Об асимптотике преобразования типа Розенблатта гауссовской смеси

Евгений Савинов
easavinov@fa.ru
ORCID: 0000-0001-9414-8820

Финансовый университет при Правительстве РФ, Москва,
125993, Ленинградский проспект 49, ГСП-3.

Аннотация: Мы показываем, что преобразование перекрестной независимости (CI-преобразование) гауссовской смеси имеет асимптотически гауссовское распределение, связанное с гауссовым ядром смеси тем же преобразованием. Дополнительно мы изучаем поведение экстремальных значений в соответствующих схемах серий.

Abstract: We show that a cross independence (CI) transformation of some Gaussian mixture has an asymptotically Gaussian distribution connected with the Gaussian core of the mixture by the same type of transform. In addition we study a behavior of extreme values in related triangular arrays.

1. Введение

В Savinov and Shamraeva (2023) было введено CI-преобразование копул, основанное на преобразованиях случайных величин, которые ранее рассматривались в работах Шатских С.Я. (см., например, Shatskikh (2006)) и которые являются так называемыми преобразованиями розенблаттовского типа Rosenblatt (1952). В данной работе показано, как это преобразование может быть использовано в задаче определения принадлежности выборки смеси распределений с гауссовской структурой.

Опишем постановку задачи. Рассмотрим следующую общую задачу определения компонент смеси. Пусть имеется система случайных величин $X_1, X_2, \dots, X_n$, описывающая продолжительности работы n различных компонентов сложной системы, работающей в случайной среде. Мы предполагаем, что при заданном состоянии среды (t) компоненты зависимы, имеют различные характеристики, и совместная функция распределения времени их работы определяется как $G_n^{(t)}(x_1, x_2, \dots, x_n)$. Таким образом, продолжительности работы компонентов системы в случайной среде описываются смесью

$$F_n(x_1, x_2, \dots, x_n) = \int_{-\infty}^{\infty} G_n^{(t)}(x_1, x_2, \dots, x_n)\ \mu(dt), \tag{1}$$

где предполагается, что параметр состояния среды одномерный, а мера μ описывает ее вероятностное поведение. Интересует вопрос, можно ли по выборке из распределения F_n что-то узнать о распределении $G_n^{(t)}$, ничего не зная о μ. В Теореме 1 раздела 3 данной работы установлено, что CI-преобразование, как минимум, позволяет определить гауссовскую структуру (тип зависимости) функций $G_n^{(t)}$ в смеси (1), если последняя является одним из вариантов многомерного распределения Стьюдента (Kshirsagars Multivariate t-Distribution, см. Kotz and Nadarajah (2004), стр. 87) с числом степеней свободы r. Более того, проведенные эксперименты по моделированию, описанные в разделе 4, согласуются с предположением, что определить гауссовскую структуру компонент можно для более широкого класса гуссовских смесей типа (1). В то же время существуют выборки из многомерных распределений, CI-преобразования которых не приводят к гауссовской структуре с помощью описанной процедуры, что отрицает их извлечение из любой гауссовской смеси.

Среди исследований, посвященных оценкам компонент непрерывных (масштабных) смесей, в основном можно выделить работы об оценках весового распределения (см. например, Eltoft et al. (2006), Orellana et al. (2018), Melkumova and Shatskikh (2019)), а также о параметрах компонент смеси в случае условно независимых случайных величин (Dey (1990)).

В разделе 5 мы изучаем поведение максимума гауссовских случайных величин в схеме серий, когда совместные распределения каждой строки не являются гауссовскими, а порождаются смесью гауссовских распределений, копулы которых подверглись CI-преобразованию. Результаты раздела не связаны напрямую с процедурой определения гауссовости смеси. Здесь ставится вопрос, будут ли, ввиду выводов из Теоремы 1, полученные асимптотически гауссовские случайные векторы в строках схемы серий вести себя как гауссовские векторы с точки зрения асимптотики максимума их компонент. В рассмотренном здесь случае сильной зависимости показано, что поведение максимумов в таких схемах серий аналогично поведению в гауссовских схемах. Отметим, что аналогичная задача для случая слабой зависимости рассматривалась в Savinov (2015).

2. Основные понятия.

Прежде чем напомнить определение копулы введем некоторые обозначения.
$\mathbf{I} := [0,1]$, $\mathbf{u} := (u_1, u_2, \ldots, u_n) \in \mathbf{I}^n$. Для $\mathbf{a}, \mathbf{b} \in \mathbf{I}^n$ будем писать $\mathbf{a} \leqslant \mathbf{b}$, когда $a_k \leqslant b_k$ для всех $k = 1, 2, \ldots, n$. Для $\mathbf{a} \leqslant \mathbf{b}$ будем обозначать через $[\mathbf{a}, \mathbf{b}]$ n-прямоугольник $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n] \subseteq \mathbf{I}^n$. $\Phi(\cdot)$ -

