(\ln \frac{c_i(F_{iT}; \tilde{\alpha}_i)}{c_1(F_{iT}; \tilde{\alpha}_1)} - \frac{1}{T} \sum_{s=1}^T \ln \frac{c_i(F_{sT}; \tilde{\alpha}_i)}{c_1(F_{sT}; \tilde{\alpha}_1)} + \sum_{j=1}^n (\hat{Q}_{i,j}(F_{jt}; \tilde{\alpha}_i) - \hat{Q}_{1,j}(F_{jt}; \tilde{\alpha}_1)) \right) \times \\ \times \left(\ln \frac{c_k(F_{iT}; \tilde{\alpha}_k)}{c_1(F_{iT}; \tilde{\alpha}_1)} - \frac{1}{T} \sum_{s=1}^T \ln \frac{c_k(F_{sT}; \tilde{\alpha}_k)}{c_1(F_{sT}; \tilde{\alpha}_1)} + \sum_{j=1}^n (\hat{Q}_{k,j}(F_{jt}; \tilde{\alpha}_k) - \hat{Q}_{1,j}(F_{jt}; \tilde{\alpha}_1)) \right),$$

где $\hat{Q}_{i,j}$ — состоятельная оценка $Q_{i,j}$, которая вычисляется следующим образом:

$$\hat{Q}_{i,j}(U_{js}; \hat{\alpha}_i) = \frac{1}{T} \sum_{t=1, t \neq s}^T l_{ij}(U_t; \hat{\alpha}_i) (\mathbf{1}\{U_{js} \leq U_{jt}\} - U_{jt})^3.$$

Затем можно вычислить разложение Холецкого $\hat{\Omega} = \hat{C}\hat{C}'$ и сформировать

$$\hat{Z}_b = (\hat{Z}_{b,2}, \dots, \hat{Z}_{b,M})' = \hat{C}\eta_b,$$

где $\eta_b \in N(0, I_{M-1})$, т. е. имеет $(M-1)$ -мерное стандартное нормальное распределение. После чего найдем наибольшую порядковую статистику $\zeta_{b,M-1} = \max_{k=2, \dots, M} \hat{Z}_{b,k}$, а для того, чтобы вычислить p -значение методом Монте-Карло, повторим эти итерации много раз (например, порядка $B = 100000$), см. (Chen, Fan, 2006):

$$\frac{1}{B} \sum_{b=1}^B \mathbf{1}\{\zeta_{b,M-1} \leq T_T^M\}.$$

8.3. Байесовский выбор копула-функций

В (Huard et al., 2006) предлагается байесовская процедура выбора наиболее вероятного двумерного семейства копула-функций среди заданного множества копула-функций. В этом подходе параметры копула-функций интерпретируются как случайные величины, и, следовательно, они не должны оцениваться. Кроме того, в указанной работе параметризация плотности копула-функции осуществляется в терминах τ -Кендалла (см. (Фантаццини, 2011b, п. 5.2)), так что априорные распределения параметров заменены априорными распределениями τ -Кендалла, которые концептуально являются более удобными. Априорное распре-

³ Здесь и в дальнейшем $\mathbf{1}\{\cdot\}$ означает, как и раньше, индикаторную функцию множества $\{\cdot\}$.

деление τ -Кендалла берется одним и тем же для всех тестируемых копула-функций и используется как базис для сравнения.

Пусть $\mathcal{C}$ — множество всех копула-функций. Возьмем из $\mathcal{C}$ конечное подмножество $\mathcal{C}_Q \subset \mathcal{C}$ копула-функций, которые необходимо включить в предложенный метод. Каждая копула-функция из $\mathcal{C}_Q$ обозначается C_l , $l = 1, \dots, Q$. Метод, предложенный в работе (Huard et al., 2006), на первом шаге состоит в проверке следующих Q гипотез:

$$H_l: \text{данные извлечены из копула-функции } C_l, \quad l = 1, \dots, Q.$$

Для этого вычисляются $P(H_l | D)$ — вероятности реализации гипотезы H_l при заданных наблюдениях D , которые представляют собой T независимых пар (u_t, v_t) , $t = 1, \dots, T$, с равномерно распределенными компонентами (в работе (Huard et al., 2006) называемые квантилями). Если используются нормированные ранги, то условие независимости может нарушаться, и предложенный метод следует рассматривать как аппроксимацию и применять с некоторой осторожностью. Используя теорему Байеса, получим:

$$P(H_l | D, I) = \frac{P(D | H_l, I)P(H_l | I)}{P(D | I)}, \quad (26)$$

где $P(D | H_l, I)$ — функция правдоподобия, $P(H_l | I)$ — априорное распределение для копула-функции, $P(D | I)$ — нормирующая константа; I соответствует дополнительной информации. «Правильной» копула-функцией, в терминах работы (Huard et al., 2006), называют копула-функцию, отвечающую наибольшему значению апостериорной вероятности $P(H_l | D, I)$.

В (Huard et al., 2006) плотность копула-функции перепараметризируется в терминах τ -Кендалла ($\tau = g_l(\alpha)$), который становится общим параметром для всех копула-функций из $\mathcal{C}_Q$.

Таблица 3. Выражения для τ -Кендалла для некоторых копула-функций

Копула-функция	$\tau = g_l(\alpha)$	Область изменения τ
Клейтон	$1 - 2(2 + \alpha)^{-1}$	$[0; 1] \setminus \{0\}$
Гумбель	$1 - \alpha^{-1}$	$[0; 1]$
Франк	$1 - 4\alpha^{-1} \left(1 - \alpha^{-1} \int_0^\alpha s / (e^s - 1) ds \right)$	$[-1; 1] \setminus \{0\}$
Гаусс	$2\pi^{-1} \arcsin \rho$	$[-1; 1]$

Следовательно, τ -Кендалла может использоваться в (26) как случайная величина

$$P(H_l | D, I) = \int_{-1}^{+1} P(H_l, \tau | D, I) d\tau = \int_{-1}^{+1} \frac{P(D | H_l, \tau, I)P(H_l | \tau, I)P(\tau | I) d\tau}{P(D | I)}, \quad (27)$$

где $P(H_l | \tau, I)$ — априорная вероятность гипотезы H_l для копула-функции; $P(\tau | I)$ — априорная плотность для τ -Кендалла; правдоподобие $P(D | H_l, \tau, I)$, которое теперь зависит от τ , может быть вычислено следующим образом:

$$P(D | H_l, \tau, I) = \prod_{i=1}^T P(u_i, v_i | \tau, I) = \prod_{i=1}^T c_i(u_i, v_i | g_i^{-1}(\tau)), \quad (28)$$

где $c_i(u_i, v_i | g_i^{-1}(\tau))$ — плотность l -ой копула-функции, параметризованной в терминах τ -Кендалла.

Для того чтобы выбрать априорные распределения, в работе (Huard et al., 2006) указываются некоторые требования и правила:

(I_1) τ -Кендалла принадлежит множеству Λ и каждая реализация $\tau \in \Lambda$ равновероятна;

(I_2) для заданного τ все копула-функции, для которых $\tau \in \Omega_l$, равновероятны, здесь Ω_l — область значений τ в условиях справедливости гипотезы H_l .

Область Λ необходима для того, чтобы включать дополнительную информацию относительно зависимости между переменными. Например, если известно, что τ положительно, то можно предположить, что $\Lambda = [0; 1]$; если априорная информация относительно значений τ отсутствует, то $\Lambda = [-1; 1]$. Следовательно, учитывая (I_1), априорное распределение τ есть:

$$P(\tau | I_1) = \begin{cases} \frac{1}{\lambda(\Lambda)} & \text{при } \tau \in \Lambda; \\ 0 & \text{иначе,} \end{cases} \quad (29)$$

где $\lambda(\cdot)$ обозначает меру Лебега, т.е. $\lambda(\Lambda)$ — длина интервала Λ . Кроме того, условие (I_2) определяет априорное распределение для копула-функции:

$$P(H_l | \tau, I_2) \propto \mathbf{1}\{\tau \in \Omega_l\}, \quad (30)$$

где предполагается, что все копула-функции равновероятны относительно τ . В случае, если известно, что τ концентрируется вокруг определенного значения, то в работе (Huard et al., 2006) предлагается использовать бета-распределение на интервале $[-1; 1]$ с параметрами, обеспечивающими такую форму распределения, которая соответствует априорной информации.

Если подставить (28), (30) и (29) в (27), то получим:

$$\begin{aligned} P(H_l | D, I) &= \frac{1}{P(D | I)} \int_{-1}^{+1} \prod_{i=1}^T c_i(u_i, v_i | g_i^{-1}(\tau)) \cdot \frac{\mathbf{1}\{\tau \in \Omega_l \cap \Lambda\}}{\lambda(\Lambda)} d\tau = \\ &= \frac{1}{P(D | I) \lambda(\Lambda)} \int_{\Omega \cap \Lambda} \prod_{i=1}^T c_i(u_i, v_i | g_i^{-1}(\tau)) d\tau, \end{aligned} \quad (31)$$

где нормирующая константа $P(D | I)$ в (31) вычисляется при помощи так называемого правила суммы, более подробно об этом см. (Jaynes, Bretthorst, 2003):

$$P(D | I) = \sum_{l=1}^Q P(D | H_l, I) P(H_l | I).$$

Важно отметить, что правило суммы справедливо только тогда, когда гипотезы H_i являются взаимоисключающими и исчерпывающими (в совокупности). К сожалению, эти условия обычно не выполняются по двум причинам. Во-первых, если данные сгенерированы из копула-функции, не входящей в $\mathcal{C}_Q$ (что является довольно распространенной ситуацией), то гипотезы H_i в совокупности не являются исчерпывающими. Во-вторых, если множество гипотез H_i содержит две или более похожих копула-функции, то в этом случае гипотезы H_i не являются исчерпывающими. В работе (Huard et al., 2006) весьма кратко рассматриваются возможные решения, которые бы обеспечили полноту, но более детальное изучение этого вопроса оставлено для дальнейшего исследования.

Этот метод представляется весьма интересным, поскольку не требует предварительного оценивания параметров копула-функций. Более того, отметим, что рассматриваемые копула-функции не обязаны быть вложенными.

8.4. Эмпирические приложения: пример выбора копула-функций

Критерий AIC не вычисляется напрямую в пакете R , но несложно написать небольшую функцию, которая бы вычисляла значение AIC при заданных значении максимума логарифмической функции правдоподобия и числе параметров. Эта задача остается в качестве упражнения.

В пакете MATLAB реализован код для байесовского выбора копула-функции из работы (Huard et al., 2006), он доступен на сайте <http://code.google.com/p/copula/>.

9. Критерии согласия для копула-функций

После того, как, согласно указанным выше процедурам, выбрана «наиболее подходящая» копула-функция, следующим шагом является проверка того, насколько хорошо она согласуется с имеющимися данными.

Одним из вариантов такой проверки является *неформальная графическая диагностика*, в соответствии с которой сравниваются нормированные ранги со случайными выборками, сгенерированными в соответствии с используемой копула-функцией. А именно, графически сравнивается эмпирическая оценка копула-функции с параметрической моделью. Однако эту процедуру можно рассматривать в качестве предварительного и приблизительно-го анализа.