Определение 1. *Пусть $H : \boldsymbol{I}^n \to \boldsymbol{I}$, $\boldsymbol{a}, \boldsymbol{b} \in \boldsymbol{I}^n$, $B = [\boldsymbol{a}, \boldsymbol{b}]$ – n-прямоугольник в $\boldsymbol{I}^n$.*
*H-*объемом *n-прямоугольника B будем называть разность порядка n функции H на B*

$$V_H(B) = \Delta_{\boldsymbol{a}}^{\boldsymbol{b}} H(\boldsymbol{t}) = \Delta_{a_n}^{b_n} \Delta_{a_{n-1}}^{b_{n-1}} \ldots \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} H(\boldsymbol{t}),$$

где разность первого порядка определяется как

$$\Delta_{a_k}^{b_k} H(\boldsymbol{t}) = H(t_1, \ldots, t_{k-1}, b_k, t_{k+1}, \ldots, t_n) - H(t_1, \ldots, t_{k-1}, a_k, t_{k+1}, \ldots, t_n).$$

Определение 2 (Copula)**.** *Функция $C : \boldsymbol{I}^n \to \boldsymbol{I}$ называется* n-копулой*, если она обладает следующими свойствами*

1) Для каждого $\boldsymbol{u} \in \boldsymbol{I}^n$

$$C(\boldsymbol{u}) = 0, \quad \text{если} \quad u_1 u_2 \cdots u_n = 0;$$

2) Если все координаты $\boldsymbol{u}$ кроме u_k равны 1, то

$$C(\boldsymbol{u}) = u_k;$$

3) Для всех $\boldsymbol{a}$ и $\boldsymbol{b}$ из $\boldsymbol{I}^n$ таких, что $\boldsymbol{a} \leqslant \boldsymbol{b}$

$$V_C([\boldsymbol{a}, \boldsymbol{b}]) \geqslant 0.$$

Как известно, копула связывает многомерную функцию распределения с одномерными маргинальными распределениями (см., например, Nelsen (2006)).
Далее вводится понятие CI-преобразования.
Рассмотрим случайный вектор X_n с абсолютно-непрерывной функцией распределения

$$F(x_1, \ldots, x_n) = C\left(F_1(x_1), \ldots, F_n(x_n)\right)$$

на вероятностном пространстве $\{\Omega, \mathfrak{B}, \mathsf{P}\}$. Здесь $F_i(x_i)$ маргинальные распределения и C копула. Обозначим через $f_{1\ldots n}(x_1, \ldots, x_n)$ плотность распределения вектора X_n, $f_i(x_i)$ маргинальные плотности, и $F_{i\,|\,1\ldots\hat{\imath}\ldots n}(x_i | x_1, \ldots, \widehat{x_i}, \ldots, x_n)$ условную функцию распределения случайной величины X_i относительно всех остальных компонент (знак ˆ означает пропуск соответствующего компонента). Рассмотрим случайные величины

$$X_{i,n}^* = F_{i\,|\,1\ldots\hat{\imath}\ldots n}(X_i | X_1, \ldots, \widehat{X_i}, \ldots, X_n). \tag{2}$$

Как показано в Savinov and Shamraeva (2023), их совместная функция распределения является копулой, и следующее определение корректно.

Определение 3. *Будем назвать отображение $C \mapsto C^{ci}$*

$$C^{ci}(u) = \mathsf{P}\{X_{1,n}^* \leqslant u_1, \ldots, X_{n,n}^* \leqslant u_n\}, \quad u = (u_1, \ldots, u_n),$$

*где $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ абсолютно-непрерывный вектор с копулой C, CI-*преобразованием *(*Cross-Independence*) абсолютно-непрерывной копулы C. Копулу $C^{ci}(u)$ будем называть CI-копулой или CI-образом копулы C.*

Далее мы вводим меру на гильбертовом пространстве только с одной целью – с помощью проекций меры на некоторый ортонормированный базис получить согласованное семейство копул.

Рассмотрим измеримое пространство $\{\mathbb{H}, \mathcal{B}(\mathbb{H})\}$, где $\mathbb{H}$ вещественное сепарабельное гильбертово пространство со счетным оронормированным базисом $\{e_i\}_{i=1}^{\infty}$, борелевской σ-алгеброй и скалярным произведением $\langle \cdot, \cdot \rangle$. Будем рассматривать на нем счетно-аддитивную меру μ с характеристическим функционалом

$$\Psi_\mu(y) = \int_0^\infty \exp\left\{ -\frac{t}{2} \langle By, y \rangle \right\} g_r(t)\, dt, \qquad y \in \mathbb{H}, \tag{3}$$

где B линейный самосопряженный положительно определенный ядерный оператор с собственными векторами $\{e_i\}_{i=1}^{\infty}$ и

$$g_r(t) = \frac{r^{r/2}}{2^{r/2}\Gamma(r/2)} t^{-r/2-1} \exp\left\{ -\frac{r}{2t} \right\}, \qquad t > 0.$$

Следует отметить, что эту меру можно назвать мерой Стьюдента на $\{\mathbb{H}, \mathcal{B}(\mathbb{H})\}$ с r степенями свободы. Действительно, выберем $\{f_k\}$ – произвольный ортонормированный базис в гильбертовом пространстве $\mathbb{H}$ и рассмотрим случайные величины $X_i = \langle \cdot, f_i \rangle$ на вероятностном пространстве $\{\mathbb{H}, \mathcal{B}(\mathbb{H}), \mu\}$. Ясно, что соответствующие проекции меры μ (распределения случайных векторов $\mathbf{X}_n = (X_1, \ldots, X_n)$) имеют характеристические функции

$$\psi_{1\ldots n}(y_1, \ldots, y_n) = \int_0^\infty \exp\left\{ -\frac{t}{2} \sum_{i,j=1}^n y_i y_j \langle B f_i, f_j \rangle \right\} g_r(t)\, dt. \tag{4}$$

Покажем, что характеристическая функция (4) соответствует одному из вариантов многомерного распределения Стьюдента Kshirsagars Multivariate t-distribution (см. Kotz and Nadarajah (2004), стр. 87).

Лемма 1. *Предположим, что вектор Y имеет n-мерное гауссовское распределение с нулевым средним и ковариационной матрицей $C = (\langle Bf_i, f_j \rangle)$, и S_r^2 независимая от Y случайная величина с распределением $\chi^2(r)$ $(r \in \mathbb{N})$.). Тогда случайный вектор $T = Y/\sqrt{S^2/r}$ обладает характеристической функцией (4).*

Доказательство. Характеристическая функция случайного вектора T имеет вид

$$\begin{aligned}\varphi_T(y) = \mathrm{E} \exp\left\{ i \frac{\langle Y, y \rangle}{\sqrt{S_r^2/r}} \right\} &= \mathrm{E}\left[\mathrm{E}\left(\exp\left\{ i \frac{\langle Y, y \rangle}{\sqrt{S_r^2/r}} \right\} \middle| \frac{1}{S_r^2/r} \right) \right] \\ &= \mathrm{E}\left[\mathrm{E}\left(\exp\left\{ i \langle Y, y \rangle \sqrt{U} \right\} \middle| U \right) \right],\end{aligned}$$

где $U \sim inv\Gamma\left(\frac{r}{2}, \frac{r}{2}\right)$ (обратное гамма-распределение с параметрами $(r/2, r/2)$). Тогда

$$\begin{aligned}\varphi_T(y) = \int_0^\infty \mathrm{E}\left(\exp\left\{ i \langle Y, y\sqrt{t} \rangle \right\} \right) g_r(t)\, dt &= \int_0^\infty \varphi_Y\left(y\sqrt{t} \right) g_r(t)\, dt \\ &= \int_0^\infty \exp\left\{ -\frac{t}{2} \langle Cy, y \rangle \right\} g_r(t)\, dt.\end{aligned}$$

□

Замечание 1. *Конечно, такое представление распределения Стьюдента как гауссовской смеси хорошо известно, и его структура зависимости (как и других нормальных смесей в случае размерности два) изучались с использованием копул, например, в Heinen and Valdesogo (2020).*

Введем некоторые вспомогательные понятия.

Обозначим $B_n = \pi_n B \pi_n$, где π_n — ортопроектор $\mathbb{H} \to \mathbb{H}_n = span\{f_1, \ldots, f_n\}$. Введем обозначения для следующих квадратичных форм:

$$\begin{aligned} s_n^2 &:= s_n^2(h) = \frac{1}{n} \langle B_n^{-1} \pi_n h, \pi_n h \rangle, \\ s_\infty^2 &:= s_\infty^2(h) = \lim_{n\to\infty} s_n^2(h), \end{aligned}$$

Случайные величины

$$\zeta_{i,n} := \zeta_{i,n}(h) = \frac{\langle B_n^{-1} f_i, h \rangle}{s_\infty \langle B_n^{-1} f_i, f_i \rangle^{1/2}}, \qquad i = 1, \ldots, n, \tag{5}$$

относительно меры μ (см. доказательство теоремы 1 в Savinov (2015)) совместно гауссовские с ковариациями

$$c_{ij}^{(n)} := cov\big(\zeta_{i,n}, \zeta_{j,n}\big) = \frac{\langle B_n^{-1} f_i, f_j \rangle}{[\langle B_n^{-1} f_i, f_i \rangle \langle B_n^{-1} f_j, f_j \rangle]^{1/2}}. \tag{6}$$

3. Сходимость CI-копул

Теорема 1. *Пусть C_n стьюдентовские копулы распределений с характеристическими функциями (4). Обозначим через $C_{gB_n^-}$ гауссовскую n-копулу с матрицей корреляций $\left(c_{ij}^{(n)}\right)$, элементы которой определяются уравнением (6). Тогда для любых $k \in \mathbb{N}$ и $(u_1, \ldots, u_k) \in (0,1)^k$ выполняется сходимость*

$$C_n^{ci}(u_1, \ldots, u_k, 1 \ldots, 1) \to C_{gB_k^-}(u_1, \ldots, u_k), \quad n \to \infty.$$

Доказательство. Без потери общности, мы можем предполагать C_n копулами семейства стьюдентовских векторов $\mathbf{X}_n = \{X_1, X_2, \ldots, X_n\}$, $n = 1, 2, \ldots$, определенных на измеримом пространстве $\{\mathbb{H}, \mathcal{B}(\mathbb{H})\}$, где $\mathbb{H}$ — вещественное сепарабельное гильбертово пространство со счетным ортонормированным базисом $\{f_i\}_{i=1}^{\infty}$, борелевской σ-алгеброй и скалярным произведением $\langle\cdot,\cdot\rangle$, счетно-аддитивной мерой Стьюдета μ с r степенями свободы и характеристическим функционалом (3), $X_i = \langle\cdot, f_i\rangle$.