С недавнего времени начали появляться работы, в которых изучаются статистические тесты, используемые в критериях согласия. Вот лишь некоторые из них: (Genest, Rivest, 1993; Shih, 1998; Breymann et al., 2003; Malevergne, Sornette, 2003; Scaillet, 2005; Fermanian, 2005; Panchenko, 2005; Genest et al., 2006a; Berg, Bakken, 2007; Dobrić, Schmid, 2007; Quessy et al., 2007; Genest et al., 2009; Genest, Rémillard, 2008; Berg, 2009).

Пусть $C(\cdot)$ — n -мерная копула-функция для рассматриваемой выборки. Необходимо протестировать следующую гипотезу:

$$H_0 : C \in \mathcal{C} = \{C_\alpha; \alpha \in \Theta\} \text{ против } H_1 : C \notin \mathcal{C},$$

где $C_\alpha(\cdot)$ — параметрическая копула-функция, Θ — пространство параметров.

В случае, когда выбирается наиболее приемлемая *одномерная* модель, обычно используются такие критерии согласия, как критерий Колмогорова–Смирнова или критерий Андерсона–Дарлингга.

Возможным вариантом построения критерия согласия для копула-функций, используемых для описания многомерных распределений, являются подходы, основанные на многомерном группировании, см., например, (Dobrić, Schmid, 2005). К сожалению, при этом, как и во всех подходах, основанных на группировании, используется дискретизация вероятностного пространства, которая практически нереализуема в случае задач большой размерности. Более того, группирование данных в некотором смысле произвольно и весьма нетривиально. Аналогично, подходы, основанные на использовании многомерных ядерных оценок копула-функций, представленные, например, в работах (Fermanian, 2005; Scaillet, 2005), являются в случае большой размерности (характерной для задач из области финансов и страхования) чрезмерно затратными с вычислительной точки зрения. Поэтому далее, характеризуя многомерный случай, будут рассматриваться некоторые другие подходящие подходы.

Перед тем как приступить к анализу различных тестов, кратко определим, что представляет собой *преобразование Розенблатта*. Изначально оно было предложено в работе (Rosenblatt, 1952) и стало известным как *условно-вероятностное интегральное преобразование* (Conditional Probability Integral Transform, CPIT). Оно позволяет преобразовать набор зависимых случайных величин, описываемых некоторым многомерным распределением, в набор независимых равномерно распределенных на $[0; 1]$ случайных величин. Таким образом, если для многомерных случайных величин имеется тест на независимость и равномерность, то, используя преобразование Розенблатта, можно протестировать адекватность любой модели.

Если взять случайную величину $\nu = F(X)$, где $F(\cdot)$ — функция распределения случайной величины X , то ν будет иметь равномерное распределение на $[0; 1]$. Если в качестве $F(\cdot)$ взять иную функцию распределения, скажем $G(\cdot)$, то $\nu = G(X)$ уже не будет равномерно распределенной. Это простое свойство лежит в основе большинства работ, в которых изучаются вопросы, связанные с тестированием распределений и копула-функций.

Дадим формальное определение преобразования Розенблатта.

Определение 13 (преобразование Розенблатта). Пусть $X = (X_1, \dots, X_n)$ — случайный вектор с функцией распределения $F_X(x_1, x_2, \dots, x_n)$; $F_{X_i} = P(X_i \leq x_i)$ — функция распределения случайной величины X_i , $i = 1, \dots, n$. Рассмотрим следующее вероятностно-интегральное преобразование, определяющее n случайных величин $V_i = T(X_i)$:

$$\begin{aligned} T(x_1) &= P(X_1 \leq x_1) = F_{X_1}(x_1); \\ T(x_2) &= P(X_2 \leq x_2 \mid X_1 = x_1) = F_{X_2|X_1}(x_2 \mid x_1); \\ &\vdots \\ T(x_n) &= P(X_n \leq x_n \mid X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = \\ &= F_{X_n|X_1, \dots, X_{n-1}}(x_n \mid x_1, \dots, x_{n-1}). \end{aligned}$$

Тогда случайные величины $V_i = T(X_i)$ ($i = 1, 2, \dots, n$) являются равномерно распределенными на отрезке $[0; 1]$ и взаимно независимыми, а сами преобразования $T(X_i)$ называются

ся преобразованиями Розенблатта или условно-вероятностными интегральными преобразованиями (СПИТ).

Предположим, что многомерная функция распределения $F_X(\cdot)$ имеет копула-функцию $C(\cdot)$, т. е.

$$F_X(x_1, x_2, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)).$$

Обозначим через $C_i(u_1, \dots, u_i)$ совместное i -мерное распределение, такое, что:

$$C_i(u_1, \dots, u_i) = C(u_1, \dots, u_i, 1, \dots, 1), \quad i = 1, 2, \dots, n-1.$$

При этом $C_1(u_1) = u_1$ и $C_n(u_1, \dots, u_n) = C(u_1, \dots, u_n)$. Пусть случайный вектор $(U_1, \dots, U_i)$ имеет распределение $C_i(u_1, \dots, u_i)$. Тогда плотность условного распределения U_i в точке u_i при заданных значениях $u_1, u_2, \dots, u_{i-1}$ случайных величин $U_1, \dots, U_{i-1}$ соответственно равна:

$$c_i(u_i | u_1, \dots, u_{i-1}) = \frac{\partial^{i-1} C_i(u_1, \dots, u_i)}{\partial u_1 \dots \partial u_{i-1}} \bigg/ \frac{\partial^{i-1} C_{i-1}(u_1, \dots, u_{i-1})}{\partial u_1 \dots \partial u_{i-1}}$$

при $i = 1, \dots, n$. Теперь с использованием условных распределений $C_i(\cdot)$ можно определить случайные величины V_i при $i = 2, \dots, n$:

$$V_i = C_i(F_{X_i}(X_i) | F_{X_1}(X_1), \dots, F_{X_{i-1}}(X_{i-1})).$$

Как указано в работе (Berg, 2009), преобразование Розенблатта дает значительное преимущество при тестировании критериев согласия. В (Hong, Li, 2005) представлены результаты многомерных тестов с использованием имитационного моделирования и преобразованных исходных случайных величин, которые оказались лучше, чем результаты с использованием *непреобразованных* случайных величин. В работе (Chen et al., 2004) высказывается предположение, что аналогичный вывод справедлив и для критериев согласия для копула-функций.

В качестве недостатка тестов, основанных на преобразовании Розенблатта, следует отметить отсутствие инвариантности относительно принятого порядка нумерации рассматриваемых случайных величин. Общее число возможных преобразований случайных величин тогда будет равно $n!$. В (Berg, 2009) показано, что для некоторых подходов, основанных на преобразовании Розенблатта, оценки p -значений критической статистики зависят от того, какой порядок преобразований выбран для рассматриваемых случайных величин. Однако в этой работе также указывается, что по мере роста числа наблюдений различия в p -значениях уменьшаются, причем эти p -значения асимптотически сходятся к некоторым общим критическим уровням. Таким образом, в работе (Berg, 2009) высказывается предположение, что выбор порядка преобразований лишь незначительным образом влияет на результаты теста.

Следуя схеме, представленной в (Berg, 2009), критерии согласия для копула-функций могут быть классифицированы в девять семейств:

$\mathcal{A}_1$: Тесты, основанные на преобразовании Розенблатта. Предложены в работе (Berg, Bekken, 2007). Этот подход в качестве частных случаев включает в себя подходы, предложенные в работах (Malevergne, Sornette, 2003; Breymann et al., 2003; Chen et al., 2004).

$\mathcal{A}_2$: Тесты, основанные на эмпирических оценках копула-функций. Предложены в работах (Kole et al., 2007; Genest, Rémillard, 2008).

$\mathcal{A}_3$: Тесты, основанные на подходе $\mathcal{A}_2$ и преобразовании Розенблатта. Предложены в (Genest et al., 2009).

$\mathcal{A}_4$: Тесты, основанные на эмпирических оценках копула-функций и их функций распределения. Предложены в работах (Genest, Rivest, 1993; Wang, Wells, 2000; Savu, Trede, 2008; Genest et al., 2006a).

$\mathcal{A}_5$: Тесты, основанные на функциях спирменовской меры зависимости (см. (Фантаццини, 2001b, п. 5.2)). Предложены в (Quessy et al., 2007).

$\mathcal{A}_6$: Тесты, основанные на критерии Шиха (Shih, 1998) для двумерной модели Клейтона. Обобщены на случай произвольной размерности в работе (Berg, 2009).

$\mathcal{A}_7$: Тесты, основанные на внутреннем произведении векторов, как мере расстояния между ними. Предложены в работе (Panchenko, 2005).

$\mathcal{A}_8$: Тесты, основанные на подходе $\mathcal{A}_7$ и преобразовании Розенблатта. Предложены в (Berg, 2009).

$\mathcal{A}_9$: Тесты, основанные на объединении указанных выше подходов. Предложены в (Berg, 2009).

Подходы $\mathcal{A}_1 - \mathcal{A}_5$ основаны на снижении размерности, поскольку статистики этих тестов являются одномерными, тогда как статистика подхода $\mathcal{A}_6$ строится с использованием моментов; критерии $\mathcal{A}_7, \mathcal{A}_8$ представляют собой многомерные подходы. В следующих разделах 9.1–9.4 будем придерживаться структуры, предложенной в работах (Berg, 2009; Berg, Bekken, 2007).

9.1. Тесты, основанные на преобразовании Розенблатта

Подход $\mathcal{A}_1$ (CRIT-подход)

Подход $\mathcal{A}_1$ основан на преобразовании Розенблатта, примененном к нормированным рангам $\hat{\mathbf{u}}_1 = (\hat{u}_{11}, \dots, \hat{u}_{n1}), \dots, \hat{\mathbf{u}}_T = (\hat{u}_{1T}, \dots, \hat{u}_{nT})$, где компоненты $\hat{u}_{it}$ вектора $\hat{\mathbf{u}}_t$ вычисляются в предположении справедливости нулевой гипотезы о копула-функции C_α по следующей формуле (см. п. 1 из раздела 7.1 работы (Фантаццини, 2001b)):

$$\hat{u}_{it} = F_{i,T}(x_{it}), \text{ где } F_{i,T}(x) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}\{x_{it} < x\}, \quad i = 1, 2, \dots, n. \quad (32)$$

В соотношении (32) правая часть определяет непараметрическую оценку частной функции распределения переменной X_i в точке x_{it} . При справедливости нулевой гипотезы, выборка $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_T)$ представляет собой выборку из копула-функции P , отвечающей случаю независимости (также обозначаемой как $C_\perp$). К сожалению, эмпирический аналог преобразования Розенблатта предполагает использование рангов, что индуцирует зависимость эмпирических значений V_i , $i = 1, 2, \dots, n$. Таким образом, V_i , $i = 1, \dots, n$, не являются независимыми одинаково распределенными случайными величинами. Тем не менее, как показа-

но в работе (Genest et al., 2006a), можно получать достаточно надежные оценки p -значений критических статистик при помощи параметрической бутстрап процедуры. Теоретические свойства этих оценок представлены в работе (Genest, Rémillard, 2008)⁴.