Обозначим $x_i := \Phi^{-1}(u_i)$. Тогда

$$\begin{aligned} C_n^{\mathrm{ci}}(u_1, \ldots, u_k, 1 \ldots, 1) &= C_n^{\mathrm{ci}}(\Phi(x_1), \ldots, \Phi(x_k), 1 \ldots, 1) \\ &= \mu\left\{\Phi^{-1}\left(X_{1,n}^*\right) \leqslant x_1, \ldots, \Phi^{-1}\left(X_{k,n}^*\right) \leqslant x_k\right\}. \end{aligned} \tag{7}$$

Из пункта 3 леммы 3 работы Savinov (2015), следует, что

$$\lim_{n\to\infty} \mu\left\{\Phi^{-1}\left(X_{1,n}^*\right) \leqslant x_1, \ldots, \Phi^{-1}\left(X_{k,n}^*\right) \leqslant x_k\right\} = \lim_{n\to\infty} \mu\left\{\zeta_{1,n} \leqslant x_1, \ldots, \zeta_{k,n} \leqslant x_k\right\}. \tag{8}$$

Заметим, что для любого $i \leqslant k < n$ выполняется

$$B_n^{-1} f_i = B_k^{-1} f_i. \tag{9}$$

Действительно, положим

$$h := B_n^{-1} f_i \in \mathbb{H}_n, \quad h' := B_k^{-1} f_i \in \mathbb{H}_n,$$

следовательно,

$$\begin{gathered} f_i = B_k h' = \pi_k B h' = \pi_k \sum_{\tilde{i}=1}^{k} \langle h', f_{\tilde{i}}\rangle B f_{\tilde{i}} = \sum_{\tilde{k}=1}^{k}\sum_{\tilde{i}=1}^{k} \langle h', f_{\tilde{i}}\rangle \left\langle B f_{\tilde{i}}, f_{\tilde{k}}\right\rangle f_{\tilde{k}} \quad \Rightarrow \\ \sum_{\tilde{i}=1}^{k} \langle h', f_{\tilde{i}}\rangle \left\langle B f_{\tilde{i}}, f_{\tilde{k}}\right\rangle = \delta_{i\tilde{k}}, \\ B_n h' = \pi_n B h' = \sum_{\tilde{k}=1}^{n}\sum_{\tilde{i}=1}^{k} \langle h', f_{\tilde{i}}\rangle \left\langle B f_{\tilde{i}}, f_{\tilde{k}}\right\rangle f_{\tilde{k}} = \sum_{\tilde{k}=1}^{n} \delta_{i\tilde{k}} f_{\tilde{k}} = f_i = B_n h \quad \Rightarrow \\ B_n(h - h') = 0 \quad \Rightarrow \quad h = h', \end{gathered}$$

что и доказывает равенство (9).

Из (9) следует, что $c_{ij}^{(n)} = c_{ij}^{(k)}$, и учитывая (5) и (6), получим

$$\begin{aligned} \mu\left\{\zeta_{1,n} \leqslant x_1, \ldots, \zeta_{k,n} \leqslant x_k\right\} &= \mu\left\{\zeta_{1,k} \leqslant x_1, \ldots, \zeta_{k,k} \leqslant x_k\right\} = \\ &= C_{gB_k^-}(\Phi(x_1), \ldots, \Phi(x_k)) = C_{gB_k^-}(u_1, \ldots, u_k), \end{aligned} \tag{10}$$

и доказательство теоремы следует из равенств (7), (8) и (10). □

Замечание 2. *Пример CI-преобразования n-мерной гауссовской копулы в Savinov and Shamraeva (2023) показывает, что*

$$C_{gB_k^-}(u_1, \ldots, u_k) = C_{gB_k}^{ci}(u_1, \ldots, u_k),$$

где C_{gB_k} копула гауссовского распределения с ковариационной матрицей $\langle Bf_i, f_j\rangle$, $i, j = 1, 2, \ldots, k$. Таким образом, из этого примера и теоремы 1 следует, что

$$C_n^{ci}(u_1, \ldots, u_k, 1 \ldots, 1) \to C_{gB_k}^{ci}(u_1, \ldots, u_k).$$

4. Моделирование и проверка согласия выборки большой размерности с нормальной смесью

Для проверки согласия некоторой n-мерной выборки (для достаточно большого n) с семейством распределений (1) предлагается процедура, состоящая из последовательного CI-преобразования, нормализации (в смысле приведения координат элементов полученной выборки к стандартному гауссовскому распределению), редукции (взятия проекции выборки на меньшую размерность) и проверки полученной выборки на согласие с многомерным нормальным распределением.

Для участие в эксперименте были смоделированы выборки размерности $n = 40$ объема $N = 200$ из следующих шести распределений: многомерного распределения Kshirsagars Multivariate t-distribution вида (4) с недиагональными матрицами $\mathrm{C} = (\langle Bf_i, f_j\rangle)$ с $r = 1, 2, 4$ и 8-ю степенями свободы, гауссовской смеси вида (1), где мера μ определяется показательным распределением с плотностью $e^{-t},\ t>0$, и многомерного распределения, в котором одномерные гауссовские компоненты связаны D-vine копулами (см. например, Joe (1996), Bedford and Cooke (2001), Bedford and Cooke (2002)), в которых в качестве парных условных копул были выбраны двумерные гауссовские и копулы Клэйтона (с разными параметрами).

При вычислении CI-преобразования для оценок условных функций распределения применялся подход, использующий снижение размерности (см. Hall and Yao (2005)).

Для проверки многомерной нормальности использовался критерий, описанный в Henze and Zirkler (1990), с уровнем значимости $\alpha = 0.05$. Все двенадцать выборок, собранные в Таблицах 1 и 2, предварительно были приведены к одномерным маргинальным нормальным распределениям и имеют соответствующую указанному распределению структуру зависимости (например, $\tilde{t}(8)$ – выборка из распределения с копулой стьюдентовского распределения $t(8)$ и маргинальными $N(0,1)$, а $ci\ \tilde{t}(8)$ распределение с CI-образом стьюдентовской копулы и теми же маргинальными нормальными распределениями). Таким образом, тест многомерной нормальности фактически проверяет наличие гауссовской структуры зависимости.

В ячейках Таблиц указаны p-значения и результат проверки на нормальность (True/False) проекции соответствующей размерности, указанной в наименовании столбца. Как можно видеть, проекции исходных выборок любой размерности продемонстрировали отклонение от гауссовской структуры. В то же время проекции низкой (2-5) размерности CI-образов показали согласие с гауссовским распределением для всех гауссовских смесей, включая вариант с экспоненциальным смешиванием (см. последнюю строку в Таблице 2), и преимущественное отклонение от гауссовской структуры для проекций распределения на основе CI-образа D-vine копулы.

Таблица 1.

n=200 d=20	**2**	**3**	**5**	**8**
$\tilde{t}(8)$	$7.7e-07(False)$	$7.6e-16(False)$	$3.9e-33(False)$	$9.6e-125(False)$
$\tilde{t}(4)$	$1.8e-05(False)$	$3.4e-09(False)$	$1.3e-31(False)$	$2.7e-185(False)$
$\tilde{t}(2)$	$1.1e-33(False)$	$1.8e-93(False)$	$0.0e+00(False)$	$0.0e+00(False)$
$\tilde{t}(1)$	$9.1e-56(False)$	$1.4e-145(False)$	$0.0e+00(False)$	$0.0e+00(False)$
$\widetilde{exp\_mix}$	$6.3e-30(False)$	$3.0e-88(False)$	$0.0e+00(False)$	$0.0e+00(False)$
$vine$	$1.3e-04(False)$	$6.8e-05(False)$	$1.5e-29(False)$	$2.4e-178(False)$
$ci\ \tilde{t}(8)$	$3.1e-01(True)$	$3.6e-02(False)$	$2.9e-01(True)$	$2.4e-01(True)$
$ci\ \tilde{t}(4)$	$4.2e-01(True)$	$8.6e-01(True)$	$7.6e-01(True)$	$6.8e-01(True)$
$ci\ \tilde{t}(2)$	$3.6e-01(True)$	$2.4e-01(True)$	$4.7e-01(True)$	$7.0e-02(True)$
$ci\ \tilde{t}(1)$	$4.1e-01(True)$	$1.5e-01(True)$	$1.3e-01(True)$	$6.3e-03(False)$
$ci\ \widetilde{exp\_mix}$	$1.3e-01(True)$	$1.1e-04(False)$	$1.1e-14(False)$	$2.5e-70(False)$
$ci\ vine$	$2.4e-02(False)$	$2.9e-01(True)$	$1.4e-02(False)$	$5.6e-08(False)$

Таблица 2.