Общий вид критической статистики критериев согласия для подхода $\mathcal{A}_1$ следующий:

$$W_{1t} = \sum_{i=1}^n \Gamma\{v_{it}\}, \quad t=1, \dots, T, \quad (33)$$

где $\Gamma\{\cdot\}$ — весовая функция, используемая для взвешивания информации из $(v_1, \dots, v_T)$. В качестве весовой функции может быть использована, например, $\Gamma\{v_{it}\} = (\Phi^{-1}(v_{it}))^2$, что соответствует подходу из (Breumann et al., 2003). Более того, если нулевая гипотеза соответствует гауссовской копула-функции, тогда имеем подход, предложенный в работе (Malevergne, Sornette, 2003). В двух последних упомянутых работах используется статистика Андерсона–Дарлинга (Anderson, Darling, 1954). В работе (Berg, Bekken, 2007) показано, что статистика Андерсона–Дарлинга с $\Gamma\{v_{it}\} = |v_{it} - 0.5|$ дает достаточно хорошие результаты при тестировании нулевой гипотезы, соответствующей гауссовской копула-функции. Следовательно, для двух упомянутых выше вариантов выбора весовой функции имеем два подхода:

$$\mathcal{A}_{1t}^{(a)} = \sum_{i=1}^n (\Phi^{-1}(v_{it}))^2 \quad \text{и} \quad \mathcal{A}_{1t}^{(b)} = \sum_{i=1}^n |v_{it} - 0.5|.$$

Можно показать, что для подхода $\mathcal{A}_{1t}^{(a)}$ функция распределения $F_1(\cdot)$ статистики W_{1t} для любых t представляет собой χ_n^2 -распределение. Однако, как говорилось выше, если использовать нормированные ранги, распределение W_{1t} неизвестно. Но для аппроксимации функции распределения $F_1(\cdot)$ при справедливости нулевой гипотезы необходимо обратиться к бутстрап процедуре.

Так называемый «тестовый датчик» (test observator) S_1 подхода $\mathcal{A}_1$ определяется как функция распределения для $F_1(W_{1t})$, см. (Berg, 2009):

$$S_{1t}(w) = P(F_1(W_{1t}) \leq w), \quad w \in [0; 1].$$

При справедливости нулевой гипотезы имеем $S_{1t}(w) = w$ при всех t . Эмпирическая версия тестового датчика может быть вычислена в виде:

$$\hat{S}_1(w) = \frac{1}{T+1} \sum_{i=1}^T \mathbf{1}\{F_1(W_{1t}) \leq w\}, \quad w = \frac{1}{T+1}, \dots, \frac{T}{T+1}. \quad (34)$$

Для всех подходов, которые предполагают снижение размерности, в работе (Berg, 2009) рассматривается статистика Крамера–фон Мизеса, но следует иметь в виду, что могут быть использованы и другие статистики (например, Андерсона–Дарлинга, Колмогорова–Смир-

⁴ Асимптотические свойства этой бутстрап процедуры до сих пор были получены лишь для подходов $\mathcal{A}_2$ и $\mathcal{A}_4$. Однако результаты, полученные в работах (Berg, Bekken, 2007; Dobrić, Schmid, 2007; Berg, 2009), указывают на справедливость упомянутых асимптотических свойств и для других подходов.

нова, см. (Berg, Bekken, 2007; Marsaglia, Marsaglia, 2004)). Если использовать статистику Крамера–фон Мизеса, то критическая статистика имеет вид:

$$\hat{T}_1 = \frac{T}{3} + \frac{T}{T+1} \sum_{t=1}^T \hat{S}_1 \left(\frac{t}{T+1} \right)^2 - \frac{T}{(T+1)^2} \sum_{t=1}^T (2t+1) \hat{S}_1 \left(\frac{t}{T+1} \right)^2 \quad (35)$$

(см. доказательство в (Berg, 2009)).

Некоторое расширение: СРІТ-2 подход. В работе (Berg, Bekken, 2007) показано, что описанный выше подход, в котором используется весовая функция $(\Phi^{-1}(\cdot))^2$, не всегда является адекватным, поскольку данным, расположенным на границе n -мерного единичного гиперкуба, присваиваются большие веса: при небольших выборках такой способ взвешивания делает подход менее устойчивым и менее мощным, поскольку будут присутствовать несколько наблюдений в граничных областях. Более того, в (Berg, Bekken, 2007) показано, что СРІТ-подход может давать плохие результаты в случаях, когда рассматриваемые данные расположены радиально асимметрично.

Учитывая указанные недостатки, в (Berg, Bekken, 2007) был предложен новый подход, названный СРІТ-2, который позволяет использовать любую весовую функцию в (33) и при помощи дополнительного интегрального условно-вероятностного преобразования, построенного с использованием порядковых статистик, идентифицировать радиальную асимметрию. По существу, в работе (Berg, Bekken, 2007) предлагается на первом шаге применить условное интегрально-вероятностное преобразование к нормированным рангам $\hat{\mathbf{U}}$, а на втором шаге применить условное интегрально-вероятностное преобразование к $\mathbf{V}$, полученному на первом шаге.

Пусть $\mathbf{V} = (V_1, \dots, V_n)$, где $V_1, \dots, V_n$ — независимые равномерно распределенные на $[0; 1]$ случайные величины (будем обозначать это $\mathbf{V} \in U(0; 1)^n$). Вектор $\mathbf{V}$ получен путем применения интегрального условно-вероятностного преобразования к $\hat{\mathbf{U}}$. В работе (Berg, Bekken, 2007) соответствующие порядковые статистики обозначаются как:

$$V_{(1)} \leq V_{(2)} \leq \dots \leq V_{(n-1)} \leq V_{(n)}.$$

Если $V_{(1)}, \dots, V_{(n)}$ — порядковые статистики выборки независимых одинаково распределенных из $U(0; 1)$ случайных величин, тогда $V_{(i)}$ имеет бета-распределение с параметрами $(i, n - (i - 1))$, см. (D'Agostino, Stephens, 1986, глава 8). Для того чтобы вычислить интегральное условно-вероятностное преобразование для порядковых статистик, Берг и Беккен используют результаты теоремы 2.7 из (David, 1981). С использованием результатов теоремы 1 из (Deheuvels, 1984) и того факта, что в условиях справедливости нулевой гипотезы $\mathbf{V} \in U(0; 1)^n$, в работе (Berg, Bekken, 2007) получено следующее выражение для интегрального условно-вероятностного преобразования порядковой статистики $\mathbf{V}$:

$$H_1 = F_{V_{(i)}|V_{(i-1)}}(v_{(i)}) = 1 - \left(\frac{1 - v_{(i)}}{1 - v_{(i-1)}} \right)^{n-(i-1)}, \quad i = 1, \dots, n; \quad v_{(0)} = 0. \quad (36)$$

Слишком большое или слишком малое значение H_i может означать ошибочность нулевой гипотезы, см. (Glen et al., 2001). Позднее в работе (Berg, Bekken, 2007) получена статистика, основанная на одновременном использовании $\mathbf{V}$ и $\mathbf{H}$:

$$W_{2t} = \sum_{i=1}^n \Gamma_V(V_{(i,t)}) \cdot \Gamma_H(H_{it}), \quad t = 1, \dots, T, \quad (37)$$

где $\Gamma_V(\cdot)$ и $\Gamma_H(\cdot)$ — весовые функции, используемые для взвешивания информации из $\mathbf{V}$ и $\mathbf{H}$ соответственно. Выбор весовых функций зависит от того, какой области определения копула-функции уделяется наибольшее внимание. Например, если $\Gamma_V = (\Phi^{-1}(\cdot))^2$ и $\Gamma_H = 1$, то получим исходный CRIT-подход (33) из (Breymann et al., 2003).

В работе (Berg, Bekken, 2007) указывается на то, что во многих тестах могут быть получены достаточно хорошие результаты при использовании следующей комбинации двух весовых функций:

- а) $\Gamma_V(X) = |X - 0.5|$, $\Gamma_H(X) = 1$,
- б) $\Gamma_V(X) = (X - 0.5)^2$, $\Gamma_H(X) = 1$.

Более того, в случае сильной асимметрии рекомендуется выбрать $\Gamma_V(X) = |X - 0.5|$, $\Gamma_H(X) = |X - 0.5|$.

Наконец, важно отметить, что в целом распределение W_{2t} неизвестно, и для его оценки необходимо использовать бутстрап приближение. Если вычислить W_{2t} с использованием некоторых весовых функций Γ_V и Γ_H , то можно смоделировать n независимых одинаково распределенных $U(0; 1)$ случайных величин $\tilde{V}$, вычислить $\tilde{W}_{2t}$, используя те же самые весовые функции, что и для W_{2t} . Повторяя эту процедуру большое число раз в предположении справедливости нулевой гипотезы, можно аппроксимировать функцию распределения F_{2t} для W_{2t} .

Реализация тестов

Далее представлено подробное описание тестовых процедур, включая вычисление оценок соответствующих p -значений в предположении справедливости нулевой гипотезы о параметрическом виде копула-функции. Среди описанных процедур — CRIT и CRIT-2 подходы.

Процедура 9.1. Реализация теста в подходе $\mathcal{A}_1$

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).
2. Оценить параметры копула-функции α с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, где l — логарифмическая функция правдоподобия наблюдений $\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T$, т.е. полупараметрического метода.
3. Вычислить $(\mathbf{v}_1, \dots, \mathbf{v}_T)$, полученные путем применения интегрального условно-вероятностного преобразования к $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$, в предположении справедливости нулевой гипотезы о том, что копула-функция равна $C_{\hat{\alpha}}$.
4. Вычислить CRIT-2 выборку $(\mathbf{h}_1, \dots, \mathbf{h}_T)$, применяя интегральное условно-вероятностное преобразование (36) к $(\mathbf{v}_1, \dots, \mathbf{v}_T)$.
5. Используя весовые функции Γ_V и Γ_H (если $\Gamma_H = 1$, то W_{1t} и W_{2t} совпадают), вычислить W_{1t} или W_{2t} по формулам (33) или (37) соответственно.

6. Если при справедливости нулевой гипотезы для W_{1l} или W_{2l} функция распределения известна, вычислить $F_1(W_1)$ или $F_2(W_2)$ и перейти к шагу 8. Если распределение неизвестно, аппроксимировать $F_1(W_1)$ или $F_2(W_2)$ следующим образом: для некоторого достаточно большого целого N_b итеративно повторять нижеприведенную процедуру для каждого $l \in \{1, \dots, N_b\}$.

а) сгенерировать случайную выборку $\mathbf{v}_l^* = (v_{1,l}^*, \dots, v_{n,l}^*)$, компоненты которой представляют собой компоненты $U(0; 1)^n$ случайного вектора;

б) применяя интегральное условно-вероятностное преобразование (36) к $(v_{1,l}^*, \dots, v_{n,l}^*)$, вычислить $\mathbf{h}_l^* = (h_{1,l}^*, \dots, h_{n,l}^*)$;

в) в соответствии с (33) или (37), используя те же весовые функции, что и на шаге 5, и беря $(v_{1,l}^*, \dots, v_{n,l}^*)$ и $(h_{1,l}^*, \dots, h_{n,l}^*)$, вычислить $\hat{W}_{1,l}^*$ или $\hat{W}_{2,l}^*$ соответственно.

7. Вычислить $F_1(W_1) = \frac{1}{N_b + 1} \sum_{l=1}^{N_b} \mathbf{1}\{\hat{W}_{1,l}^* > W_1\}$. $F_2(W_2)$ вычисляется аналогичным образом.

8. В соответствии с (34) и (35) вычислить $\hat{T}_1$ (относительно статистики Андерсона–Дарлинга см. (Berg, Bekken, 2007)). $\hat{T}_2$ вычисляется аналогично.

9. Для некоторого достаточно большого целого K повторить следующие шаги для каждого $k = 1, \dots, K$:

а) получить псевдо-наблюдения $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$ путем преобразования выборочных данных $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ в нормированные ранги с использованием (32);

б) по полученным таким образом псевдо-наблюдениям оценить параметры копула-функции α^0 с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha)$, т. е. полупараметрического метода;

в) вычислить $(\mathbf{v}_{1,k}^0, \dots, \mathbf{v}_{T,k}^0)$, полученные путем применения интегрального условно-вероятностного преобразования к $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$ в предположении справедливости нулевой гипотезы о том, что копула-функция равна $C_{\alpha_k^0}$;

г) вычислить СРПТ-2 выборку $(\mathbf{h}_{1,k}^0, \dots, \mathbf{h}_{T,k}^0)$ путем применения интегрального условно-вероятностного преобразования (36) к $(\mathbf{v}_{1,k}^0, \dots, \mathbf{v}_{T,k}^0)$;

д) используя такие же весовые функции, как на шаге 5, примененные к $(\mathbf{v}_{1,k}^0, \dots, \mathbf{v}_{T,k}^0)$ и $(\mathbf{h}_{1,k}^0, \dots, \mathbf{h}_{T,k}^0)$, вычислить $W_{1,k}^0$ или $W_{2,k}^0$ по формулам (33) или (37) соответственно;

е) если при справедливости нулевой гипотезы для $W_{1,k}^0$ или $W_{2,k}^0$ функция распределения известна, вычислить $F_1(W_{1,k}^0)$ или $F_2(W_{2,k}^0)$ и перейти к шагу (з);

• если распределение неизвестно, аппроксимировать $F_1(\cdot)$ или $F_2(\cdot)$, итеративно повторяя (для некоторого достаточно большого целого N_b) нижеприведенную процедуру для каждого $l \in \{1, \dots, N_b\}$:

• сгенерировать случайную выборку $\mathbf{v}_{l,k}^{0*} = (v_{1,l,k}^{0*}, \dots, v_{n,l,k}^{0*})$ из $U(0; 1)^n$ случайного вектора;

• применяя интегральное условно-вероятностное преобразование (36) к $(v_{1,l,k}^{0*}, \dots, v_{n,l,k}^{0*})$, вычислить $\mathbf{h}_l^{0*} = (h_{1,l,k}^{0*}, \dots, h_{n,l,k}^{0*})$;

• в соответствии с (33) или (37), используя те же весовые функции, что и на шаге 5, и беря $(v_{1,l,k}^{0*}, \dots, v_{n,l,k}^{0*})$ и $(h_{1,l,k}^{0*}, \dots, h_{n,l,k}^{0*})$, вычислить $\hat{W}_{1,l,k}^{0*}$ или $\hat{W}_{2,l,k}^{0*}$ соответственно;

ж) вычислить $F_1(W_{1,k}^0) = \frac{1}{N_b + 1} \sum_{l=1}^{N_b} \mathbf{1}\{\hat{W}_{1,l,k}^{0*} > W_{1,k}^0\}$. $F_2(W_{2,k}^0)$ вычисляется аналогичным образом;

з) в соответствии с (34) и (35) вычислить $\hat{T}_{1,k}^0$ (относительно статистики Андерсона–Дарлинга см. (Berg, Bekken, 2007)); $\hat{T}_{2,k}^0$ вычисляется аналогично.

10. Для оценки p -значения для подхода $\mathcal{A}_1$ воспользоваться формулой:

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}\{\hat{T}_{1,k}^0 > \hat{T}_1\}.$$

9.2. Тесты, основанные на использовании эмпирических копула-функций

Подход $\mathcal{A}_2$

Подход $\mathcal{A}_2$, описанный в (Berg, 2009), основан на тесте, предложенном в работе (Genest, Rémillard, 2008). А именно, этот подход основан на использовании эмпирической копула-функции, впервые представленной в (Deheuvels, 1979):

$$\hat{\mathcal{C}}(\mathbf{u}) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}\{\hat{u}_{1t} \leq u_1, \dots, \hat{u}_{nt} \leq u_n\}, \quad (38)$$

где $\mathbf{u} = (u_1, \dots, u_n) \in [0; 1]^n$. Идея, лежащая в основе этого теста, состоит в сравнении $\hat{\mathcal{C}}(\cdot)$ с параметрической оценкой копула-функции $C_{\hat{\alpha}}(\cdot)$. Это классический подход в статистике, который состоит в построении одномерного критерия согласия, основанного на расстоянии между эмпирической функцией распределения и распределением, предполагаемым при нулевой гипотезе. С использованием статистики Крамера-фон Мизеса, в работе (Genest et al., 2009) получена следующая критическая статистика:

$$\hat{T}_2 = T \int_{[0,1]^d} (\hat{\mathcal{C}}(\hat{\mathbf{u}}) - C_{\hat{\alpha}}(\hat{\mathbf{u}}))^2 d\hat{\mathcal{C}}(\hat{\mathbf{u}}) = \sum_{t=1}^T (\hat{\mathcal{C}}(\hat{\mathbf{u}}_t) - C_{\hat{\alpha}}(\hat{\mathbf{u}}_t))^2. \quad (39)$$

Далее более подробное представлено описание параметрической бутстрап процедуры вычисления оценок p -значений при тестировании параметрической нулевой гипотезы для копула-функции.

Процедура 9.2. Реализация теста в подходе $\mathcal{A}_2$

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).
2. Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т. е. полупараметрического метода.
3. В соответствии с формулой (38) вычислить $\hat{\mathcal{C}}(\hat{\mathbf{u}})$.
4. Если имеется аналитическое выражение для $C_{\alpha}(\cdot)$, то, подставляя $\hat{\mathcal{C}}(\hat{\mathbf{u}})$ и $C_{\hat{\alpha}}(\hat{\mathbf{u}})$ в (39), вычислить оценку $\hat{T}_2$ и перейти к шагу 5.

Если аналитического выражения для $C_{\alpha}(\cdot)$ нет, то выбрать достаточно большое натуральное $N_b \geq T$ и реализовать следующие шаги:

- а) согласно предположению о справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_1^*, \dots, \mathbf{x}_{N_b}^*)$ и, согласно (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_1^*, \dots, \hat{\mathbf{u}}_{N_b}^*)$;
- б) вычислить аппроксимацию $C_{\hat{\alpha}}$ при помощи

$$C_{\hat{\alpha}}^*(\mathbf{u}) = \frac{1}{N_b + 1} \sum_{l=1}^{N_b} \mathbf{1}\{\hat{\mathbf{u}}_l^* \leq \mathbf{u}\}, \quad \mathbf{u} \in [0; 1]^n;$$

в) по формуле (39) вычислить приближение статистики Крамера–фон Мизеса:

$$\hat{T}_2 = \sum_{t=1}^T \left(\hat{C}(\hat{\mathbf{u}}_t) - C_{\hat{\alpha}}^*(\hat{\mathbf{u}}_t) \right)^2.$$

5. При некотором достаточно большом натуральном K для $k = 1, \dots, K$ последовательно повторить следующие шаги:

а) согласно предположению о справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

б) по полученным таким образом псевдо-наблюдениям оценить параметры α^0 копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha^0} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha^0)$, т. е. с использованием полупараметрического метода;

в) определить $\hat{C}_k^0(\mathbf{u})$ по формуле $\hat{C}_k^0(\mathbf{u}) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}\{\hat{\mathbf{u}}_{t,k}^0 \leq \mathbf{u}\}, \quad \mathbf{u} \in [0; 1]^n$;

г) если определено аналитическое выражение для $C_{\alpha}(\cdot)$, то положить:

$$\hat{T}_{2,k}^0 = \sum_{t=1}^T \left(\hat{C}_k^0(\hat{\mathbf{u}}_{t,k}^0) - C_{\hat{\alpha}_k^0}(\hat{\mathbf{u}}_{t,k}^0) \right)^2$$

и перейти к шагу 6; если аналитическое выражение для $C_{\alpha}(\cdot)$ не определено, то выбрать достаточно большое натуральное $N_b \geq T$ и реализовать следующие шаги:

- согласно предположению о справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^{0*}, \dots, \mathbf{x}_{N_b,k}^{0*})$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^{0*}, \dots, \hat{\mathbf{u}}_{N_b,k}^{0*})$;

- вычислить аппроксимацию $C_{\hat{\alpha}_k^0}$ при помощи:

$$C_{\hat{\alpha}_k^0}^{0*}(\mathbf{u}) = \frac{1}{N_b + 1} \sum_{l=1}^{N_b} \mathbf{1}\{\hat{\mathbf{u}}_{l,k}^{0*} \leq \mathbf{u}\};$$

- по формуле (39) вычислить приближение статистики Крамера–фон Мизеса:

$$\hat{T}_{2,k}^* = \sum_{t=1}^T \left(\hat{C}_k^0(\hat{\mathbf{u}}_{t,k}^0) - C_{\hat{\alpha}_k^0}^{0*}(\hat{\mathbf{u}}_{t,k}^0) \right)^2.$$

6. Для оценки p -значения в подходе $\mathcal{A}_2$ воспользоваться формулой

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}\{\hat{T}_{2,k}^0 > \hat{T}_2\}.$$

Некоторое расширение. В работе (Kole et al., 2007) предложен тест, который можно рассматривать в качестве расширения подхода $\mathcal{A}_2$ на случаи, когда аналитическое выражение (гауссовское, студентовское и т. п.) для копула-функции отсутствует.

В основе этого теста лежит та же идея, что и в (Genest et al., 2009), состоящая в вычислении расстояния между эмпирической и параметрической копула-функциями. Однако формальное представление функции распределения эллиптического типа, как правило, невозможно, и вычисление их значений с вычислительной точки зрения очень трудоемко в условиях большой размерности. В работе (Kole et al., 2007) используется свойство постоянства функции плотности эллиптических распределений на эллипсоидах. Каждой случайной величине, имеющей эллиптическое распределение, отвечает одномерная случайная величина со специфическим распределением, которое соответствует радиальным эллипсоидам, для которых плотность постоянна (см. (Fang et al., 1990)). Вместо того чтобы рассматривать исходные наблюдения, в работе (Kole et al., 2007) рассматриваются квадраты радиусов эллипсоида, отвечающего постоянной плотности. Как следствие, авторы последней упомянутой работы сравнивают эмпирические распределения квадратов радиусов с их теоретическими распределениями, которые представляют собой некоторые стандартные распределения в случае гауссовской и студентовской копула-функции.