n=200 d=40	**2**	**3**	**5**	**8**
$\tilde{t}(8)$	$7.3e-04(False)$	$1.8e-07(False)$	$2.7e-19(False)$	$2.1e-38(False)$
$\tilde{t}(4)$	$7.4e-11(False)$	$1.9e-23(False)$	$1.8e-83(False)$	$0.0e+00(False)$
$\tilde{t}(2)$	$4.0e-28(False)$	$1.0e-70(False)$	$0.0e+00(False)$	$0.0e+00(False)$
$\tilde{t}(1)$	$1.4e-56(False)$	$5.3e-162(False)$	$0.0e+00(False)$	$0.0e+00(False)$
$\widetilde{exp\_mix}$	$5.7e-31(False)$	$5.7e-81(False)$	$0.0e+00(False)$	$0.0e+00(False)$
$ci\ \tilde{t}(8)$	$9.0e-02(True)$	$4.0e-01(True)$	$3.0e-01(True)$	$3.1e-01(True)$
$ci\ \tilde{t}(4)$	$7.7e-01(True)$	$7.2e-01(True)$	$7.4e-01(True)$	$3.0e-01(True)$
$ci\ \tilde{t}(2)$	$6.6e-01(True)$	$4.5e-01(True)$	$8.9e-02(True)$	$9.8e-02(True)$
$ci\ \tilde{t}(1)$	$1.3e-01(True)$	$7.1e-02(True)$	$2.0e-05(False)$	$1.6e-02(False)$
$ci\ \widetilde{exp\_mix}$	$8.7e-01(True)$	$6.3e-01(True)$	$4.0e-01(True)$	$2.1e-02(False)$

Замечание 3. *Относительно небольшой объем выборки ($N = 200$) объясняется тем, что при фиксированной базовой размерности n увеличение выборки ведет к увеличению мощности критерия, который начинает демонстрировать отклонение от нормальности проекций всех CI-образов, что естественно, поскольку они лишь асимптотически гауссовские.*

5. Экстремальная предельная теорема

Теорема 1 и Замечание 2 показывают, что CI-образ распределения Стьюдента оказывается асимптотически гауссовским распределением, которое, в свою, очередь, является результатом CI-преобразования гауссовского ядра исходной смеси.

Сейчас нас будет интересовать следующий вопрос: верно ли, что экстремальные значения случайных величин с CI-копулами распределения Стьюдента и гауссовскими маргинальными распределениями ведут себя так же, как экстремумы гауссовских векторов в схеме серий?

Стоит заметить, что экстремальные значения в схеме серий зависимых случайных величин, связанные различными семействами копул, изучались в Lebedev (2015). В работе также представлены обобщенные результаты об экстремальных индексах для схем серий с архимедовыми копулами. Кроме того, в Goldaeva and Lebedev (2018) изучались вопросы серий случайной длины.

Используем классические для экстремальной теории (см. Leadbetter et al. (1983)) числовые последовательности ($n \geqslant 2$)

$$\alpha_n = (2\ln n)^{1/2}, \quad \beta_n = (2\ln n)^{1/2} - \frac{1}{2}(2\ln n)^{-1/2}(\ln\ln n + \ln(4\pi)).$$

Теорема 2. *Рассмотрим семейство стьюдентовских случайных векторов $\boldsymbol{X}_n = \{X_1, X_2, \ldots, X_n\}$ $n = 1, 2, \ldots,$ — определенных на вероятностном пространстве $\{\mathbb{H}, \mathcal{B}(\mathbb{H}), \mu\}$, описанное выше. Пусть C_n соответствующие t-копулы, $\Phi(x)$ и $\varphi(x)$ обозначают стандартную нормальную функцию распределения и плотность, соответственно. Введем схему серий случайных величин $\left\{X_i^{(n)}\right\}_{i=1}^{n}$ $n = 1, 2, \ldots,$ определенную на подходящем вероятностном пространстве $\{\Omega_0, \mathfrak{B}_0, \mathsf{P}_0\}$ и имеющую совместную функцию распределения*

$$\mathsf{P}_0\left\{X_1^{(n)} \leqslant x_1, \ldots, X_n^{(n)} \leqslant x_n\right\} = C_n^{ci}(\Phi(x_1), \ldots, \Phi(x_n)).$$

Если базис $\{f_k\}$ такой, что

$$\delta := \limsup_{n\to\infty} \max_{i\neq j} \left|c_{ij}^{(n)}\right| < 1, \tag{11}$$

и для $\gamma > 0$ существует $0 < \alpha < \frac{1-\delta}{1+\delta}$ такой, что

$$\max_{n^\alpha < j-i < n} \left|c_{ij}^{(n)} \ln(j-i) - \gamma\right| \to 0, \quad n \to \infty, \tag{12}$$

то для любого $x \in \mathbb{R}$ имеет место сходимость

$$\mathsf{P}_0\left\{\alpha_n\left(\max_{1\leqslant i\leqslant n} X_i^{(n)} - \beta_n\right) \leqslant x\right\} \to \int_{-\infty}^{\infty} \exp\left\{-e^{-x-\gamma+\sqrt{2\gamma}\,z}\right\}\varphi(z)\,dz. \tag{13}$$

Доказательство.

В доказательстве используется Теорема 1 из работы Savinov (2024) и леммы 2, 3, 4 из работы Savinov (2015).