Пусть имеется случайный вектор $U = (u_{1t}, \dots, u_{nt})'$, $t = 1, \dots, T$, с частными равномерными на $[0; 1]$ распределениями, зависимости между компонентами которого определяются гауссовской копула-функцией с корреляционной матрицей Σ . В работе (Kole et al., 2007) строится квадрат радиуса следующим образом:

$$Z_\Phi = \xi' \Sigma^{-1} \xi, \quad (40)$$

где $\xi = (\Phi^{-1}(u_{1t}), \dots, \Phi^{-1}(u_{nt}))'$ — вектор значений обратной функции гауссовского одномерного распределения в точках $u_{1t}, \dots, u_{nt}$. Легко показать, что случайная величина Z_Φ имеет χ_n^2 -распределение.

Пусть $V = (v_1, \dots, v_n)'$ — случайный вектор, в котором v_i имеет равномерное распределение на $[0; 1]$, а зависимости между компонентами совместного распределения определяются копула-функцией Стюдента с корреляционной матрицей Σ и числом степеней свободы ν . В (Kole et al., 2007) квадрат радиуса строится следующим образом:

$$Z_\nu = \xi_\nu' \Sigma^{-1} \xi_\nu / n, \quad (41)$$

где $\xi_\nu = (t_\nu^{-1}(v_1), \dots, t_\nu^{-1}(v_n))'$ — вектор значений обратной функции одномерного распределения Стюдента с числом степеней свободы ν . Случайная величина Z_ν имеет F -распределение с параметрами n и ν , т. к. ξ_ν имеет распределение Стюдента и может быть представлено как $\xi_\nu = W / \sqrt{S / n}$, где W — n -мерная нормальная случайная величина с ковариационной матрицей Σ , а S — одномерная случайная величина с χ_ν^2 -распределением. Как следствие, для Z_ν из (41) имеем:

$$Z_\nu = \frac{W' \Sigma^{-1} W / n}{S / \nu}.$$

Это представление показывает, что Z_v есть отношение двух независимых χ^2 -распределенных случайных величин, каждая из которых делится на соответствующее ей количество степеней свободы. Тем самым Z_v имеет $F_{n,v}$ -распределение.

Практическая реализация этого расширения подхода $\mathcal{A}_2$ аналогична реализации самого $\mathcal{A}_2$ и состоит из четырех шагов.

1. При помощи IFM-метода (Фантаццини, 2011b) оценить параметры α копула-функции следующим образом:

а) для моделирования частных распределений воспользоваться полупараметрическим методом из работы (Danielsson, de Vries, 2000), который состоит в том, чтобы моделировать центральную часть функции распределения при помощи эмпирического распределения, а хвосты — при помощи одномерных распределений из теории экстремальных значений; в частности, для моделирования хвостов можно использовать распределение Парето; соответственно, хвосты определяются левым эмпирическим 0.01-квантилем и правым эмпирическим 0.99-квантилем;

б) при помощи метода максимума правдоподобия оценить параметры α копула-функции.

2. Оценить качество выбранной параметрической копула-функции, вычисляя ее расстояние до эмпирической копула-функции; если копула-функция принадлежит эллиптическому семейству (например, гауссовская или студентовская), то с использованием (40) или (41) воспользоваться одним из стандартных критериев согласия. В работе (Kole et al., 2007) используются следующие меры расстояния:

$$\begin{aligned} D_{KS}^m &= \max_t |\mathfrak{F}_E - F_H|, \\ D_{KS}^a &= \int_x |\mathfrak{F}_E - F_H| dF_H, \\ D_{AD}^m &= \max_t \frac{|\mathfrak{F}_E - F_H|}{\sqrt{F_H(1-F_H)}}, \\ D_{AD}^a &= \int_x \frac{|\mathfrak{F}_E - F_H|}{\sqrt{F_H(1-F_H)}} dF_H, \end{aligned}$$

где $\mathfrak{F}_E$ — эмпирическая функция распределения, а F_H — гипотетическое распределение. Первая мера расстояния — расстояние Колмогорова–Смирнова, второе расстояние — среднее абсолютное отклонение, третья мера — расстояние Андерсона–Дарлингга, четвертое — аналог расстояния Андерсона–Дарлингга с усреднением. Для того чтобы уменьшить влияние выбросов при использовании расстояний Андерсона–Дарлингга, в работе (Kote et al., 2007), следуя (Malevergne, Sornette, 2003), исходное $|\mathfrak{F}_E - F_H|$ заменено на $(\mathfrak{F}_E - F_H)^2$.

3. Если форма меры расстояния известна, а параметры копула-функции оценены, для вычисления распределения случайных величин, отвечающих указанным выше четырем расстояниям, в работе (Kole et al., 2007) используется параметрическая бутстрап процедура с K итерациями, это отвечает в точности шагу 5 в реализации подхода $\mathcal{A}_2$. Отме-

тим, что двухшаговый IFM-метод также должен быть использован в упомянутой бутстрап процедуре.

4. Подобно шагу 6 в реализации подхода $\mathcal{A}_2$, для оценки p -значений использовать полученное на предыдущем шаге распределение для мер расстояний, т. е. подсчитать, сколько раз значения меры расстояния, вычисленные с использованием бутстрап выборок, превышают значение, отвечающее исходной выборке, и поделить это число на $K + 1$.

Подход $\mathcal{A}_3$

В работе (Genest et al., 2009) предложено применить предыдущий подход $\mathcal{A}_2$ к случайным величинам, подвергнутым преобразованию Розенблатта, т. е. к $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_T)$. Следующий шаг состоит в сравнении $\hat{\mathcal{C}}(\mathbf{v})$ с копула-функцией $C_{\perp}(\mathbf{v})$, отвечающей независимости, с использованием статистики Крамера–фон Мизеса, см. (Genest et al., 2009):

$$\hat{T}_3 = T \int_{[0,1]^d} (\hat{\mathcal{C}}(\mathbf{v}) - C_{\perp}(\mathbf{v}))^2 d\hat{\mathcal{C}}(\mathbf{v}) = \sum_{i=1}^T (\hat{\mathcal{C}}(\mathbf{v}_i) - C_{\perp}(\mathbf{v}_i))^2. \quad (42)$$

Процедура 9.3. Реализация теста в подходе $\mathcal{A}_3$

Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).

Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т. е. с использованием полупараметрического метода.

В предположении справедливости нулевой гипотезы относительно копула-функции C_{α} вычислить интегральное условно-вероятностное преобразование (СРПТ) выборочных данных $(\mathbf{v}_1, \dots, \mathbf{v}_T)$ путем применения СРПТ к $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$.

Согласно формуле (38) вычислить $\hat{\mathcal{C}}(\mathbf{v})$.

Согласно формуле (42) вычислить $\hat{T}_3$.

При некотором достаточно большом натуральном K для $k = 1, 2, \dots, K$ последовательно повторить следующие шаги:

а) отправляясь от справедливости нулевой гипотезы, сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

б) по сгенерированной таким образом псевдовыборке оценить параметры α^0 копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha^0} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha^0)$, т. е. с использованием полупараметрического метода;

в) в предположении справедливости нулевой гипотезы о параметрическом виде копула-функции $C_{\hat{\alpha}_k^0}$ вычислить интегральное условно-вероятностное преобразование выборочных данных $(\mathbf{v}_{1,k}^0, \dots, \mathbf{v}_{T,k}^0)$, применяя СРПТ к $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

г) определить $\hat{\mathcal{C}}_k^0(\mathbf{u}) = \frac{1}{T+1} \sum_{i=1}^T \mathbf{1}\{\mathbf{v}_{i,k}^0 \leq \mathbf{u}\}$, $\mathbf{u} \in [0, 1]^n$;

д) вычислить $\hat{T}_{3,k}^0 = \sum_{i=1}^T (\hat{\mathcal{C}}_k^0(\mathbf{v}_{i,k}^0) - C_{\perp}(\mathbf{v}_{i,k}^0))^2$.

Вычислить оценку p -значения следующим образом:

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}_{\{\hat{T}_{3,k}^0 > \hat{T}_3\}}.$$

Подход $\mathcal{A}_4$

В работах (Genest, Rivest, 1993; Wang, Wells, 2000; Savu, Trede, 2008; Genest et al., 2006a) для построения критериев согласия для копула-функций предлагается использовать так называемую функцию зависимости Кендалла $K(w) = P(C(\hat{\mathbf{U}}) \leq w)$ (см. (Фантаццини, 2011b)). Тестовый датчик S_4 для подхода $\mathcal{A}_4$ определяется в форме этой функции, т. е.

$$S_4(w) = P(C(\hat{\mathbf{U}}) \leq w), \quad w \in [0; 1],$$

где $\hat{\mathbf{U}}$ — псевдовектор, каждая компонента которого составлена из соответствующих нормированных рангов. При справедливости нулевой гипотезы $S_4(w) = S_{4,\hat{\alpha}}(w)$, где $S_{4,\hat{\alpha}}(w)$ — функция зависимости Кендалла, рассчитанная в условиях нулевой гипотезы H_0 и зависящая от параметров соответствующей копула-функции. Непараметрическая оценка для тестового датчика S_4 равна:

$$\hat{S}_4(w) = \frac{1}{T+1} \sum_{i=1}^T \mathbf{1}_{\{\hat{\mathcal{C}}(\hat{\mathbf{u}}_i) \leq w\}}. \quad (43)$$

Статистика Крамера–фон Мизеса вычисляется следующим образом:

$$\hat{T}_4 = T \int_0^1 (\hat{S}_4(w) - S_{4,\hat{\alpha}}(w))^2 d\hat{S}_4(w) = \sum_{i=1}^T \left(\hat{S}_4\left(\frac{t}{T+1}\right) - S_{4,\hat{\alpha}}\left(\frac{t}{T+1}\right) \right)^2. \quad (44)$$

Процедура 9.4. Реализация теста в подходе $\mathcal{A}_4$

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).

2. Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т. е. полупараметрического метода.

3. В соответствии с формулой (38) вычислить $\hat{\mathcal{C}}(\hat{\mathbf{u}})$.

4. Если для $S_{4,\alpha}$ можно получить аналитическое выражение, то в соответствии с (43) и (44) вычислить статистику $\hat{T}_4$, а затем перейти к шагу 5. Если аналитическое выражение для $S_{4,\alpha}$ получить невозможно, то выбрать $N_b \geq T$ и реализовать следующие шаги:

а) согласно предположению о справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_1^*, \dots, \mathbf{x}_{N_b}^*)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_1^*, \dots, \hat{\mathbf{u}}_{N_b}^*)$;

б) вычислить аппроксимацию $S_{4,\hat{\alpha}}$ при помощи $\hat{S}_4^*(w) = \frac{1}{N_b+1} \sum_{i=1}^{N_b} \mathbf{1}_{\{\hat{\mathcal{C}}^*(\hat{\mathbf{u}}_i^*) \leq w\}}$, где

$$\hat{\mathcal{C}}^*(\mathbf{u}) = \frac{1}{N_b+1} \sum_{i=1}^{N_b} \mathbf{1}_{\{\hat{\mathbf{u}}_i^* \leq \mathbf{u}\}}, \quad \mathbf{u} \in [0; 1]^n;$$

в) по формуле (44) вычислить приближение для статистики Крамера–фон Мизеса:

$$\hat{T}_4 = \frac{T}{N_b} \sum_{l=1}^{N_b} \left(\hat{S}_4 \left(\hat{\mathcal{C}}^* (\hat{\mathbf{u}}_l^*) \right) - \hat{S}_4^* \left(\hat{\mathcal{C}}^* (\hat{\mathbf{u}}_l^*) \right) \right)^2.$$

5. При некотором достаточно большом натуральном K для $k = 1, \dots, K$ последовательно повторить следующие шаги:

а) согласно предположению нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

б) оценить параметры α^0 копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha^0} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha^0)$, т.е. с использованием полупараметрического метода;

в) определить $\hat{S}_{4,k}^0(w) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}\{\hat{\mathcal{C}}_k^0(\hat{\mathbf{u}}_{t,k}^0 \leq w)\}$, где $\hat{\mathcal{C}}_k^0(\mathbf{u}) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}\{\hat{\mathbf{u}}_{t,k}^0 \leq \mathbf{u}\}$, $\mathbf{u} \in [0; 1]^n$;

г) если для $S_{4,\alpha}$ можно получить аналитическое выражение, то с использованием $\hat{S}_{4,k}^0$ и $S_{4,\hat{\alpha}_k^0}$ в формуле (44) вычислить $\hat{T}_{4,k}^0$ и перейти к шагу б; если аналитическое выражение для $S_{4,\alpha}$ получить невозможно, тогда выбрать достаточно большое натуральное $N_b \geq T$ и реализовать следующие шаги:

• в условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^{0*}, \dots, \mathbf{x}_{N_b,k}^{0*})$ и в соответствии с (32) вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^{0*}, \dots, \hat{\mathbf{u}}_{N_b,k}^{0*})$;

• вычислить аппроксимацию $S_{4,\hat{\alpha}_k^0}$ по формуле: $\hat{S}_{4,k}^{0*}(w) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} \mathbf{1}\{\hat{\mathcal{C}}_k^{0*}(\hat{\mathbf{u}}_{l,k}^{0*}) \leq w\}$, где $\hat{\mathcal{C}}_k^{0*}(\mathbf{u}) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} \mathbf{1}\{\hat{\mathbf{u}}_{l,k}^{0*} \leq \mathbf{u}\}$, $\mathbf{u} \in [0; 1]^n$;

• по формуле (44) вычислить приближение для статистики Крамера–фон Мизеса:

$$\hat{T}_{4,k}^0 = \frac{T}{N_b} \sum_{l=1}^{N_b} \left(\hat{S}_{4,k}^0 \left(\hat{\mathcal{C}}_k^{0*}(\hat{\mathbf{u}}_{l,k}^{0*}) \right) - \hat{S}_{4,k}^{0*} \left(\hat{\mathcal{C}}_k^{0*}(\hat{\mathbf{u}}_{l,k}^{0*}) \right) \right)^2.$$

6. Вычислить оценку p -значения с использованием формулы:

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}\{\hat{T}_{4,k}^0 > \hat{T}_4\}.$$

Подход $\mathcal{A}_5$

В работе (Quessy et al., 2007) предлагается критерий согласия для двумерных копула-функций, который основан на функции зависимости Спирмена $L_2(w) = P(\hat{U}_1 \hat{U}_2 \leq w)$. При этом $P(\hat{U}_1 \hat{U}_2 \leq w) = P(C_\perp(\hat{U}_1 \hat{U}_2) \leq w)$. Этот подход обобщается на случай произвольной размерности n , при этом $L_n(w) = P(C_\perp(\hat{\mathbf{U}}) \leq w)$ и тестовый датчик S_5 подхода $\mathcal{A}_5$ определяется соотношением:

$$S_5(w) = P(C_\perp(\hat{\mathbf{U}}) \leq w), \quad w \in [0; 1],$$

где $\hat{\mathbf{U}}$ — псевдовектор, каждая компонента которого составлена из соответствующих нормированных рангов. При справедливости нулевой гипотезы (H_0) , $S_5(w) = S_{5,\hat{\alpha}}(w)$, где

$S_{5,\hat{\alpha}}(w)$ — функция зависимости Кендалла, рассчитанная при справедливости H_0 и зависящая от параметров соответствующей копула-функции. Возьмем непараметрическую оценку для S_5 :

$$\hat{S}_5(w) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}\{C_{\perp}(\hat{\mathbf{u}}_t) \leq w\}. \quad (45)$$

Статистика Крамера–фон Мизеса вычисляется следующим образом:

$$\hat{T}_5 = T \int_0^1 \left(\hat{S}_5(w) - S_{5,\hat{\alpha}}(w) \right)^2 d\hat{S}_5(w) = \sum_{t=1}^T \left(\hat{S}_5\left(\frac{t}{T+1}\right) - S_{5,\hat{\alpha}}\left(\frac{t}{T+1}\right) \right)^2. \quad (46)$$

Процедура 9.5. Реализация теста в подходе A_5

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).

2. Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т. е. полупараметрического метода.

3. Если для $S_{5,\alpha}$ существует аналитическое выражение, то, используя формулы (45) и (46), вычислить $\hat{T}_{5,k}^0$ и перейти к шагу 4. Если аналитическое выражение для $S_{5,\alpha}$ получить невозможно, тогда выбрать достаточно большое натуральное $N_b \geq T$ и реализовать следующие шаги:

а) согласно предположению о справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_1^*, \dots, \mathbf{x}_{N_b}^*)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_1^*, \dots, \hat{\mathbf{u}}_{N_b}^*)$;

б) вычислить аппроксимацию $S_{5,\hat{\alpha}}$ при помощи: $\hat{S}_5^*(w) = \frac{1}{N_b+1} \sum_{l=1}^{N_b} \mathbf{1}\{C_{\perp}(\hat{\mathbf{u}}_l^*) \leq w\}$;

в) по формуле (46) вычислить приближение для статистики Крамера–фон Мизеса:

$$\hat{T}_5 = \frac{T}{N_b} \sum_{l=1}^{N_b} \left(\hat{S}_5(C_{\perp}(\hat{\mathbf{u}}_l^*)) - \hat{S}_5^*(C_{\perp}(\hat{\mathbf{u}}_l^*)) \right)^2.$$

4. При некотором достаточно большом натуральном K для $k = 1, \dots, K$ последовательно повторить следующие шаги:

а) в условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

б) оценить параметры α^0 копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha^0} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha^0)$, т. е. полупараметрического метода;

в) определить $\hat{S}_{5,k}^0(w) = \frac{1}{T+1} \sum_{t=1}^T \mathbf{1}\{C_{\perp}(\hat{\mathbf{u}}_{t,k}^0) \leq w\}$;

г) если для $S_{5,\alpha}$ существует аналитическое выражение, то с использованием $\hat{S}_{5,k}^0$ и $S_{5,\hat{\alpha}_k^0}$ по формуле (46) вычислить $\hat{T}_{5,k}^0$ и перейти к шагу 5; если аналитическое выражение для $S_{5,\alpha}$

получить невозможно, то выбрать достаточно большое натуральное $N_b \geq T$ и реализовать следующие шаги:

- в условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^{0*}, \dots, \mathbf{x}_{N_b,k}^{0*})$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^{0*}, \dots, \hat{\mathbf{u}}_{N_b,k}^{0*})$;
- вычислить аппроксимацию $S_{5,\hat{\alpha}_k^0}$ при помощи формулы

$$\hat{S}_{5,k}^{0*}(w) = \frac{1}{N_b + 1} \sum_{l=1}^{N_b} \mathbf{1}\{C_{\perp}(\hat{\mathbf{u}}_{l,k}^{0*}) \leq w\};$$

- по формуле (46) вычислить приближение для статистики Крамера–фон Мизеса

$$\hat{T}_{5,k}^0 = \frac{T}{N_b} \sum_{l=1}^{N_b} \left(\hat{S}_{5,k}^0(C_{\perp}(\hat{\mathbf{u}}_{l,k}^{0*})) - \hat{S}_{5,k}^{0*}(C_{\perp}(\hat{\mathbf{u}}_{l,k}^{0*})) \right)^2.$$

5. Вычислить оценку p -значения по формуле

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}\{\hat{T}_{5,k}^0 > \hat{T}_5\}.$$

9.3. Другие подходы к построению критериев согласия

Подход $\mathcal{A}_6$

В работе (Shih, 1998) предложен основанный на моментных тождествах критерий согласия для двумерных моделей, задаваемых копула-функцией Клейтона. В ней рассматриваются невзвешенные и взвешенные оценки параметра зависимости α с использованием τ -Кендалла и взвешенной ранговой оценки, а именно:

$$\hat{\alpha}_{\tau} = \frac{2\hat{\tau}}{1-\hat{\tau}} \quad \text{и} \quad \hat{\alpha}_w = \frac{\sum_{i < j} \Delta_{ij} / W_{ij}}{\sum_{i < j} (1 - \Delta_{ij}) / W_{ij}}, \quad (47)$$

где $\hat{\tau} = -1 + 4 \sum_{i < j} \Delta_{ij} / T(T-1)$, $\Delta_{ij} = \mathbf{1}\{(\hat{U}_{1i} - \hat{U}_{1j})(\hat{U}_{2i} - \hat{U}_{2j}) > 0\}$,

а $W_{ij} = \sum_{t=1}^T \mathbf{1}\{\hat{U}_{1t} \leq \max(\hat{U}_{1i}, \hat{U}_{1j}), \hat{U}_{2t} \leq \max(\hat{U}_{2i}, \hat{U}_{2j})\}$.

Учитывая, что $\hat{\alpha}_{\tau}$ и $\hat{\alpha}_w$ являются несмещенными оценками для α , в предположении справедливости нулевой гипотезы ($C = C_{\alpha}$ для $\alpha \geq 0$) в (Shih, 1998) была предложена следующая критическая статистика:

$$\hat{T}_{\text{Shih}} = \sqrt{T}(\hat{\alpha}_{\tau} - \hat{\alpha}_w).$$

В (Shih, 1998) показано, что при справедливости нулевой гипотезы приведенная выше статистика имеет асимптотически нормальное распределение. Однако позднее в (Genest et al.,

2006b) было указано, что представленная в (Shih, 1998) формула для дисперсии этого асимптотического нормального распределения неверна, и предложен исправленный вариант этой формулы. В работе (Berg, 2009) предлагается расширить этот подход на случай произвольной размерности n путем сравнения $\hat{\alpha}_\tau$ и $\hat{\alpha}_W$ для каждой пары случайных величин. Как следствие, результирующий вектор, составленный из $n(n-1)/2$ статистик, будет иметь асимптотическое $n(n-1)/2$ -мерное нормальное распределение с невырожденной ковариационной матрицей, формула для которой пока не выведена. Если умножить вектор статистик на матрицу, обратную к квадратному корню ковариационной матрицы, то результирующий нормированный вектор имеет асимптотическое стандартное нормальное распределение, а сумма квадратов нормированных статистик будет иметь χ^2 -распределение с $n(n-1)/2$ степенями свободы.