Без ограничения общности можно считать, что вероятностное пространство $\{\Omega_0, \mathfrak{B}_0, \mathsf{P}_0\}$ есть $\{\mathbb{H}, \mathcal{B}(\mathbb{H}), \mu\}$, и

$$X_i^{(n)} = \Phi^{-1}\left(X_{i,n}^*\right). \tag{14}$$

Обозначим

$$M_n := \max_{1\leqslant i\leqslant n} X_i^{(n)}. \tag{15}$$

Далее будем использовать обозначения

$$a_i(n) := \frac{X_i^{(n)}}{\zeta_{i,n}}, \quad b_i(n) := \zeta_{i,n},$$

$$a_*(n) := \min_{1\leqslant i\leqslant n} \frac{X_i^{(n)}}{\zeta_{i,n}}, \quad a^*(n) := \max_{1\leqslant i\leqslant n} \frac{X_i^{(n)}}{\zeta_{i,n}}, \quad b^*(n) := \max_{1\leqslant i\leqslant n} \zeta_{i,n},$$

$$d(n) := a_*(n)\mathbf{1}_{\{b^*(n)\geqslant 0\}} + a^*(n)\mathbf{1}_{\{b^*(n)<0\}},\ e(n) := a^*(n)\mathbf{1}_{\{b^*(n)\geqslant 0\}} + a_*(n)\mathbf{1}_{\{b^*(n)<0\}},$$

где $\zeta_{i,n}$ определены в (7) из работы Savinov (2015).

Поскольку μ-п.н. $a_*(n) > 0$ (см. лемма 3, п. 2 в Savinov (2015)) и μ-п.н.

$$d(n)b^*(n) \leqslant M_n \leqslant e(n)b^*(n),$$

(см. лемма 4 в Savinov (2015)), то μ-п.н.

$$\alpha_n\left[d(n)b^*(n) - \beta_n\right] \leqslant \alpha_n\big(M_n - \beta_n\big) \leqslant \alpha_n\left[e(n)b^*(n) - \beta_n\right]. \tag{16}$$

Заметим (см. лемма 3, п. 3 в Savinov (2015)), что μ-п.н. $a_*(n) \to 1$, $a^*(n) \to 1$, поэтому ввиду

$$a_*(n) \leqslant d(n) \leqslant a^*(n), \quad a_*(n) \leqslant e(n) \leqslant a^*(n),$$

имеем

$$d(n) \to 1, \quad e(n) \to 1, \quad \mu - \text{a.s.} \tag{17}$$

Рассмотрим левую часть неравенства (16)

$$\alpha_n \left[d(n)b^*(n) - \beta_n\right] = d(n)\alpha_n \left[b^*(n) - \beta_n\right] + \alpha_n\beta_n \left[d(n) - 1\right]. \tag{18}$$

Известно (см. доказательство теоремы 1 в Savinov (2015)), что случайные величины $\zeta_{i,n},\ i = 1, 2, \dots, n$ совместно гауссовские с ковариациями

$$cov(\zeta_{i,n}, \zeta_{j,n}) = c_{ij}^{(n)}.$$

Поэтому в силу (17) и Теоремы 1 из Savinov (2024) для всех $x \in \mathbb{R}$

$$\mu\left\{d(n)\alpha_n \left[b^*(n) - \beta_n\right] \leqslant x\right\} \to \int\limits_{-\infty}^{\infty} \exp\left\{-e^{-x-\gamma+\sqrt{2\gamma}\,z}\right\} \varphi(z)\, dz. \tag{19}$$

Воспользуемся ранее полученными результатами. Замечая, что для $n \geqslant 2$

$$\alpha_n\beta_n = 2\ln n - \frac{1}{2}\left(\ln\ln n + \ln 4\pi\right) > 0,$$

μ-п.н. (см. лемма 3, п. 1 в Savinov (2015))

$$\begin{aligned}\alpha_n\beta_n A^{(n)} \leqslant \alpha_n\beta_n \left[a_*(n) - 1\right] \leqslant& \\ \leqslant \alpha_n\beta_n \left[d(n) - 1\right] \leqslant& \\ \leqslant \alpha_n\beta_n \left[a^*(n) - 1\right] \leqslant& \ \alpha_n\beta_n \max_{1\leqslant i\leqslant n} B_i^{(n)}.\end{aligned}$$

Таким образом, (см. лемма 2, п. 2, 3 в Savinov (2015)) μ-п.н. $\alpha_n\beta_n \left[d(n) - 1\right] \to 0$, откуда с учетом (18) и (19)

$$\mu\left\{\alpha_n \left[d(n)b^*(n) - \beta_n\right] \leqslant x\right\} \to \int\limits_{-\infty}^{\infty} \exp\left\{-e^{-x-\gamma+\sqrt{2\gamma}\,z}\right\} \varphi(z)\, dz.$$

Сходимость в правой части (16) доказывается аналогично, откуда и следует утверждение теоремы. $\square$

6. Заключение

Смеси распределений широко изучаются в различных контекстах, например, в таких как теория надежности, а также в других подобных случаях, когда мы имеем дело с группой наблюдений, состоящей из разнородных подгрупп. Как следует из настоящей работы, применяя методы оценки условных функций распределения и используя CI-преобразование, можно построить некоторые критерии согласия выборки с гауссовской смесью (точнее, со смесью распределений с гауссовской структурой), пока, правда, в относительно частном случае семейства стьюдентовских распределений.

Для иллюстрации проводились испытания на модельных данных, которые показали, что в действительности есть надежда распространить этот подход на гауссовские смеси с другими весовыми распределениями.