Не имея формулы ковариационной матрицы, Berg (2009) вычисляет ненормированные суммы квадратов и использует параметрическую бутстрап процедуру для оценки p -значения. Критическая статистика для подхода $\mathcal{A}_6$ равна:

$$\hat{T}_6 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n (\hat{\alpha}_{\tau,ij} - \hat{\alpha}_{W,ij})^2. \quad (48)$$

Отметим, что $\hat{\alpha}_W$ и в целом подход $\mathcal{A}_6$ может быть использован только для тестирования копула-функции Клейтона.

Процедура 9.6. Реализация теста в подходе $\mathcal{A}_6$

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).
2. Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т. е. полупараметрического метода.
3. В соответствии с (47) вычислить параметры $\hat{\alpha}_\tau$ и $\hat{\alpha}_W$.
4. В соответствии с (48) вычислить $\hat{T}_6$.
5. При некотором достаточно большом натуральном K для $k = 1, \dots, K$ последовательно повторить следующие шаги:

- а) в условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;
- б) в соответствии с (47) вычислить $\hat{\alpha}_{\tau,k}^0$ и $\hat{\alpha}_{W,k}^0$;
- в) в соответствии с (48) и используя полученные на предыдущем шаге $\hat{\alpha}_{\tau,k}^0$ и $\hat{\alpha}_{W,k}^0$, вычислить $\hat{T}_{6,k}^0$.

Вычислить оценку p -значения, используя формулу

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}\{\hat{T}_{6,k}^0 > \hat{T}_6\}.$$

Подход $\mathcal{A}_7$

В работе (Panchenko, 2005) предложен тест, основанный на скалярном произведении $\hat{\mathbf{U}}$ и $\hat{\mathbf{U}}_{\hat{\alpha}}$, где $\hat{\mathbf{U}}$ — псевдовектор, каждая компонента которого равна соответствующему нор-

мированному рангу, а $\hat{\mathbf{U}}_{\hat{\alpha}}$ — аналог вектора $\hat{\mathbf{U}}$, отвечающий нулевой гипотезе (при этом $\hat{\alpha}$ — состоятельная оценка параметра копула-функции). Идея этого теста состоит в использовании скалярного произведения в качестве меры расстояния между двумя векторами. Определим квадрат расстояния Q между двумя векторами формулой

$$Q = \langle \hat{\mathbf{U}} - \hat{\mathbf{U}}_{\hat{\alpha}} | \kappa_n | \hat{\mathbf{U}} - \hat{\mathbf{U}}_{\hat{\alpha}} \rangle^5,$$

где κ_n — положительно определенное симметричное ядро; в частности, в качестве κ_n может быть взято гауссовское ядро:

$$\kappa_n(\hat{\mathbf{U}}, \hat{\mathbf{U}}') = \exp \{ -\| \hat{\mathbf{U}} - \hat{\mathbf{U}}' \|^2 / (2nh^2) \},$$

где $\| \cdot \|$ обозначает евклидову норму в R^n ; $h > 0$ — ширина «окна». Очевидно, что Q будет равно нулю тогда и только тогда, когда $\hat{\mathbf{U}} = \hat{\mathbf{U}}_{\hat{\alpha}}$.

Если имеются случайные выборки $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ из $\hat{\mathbf{U}}$, то следующим шагом является генерация случайных выборок $(\hat{\mathbf{u}}_1^*, \dots, \hat{\mathbf{u}}_T^*)$ вектора $\hat{\mathbf{U}}_{\hat{\alpha}}$, отвечающего нулевой гипотезе. Скалярное произведение Q может быть представлено в виде $Q = Q_{11} - 2Q_{12} + Q_{22}$, а каждое слагаемое этого разложения может быть оценено с использованием V-статистики (в качестве введения в теорию U- и V-статистик см. (Denker, Keller, 1983)). Критическая статистика подхода $\mathcal{A}_7$ имеет вид:

$$\hat{T}_7 = \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T \kappa_n(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) - \frac{2}{T^2} \sum_{i=1}^T \sum_{j=1}^T \kappa_n(\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j^*) + \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T \kappa_n(\hat{\mathbf{u}}_i^*, \hat{\mathbf{u}}_j^*). \quad (49)$$

Процедура 9.7. Реализация теста в подходе $\mathcal{A}_7$

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).
2. Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т.е. полупараметрического метода.
3. В условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_1^*, \dots, \mathbf{x}_T^*)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_1^*, \dots, \hat{\mathbf{u}}_T^*)$.
4. В соответствии с (49), используя $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ и $(\hat{\mathbf{u}}_1^*, \dots, \hat{\mathbf{u}}_T^*)$, вычислить $\hat{T}_7$.
5. При некотором достаточно большом натуральном K для $k = 1, \dots, K$ последовательно повторить следующие шаги:
 - а) в условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;
 - б) оценить параметры α^0 копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha)$, т.е. полупараметрического метода;

⁵ Определение и подробности см. в (Panchenko, 2005).

в) используя копула-функцию $C_{\hat{\alpha}_k^0}$, отвечающую нулевой гипотезе, сгенерировать случайную выборку $(\mathbf{x}_{1,k}^{0*}, \dots, \mathbf{x}_{T,k}^{0*})$ и, используя (32), вычислить ассоциированную с ней псевдо-выборку $(\hat{\mathbf{u}}_{1,k}^{0*}, \dots, \hat{\mathbf{u}}_{T,k}^{0*})$;

г) в соответствии с (49) и с использованием $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$ и $(\hat{\mathbf{u}}_{1,k}^{0*}, \dots, \hat{\mathbf{u}}_{T,k}^{0*})$ вычислить $\hat{T}_{7,k}^0$.

6. Вычислить оценку p -значения по формуле

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}\{\hat{T}_{7,k}^0 > \hat{T}_7\}.$$

Подход $\mathcal{A}_8$

В работе (Berg, 2009) предложено расширение подхода $\mathcal{A}_7$, которое состоит в применении подхода $\mathcal{A}_7$ к величинам, подвергнутым преобразованию Розенблатта, т.е. к $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_T)$. Если имеются выборки $(\mathbf{v}_1^*, \dots, \mathbf{v}_T^*)$, извлеченные из копула-функции, отвечающей случаю независимости, статистика подхода $\mathcal{A}_8$ равна:

$$\hat{T}_8 = \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T \kappa_n(\mathbf{v}_i, \mathbf{v}_j) - \frac{2}{T^2} \sum_{i=1}^T \sum_{j=1}^T \kappa_n(\mathbf{v}_i, \mathbf{v}_j^*) + \frac{1}{T^2} \sum_{i=1}^T \sum_{j=1}^T \kappa_n(\mathbf{v}_i^*, \mathbf{v}_j^*). \quad (50)$$

В работе (Berg, 2009) отмечается, что может показаться странным желание строить выводы относительно отклонения от проверяемой (нулевой) гипотезы на основании единственной выборки, отвечающей нулевой гипотезе. Тем не менее, этот подход изначально рассматривался в работе (Panchenko, 2005), и Берг (2009) предпочел рассматривать этот тест именно в таком виде. Более того, подход $\mathcal{A}_8$ направлен на то, чтобы проверить эффект преобразования Розенблатта в случае использования подхода $\mathcal{A}_7$.

Процедура 9.8. Реализация теста в подходе $\mathcal{A}_8$

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).

2. Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т.е. полупараметрического метода.

3. В предположении справедливости нулевой гипотезы относительно копула-функции $C_{\hat{\alpha}}$ вычислить интегральное условно-вероятностное преобразование (СРП) выборочных данных $(\mathbf{v}_1, \dots, \mathbf{v}_T)$ путем применения СРП к $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$.

4. Сгенерировать выборку $(\mathbf{v}_1^*, \dots, \mathbf{v}_T^*)$ из копула-функции, отвечающей независимости.

5. В соответствии с (50) и с использованием $(\mathbf{v}_1, \dots, \mathbf{v}_T)$ и $(\mathbf{v}_1^*, \dots, \mathbf{v}_T^*)$ вычислить $\hat{T}_8$.

6. При некотором достаточно большом натуральном K для $k = 1, \dots, K$ последовательно повторить следующие шаги:

а) в условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдovyборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

б) оценить параметры α^0 копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha^0)$, т.е. полупараметрического метода;

в) в предположении справедливости нулевой гипотезы относительно копула-функции $C_{\alpha_k^0}$ вычислить интегральное условно-вероятностное преобразование (СПИТ) выборочных данных $(\mathbf{v}_{1,k}^0, \dots, \mathbf{v}_{T,k}^0)$ путем применения СПИТ к $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

г) используя копула-функцию, отвечающую независимости, сгенерировать случайную выборку $(\mathbf{v}_{1,k}^{0*}, \dots, \mathbf{v}_{T,k}^{0*})$;

д) в соответствии с (50), используя $(\mathbf{v}_{1,k}^0, \dots, \mathbf{v}_{T,k}^0)$ и $(\mathbf{v}_{1,k}^{0*}, \dots, \mathbf{v}_{T,k}^{0*})$, вычислить $\hat{T}_{8,k}^0$.

7. Вычислить оценку p -значения по формуле

$$\hat{p} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}_{\{\hat{T}_{8,k}^0 > \hat{T}_8\}}.$$

9.4. Тесты, основанные на усреднении критических статистик

Выше были рассмотрены некоторые критерии согласия для копула-функций. Справедливо задаться вопросом, а может ли усреднение этих тестов (с разными способами измерения отклонений от нулевой гипотезы) дать более устойчивые результаты? Ответ на этот вопрос дается в работе (Berg, 2009), где подход с усреднением обозначен как $\mathcal{A}_9$.

Понятно, что усреднение должно осуществляться по стандартизированным переменным. Более того, следует использовать оптимальные веса для расчета взвешенного среднего. Тем не менее, в силу некоторых вычислительных аспектов, в работе (Berg, 2009) представлены только два варианта усреднения:

первый вариант состоит в усреднении всех девяти указанных выше подходов; ему отвечает статистика:

$$\hat{T}_9^{(a)} = \frac{1}{9} \left\{ \hat{T}_1^{(a)} + \hat{T}_1^{(b)} + \sum_{j=2}^8 \hat{T}_j \right\}; \quad (51)$$

второй вариант состоит в усреднении лишь тех подходов, которые основаны на использовании эмпирической копула-функции, а именно подходов $\mathcal{A}_2$, $\mathcal{A}_3$ и $\mathcal{A}_4$; статистика для такого усреднения равна:

$$\hat{T}_9^{(b)} = \frac{1}{3} (\hat{T}_2 + \hat{T}_3 + \hat{T}_4). \quad (52)$$

Процедура 9.9. Реализация теста в подходе $\mathcal{A}_9$

1. Получить псевдо-наблюдения $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ путем преобразования выборочных данных $(\mathbf{x}_1, \dots, \mathbf{x}_T)$ в нормированные ранги с использованием (32).

2. Оценить параметры α копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha} = \arg \max_{\alpha} l(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T; \alpha)$, т. е. с использованием полупараметрического метода.

3. Используя $(\hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_T)$ и $\hat{\alpha}$, реализовать необходимые шаги для вычисления критических статистик процедур 9.1–9.8 и вычислить $\hat{T}_1^{(a)}$, $\hat{T}_1^{(b)}$, $\hat{T}_2 - \hat{T}_8$.

4. В соответствии с (51)–(52) вычислить $\hat{T}_9^{(a)}$ и $\hat{T}_9^{(b)}$ соответственно.