Отметим, что результат работы Savinov and Shamraeva (2023) позволяет говорить об использовании данного подхода для проверки отсутствия бесконечной перестановочности (т.е. принадлежности, в силу теоремы де Финетти, к смесям независимых случайных величин).

Наконец, было показано, что максимальные компоненты элементов преобразованной многомерной выборки из такой гауссовской смеси ведут себя аналогично тем же максимальным компонентам элементов преобразованной многомерной гауссовской выборки.

Поскольку результаты Теоремы 1 дают представление о предельной структуре зависимости и ее связи с исходной структурой, последующие исследования могут быть посвящены оценкам параметров распределений $G_k^{(t)}$ и их асимптотическим свойствам, в том числе, и для более широкого класса непрерывных смесей.

Список литературы

Bedford, T., Cooke, R.M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence 32, 245–268. URL: https://doi.org/10.1023/A:1016725902970, doi:10.1023/A:1016725902970.

Bedford, T., Cooke, R.M., 2002. Vines: A new graphical model for dependent random variables. The Annals of Statistics 30, 1031–1068. URL: http://www.jstor.org/stable/1558694.

Dey, D.K., 1990. Estimation of scale parameters in mixture distributions. The Canadian Journal of Statistics / La Revue Canadienne de Statistique 18, 171–178. URL: http://www.jstor.org/stable/3315566.

Eltoft, T., Kim, T., Lee, T.W., 2006. Multivariate scale mixture of gaussians modeling, in: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (Eds.), Independent Component Analysis and Blind Signal Separation, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 799–806.

Goldaeva, A.A., Lebedev, A.V., 2018. On extremal indices greater than one for a scheme of series. Lithuanian Mathematical Journal 58, 384–398. doi:10.1007/s10986-018-9407-2.

Hall, P., Yao, Q., 2005. Approximating conditional distribution functions using dimension reduction. The Annals of Statistics 33, 1404–1421. URL: http://www.jstor.org/stable/3448693.

Heinen, A., Valdesogo, A., 2020. Spearman rank correlation of the bivariate student t and scale mixtures of normal distributions. J. of Multivariate Analysis 179, 104650. doi:10.1016/j.jmva.2020.104650.

Henze, N., Zirkler, B., 1990. A class of invariant consistent tests for multivariate normality. Communications in Statistics - Theory and Methods 19, 3595–3617. URL: https://doi.org/10.1080/03610929008830400, doi:10.1080/03610929008830400.

Joe, H., 1996. Families of m-variate distributions with given margins and m (m-1)/2 bivariate dependence parameters. Lecture notes-monograph series , 120–141.

Kotz, S., Nadarajah, S., 2004. Multivariate t Distributions and Their Applications. Cambridge University Press. doi:10.1017/CBO9780511550683.

Leadbetter, M., Lindgren, G., Rootzen, H., 1983. Extremes and Related Properties of Random Sequences and Processes. Springer Series in Statistics, Springer-Verlag, New York. doi:10.1007/978-1-4612-5449-2.

Lebedev, A.V., 2015. Экстремальные индексы в схеме серий и их приложения. Информ. и её примен. 9, 39–54. doi:10.14357/19922264150305.

Melkumova, L., Shatskikh, S., 2019. Maximum likelihood method in de finetti's theorem. Theory of probability and its applications , 808–816doi:https://doi.org/10.1137/S0040585X97T989313.

Nelsen, R., 2006. An Introduction to Copulas. Second ed., Springer-Verlag, New York. doi:10.1007/0-387-28678-0.

Orellana, R., Carvajal, R., Agüero, J.C., 2018. Maximum likelihood infinite mixture distribution estimation utilizing finite gaussian mixtures. IFAC-PapersOnLine 51, 706–711. URL: https://www.sciencedirect.com/science/article/pii/S240589631831872X, doi:https://doi.org/10.1016/j.ifacol.2018.09.200. 18th IFAC Symposium on System Identification SYSID 2018.

Rosenblatt, M., 1952. Remarks on multivariate transformation. Ann. Math. Stat. 23, 470–472. doi:10.1214/aoms/1177729394.

Savinov, E., 2024. On gaussian triangular arrays in the case of strong dependence. Extremes URL: https://doi.org/10.1007/s10687-024-00491-3, doi:10.1007/s10687-024-00491-3.

Savinov, E., Shamraeva, V., 2023. On a rosenblatt-type transformation of multivariate copulas. Econometrics and Statistics 25, 39–48. URL: https://www.sciencedirect.com/science/article/pii/S2452306221001313, doi:https://doi.org/10.1016/j.ecosta.2021.10.016.

Savinov, E.A., 2015. Limit theorem for maximum of random variables with copulas which are it-copulas of student's t-distribution. Theory Probab. Appl. 59, 508–516. doi:10.1137/S0040585X97T987260.

Shatskikh, S.Y., 2006. The strong law of large numbers for a triangular array scheme of conditional distributions of stable elliptically contoured measures. Theory Probab. Appl. 50, 248–264. doi:10.1137/S0040585X97981652.