5. При некотором достаточно большом натуральном K для $k = 1, \dots, K$ последовательно повторить следующие шаги:

а) в условиях справедливости нулевой гипотезы сгенерировать случайную выборку $(\mathbf{x}_{1,k}^0, \dots, \mathbf{x}_{T,k}^0)$ и, используя (32), вычислить ассоциированную с ней псевдовыборку $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$;

б) оценить параметры α^0 копула-функции с использованием состоятельной ММП-оценки $\hat{\alpha}_k^0 = \arg \max_{\alpha^0} l(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0; \alpha^0)$, т. е. с использованием полупараметрического метода;

в) используя $(\hat{\mathbf{u}}_{1,k}^0, \dots, \hat{\mathbf{u}}_{T,k}^0)$ и $\hat{\alpha}_k^0$, реализовать необходимые шаги для вычисления критических статистик из процедур 9.1–9.8 и вычислить $\hat{T}_{1,k}^{0,(a)}, \hat{T}_{1,k}^{0,(b)}, \hat{T}_{2,k}^0 - \hat{T}_{8,k}^0$;

г) в соответствии с (51)–(52), используя $\hat{T}_{1,k}^{0,(a)}, \hat{T}_{1,k}^{0,(b)}, \hat{T}_{2,k}^0 - \hat{T}_{8,k}^0$, вычислить $\hat{T}_{9,k}^{0,(a)}$ и $\hat{T}_{9,k}^{0,(b)}$.

Вычислить оценку p -значений по формулам:

$$\hat{p}^{(a)} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}_{\{\hat{T}_{9,k}^{0,(a)} > \hat{T}_9^{(a)}\}}, \quad \hat{p}^{(b)} = \frac{1}{K+1} \sum_{k=1}^K \mathbf{1}_{\{\hat{T}_{9,k}^{0,(b)} > \hat{T}_9^{(b)}\}}.$$

Сравнение подходов с помощью имитационного моделирования. В работе (Berg, 2009) проведено объемное имитационное моделирование с целью сравнения описанных выше подходов к построению критериев согласия для копула-функций. И хотя показано, что не существует критерия, превосходящего по мощности любой другой из рассматриваемых подходов, подходы $\mathcal{A}_2$, $\mathcal{A}_4$ и $\mathcal{A}_9$ дают очень хорошие результаты, причем последний из них можно признать лучшим.

Интересно, что при тестировании гипотезы гауссовости (при альтернативе распределения с «тяжелыми хвостами») слабый в других ситуациях подход $\mathcal{A}_1$ дает очень хорошие результаты для высоких размерностей и больших объемов выборок. Следовательно, его целесообразно использовать в качестве предварительного теста проверки эллипсоидальности распределения с тем, чтобы решить, какой из подходов выбрать для дальнейшего исследования, см. (Huffera, Park, 2007).

9.5. Эмпирические приложения с пакетом R : пример реализации подхода $\mathcal{A}_2$

Модули пакета R дают возможность численно реализовать критерии согласия, основанные на сравнении эмпирической копула-функции с ее параметрической оценкой (в предположении справедливости проверяемой гипотезы), т. е. практически осуществить подход $\mathcal{A}_2$. Критическая статистика строится с использованием расстояния Крамера–фон Мизеса. Приближения для p -значений критической статистики вычисляются с использованием параметрических бутстрап-процедур, предложенных в (Genest, Remillard, 2008; Genest et al., 2009), или с помощью подхода, изложенного в (Kojadinovic et al., 2011; Kojadinovic, Yan, 2011). Ниже приведен код процедуры в соответствующем модуле пакета R .

```

library(copula)
x <- rcopula(claytonCopula(3), 200)
## Does the Gumbel family seem to be a good choice?
gofCopula(gumbelCopula(1), x)
## What about the Clayton family?
gofCopula(claytonCopula(1) , x)
## The same with a different estimation method
gofCopula(gumbelCopula(1), x, method="itau")
gofCopula(claytonCopula(1), x, method="itau")

## A three-dimensional example
x <- rcopula(tCopula(c(0.5, 0.6, 0.7), dim = 3, dispstr = "un"),200)
## Does the Clayton family seem to be a good choice?
gofCopula(gumbelCopula(1, dim =3), x)
## What about the t copula?
t.copula <- tCopula(rep(0, 3), dim = 3, dispstr = "un", df.fixed=TRUE)
gofCopula(t.copula, x)
## The same with a different estimation method
gofCopula(gumbelCopula(1, dim =3), x, method="itau")
gofCopula(t.copula, x, method="itau")
## The same using the multiplier approach
gofCopula(gumbelCopula(1, dim =3), x, simulation="mult")
gofCopula(t.copula, x, simulation="mult")

```

For sake of space, we report only the output relative to the first two goodness-of-fit tests:

```

> gofCopula(gumbelCopula(1), x)
Progress will be displayed every 100 iterations.
Iteration 100
...
Iteration 1000
$statistic [1] 0.2648942
$pvalue [1] 0.0004995005
$parameters [1] 1.980118
> gofCopula(claytonCopula(1), x)
Progress will be displayed every 100 iterations.
Iteration 100
...
Iteration 1000
$statistic [1] 0.01489807
$pvalue [1] 0.54995
$parameters [1] 3.317709

```

Отметим, что в пакете, созданном Даниэлем Бергом, рассмотрен более широкий диапазон критериев согласия для копула-функций. Однако к моменту написания статьи (июль 2011 г.) существовала только предварительная версия этого пакета.

Список литературы

- Фантаццини Д. (2011a). Моделирование многомерных распределений с использованием копула-функций. I. *Прикладная эконометрика*, 22 (2), 98–134.
- Фантаццини Д. (2011b). Моделирование многомерных распределений с использованием копула-функций. II. *Прикладная эконометрика*, 23 (3), 98–132.
- Anderson T. W., Darling D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49, 765–769.
- Berg D. (2009). Copula goodness-of-fit testing: An overview and power comparison. *The European Journal of Finance*, 15, 675–701.
- Berg D., Bakken H. (2007). A goodness-of-fit test for copulas based on the probability integral transform. *Technical report*, University of Oslo, Statistical research report no. 10.
- Breymann W., Dias A., Embrechts P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3, 1–14.
- 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.
- Chen X., Fan Y., Patton A. (2004). Simple tests for models of dependence between multiple financial time series, with applications to U. S. equity returns and exchange rates. London Economics Financial Markets Group, *Working Paper n. 483*.
- D'Agostino R. B., Stephens M. A. (1986). *Goodness-of-fit techniques*. New York: Marcel Dekker Inc.
- Danielsson J., de Vries C. G. (2000). Value-at-risk and extreme returns. *Annales d'Economie et de Statistique*, 60, 239–270.
- David H. A. (1981). *Order statistics*. 2 ed., Wiley.
- Deheuvels P. (1979). La fonction de dépendance empirique et ses propriétés: Un test non paramétrique d'indépendance. *Bulletin de l'Académie Royale de Belgique, Classe des Sciences*, 5e série, 65, 274–292.
- Deheuvels P. (1984). The characterization of distributions by order statistics and record values: A unified approach. *Annals of Eugenics*, 21 (2), 326–334.
- Denker M., Keller G. (1983). On U-statistics and v. Mises' statistics for weakly dependent processes. *Probability Theory and Related Fields*, 64 (4), 505–522.
- Dias A., Embrechts P. (2004). Change-point analysis for dependence structures in finance and insurance. In: *Risk Measures for the 21st Century*, ed. by Giorgio Szegoe, 321–335. Wiley Finance Series.
- Dobrić J., Schmid F. (2005). Testing goodness of fit for parametric families of copulas — application to financial data. *Communications in Statistics: Simulation and Computation*, 34 (4), 1053–1068.
- Dobrić J., Schmid F. (2007). A goodness of fit test for copulas based on Rosenblatt's transformation. *Computational Statistics and Data Analysis*, 51 (9), 4633–4642.
- Fang K. T., Kotz S., Ng K. (1990). Symmetric multivariate and related distributions. In: *Monographs on Statistics and Applied Probability*, vol. 36. Chapman and Hall: London.
- Fermanian J. (2005). Goodness of fit tests for copulas. *Journal of Multivariate Analysis*, 95, 119–152.
- Genest C., Rivest L. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88, 1034–1043.
- Genest C., Rémillard B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semi-parametric models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 44 (6), 1096–1127.

Genest C., Quessy J. F., Rémillard B. (2006a). Goodness-of-fit procedures for copula models based on the probability integral transform. *Scandinavian Journal of Statistics*, 33, 337–366.

Genest C., Quessy J. F., Rémillard B. (2006b). On the joint asymptotic behavior of two rank-based estimators of the association parameter in the gamma frailty model. *Statistics and Probability Letters*, 76, 10–18.

Genest C., Rémillard B., Beaudoin D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44 (2), 199–213.

Glen A. G., Leemis L. M., Barr D. R. (2001). Order statistics in goodness-of-fit testing. *IEEE Transactions on Reliability*, 50 (2), 209–213.

Hong Y., Li H. (2005). Nonparametric specification testing for continuous-time models with application to spot interest rates. *Review of Financial Studies*, 18, 37–84.

Huard D., Ivin G., Favre A. C. (2006). Bayesian copula selection. *Computational Statistics and Data Analysis*, 51 (2), 809–822.

Huffera F. W., Park C. (2007). A test for elliptical symmetry. *Journal of Multivariate Analysis*, 98 (2), 256–281.

Jaynes E. T., Bretthorst G. L. (2003). *Probability theory: The logic of science*. Cambridge University Press.

Kojadinovic I., Yan J. M., Holmes M. (2011). Fast large-sample goodness-of-fit tests for copulas. *Statistica Sinica*, 21 (2), 841–871.

Kojadinovic I., Yan J. (2011). A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. *Statistics and Computing*, 21 (1), 17–30.

Kole E., Koedijk K., Verbeek M. (2007). Selecting copulas for risk management. *Journal of Banking and Finance*, 31, 2405–2423.

Malevergne Y., Sornette D. (2003). Testing the gaussian copula hypothesis for financial assets dependence. *Quantitative Finance*, 3, 231–250.

Marsaglia G., Marsaglia J. (2004). Evaluating the Anderson–Darling distribution. *Journal of Statistical Software*, 9 (2), 1–5.

Panchenko V. (2005). Goodness-of-fit test for copulas. *Physica A*, 355 (1), 176–182.

Quessy J. F., Mesfioui M., Toupin M. H. (2007). A goodness-of-fit test based on Spearman's dependence function. *Working paper*, Université du Québec à Trois-Rivières.

Rosenblatt M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, 23, 470–472.

Savu C., Trede M. (2008). Goodness-of-fit tests for parametric families of Archimedean copulas. *Quantitative Finance*, 8 (2), 109–116.

Scaillet O. (2005). Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters. *FAME Research Paper Series n. 145*, International Center for Financial Asset Management and Engineering.

Shih J. H. (1998). A goodness-of-fit test for association in a bivariate survival model. *Biometrika*, 85, 189–200.

Takeuchi K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Sciences*, 153, 12–18.

Wang W., Wells M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association*, 95, 62–72.

White H. (2000). A reality check for data snooping. *Econometrica*, 68, 1097–1126.