

Основные элементы теории копул

Статья содержит важнейшие факты из теории p -копул — распределений на p -мерном гиперкубе с равномерными распределениями на $[0; 1]$ их компонент. В настоящее время копулы используются в задачах финансового рынка, макроэкономических моделях и многих других областях статистических приложений. Анализ русскоязычных статей показывает, что в своем большинстве они представляют списки копул (язык) и правила действий (грамматика) при почти полном отсутствии информации о смысле объектов и правил. В этом обзоре сделана попытка хотя бы в некоторой, малой степени исправить это положение.

Ключевые слова: копула; теорема Склара; условие Липшица; произведение копул; таблица сопряженности.

JEL classification: C19; C49; C69.

1. Введение. Что такое копула?

Пусть R^p — p -мерное пространство Эвклида, элементами которого являются вектор-столбцы $x \in R^p$ с компонентами x_1, \dots, x_p , T — знак транспонирования, так что $x^T = (x_1, \dots, x_p)$ — вектор-строка.

Функцию распределения случайного вектора $\xi = \xi(\omega)$, $\omega \in \Omega$, со значениями в R^p обозначим¹ $H(x) = P\{\xi_1 \leq x_1, \dots, \xi_p \leq x_p\}$. Пусть $F_k(x_k) = P\{\xi_k \leq x_k\}$, $k = 1, \dots, p$ — *маргинальные* функции распределения отдельных компонент (или *маргиналы*).

Скажем, что функция распределения $H \in \mathcal{H}$, если все ее маргиналы $F_k(x_k)$, $k = 1, \dots, p$, непрерывны. Для таких $F_k(x_k)$ определим λ -квантиль по формуле $F_k^{-1}(\lambda) = \inf\{u : F_k(u) \geq \lambda\}$, $\lambda \in (0; 1)$.

Пусть закон распределения $H(x)$ случайного вектора ξ принадлежит классу \mathcal{H} и пусть

$$\alpha_1 = F_1(\xi_1), \dots, \alpha_p = F_p(\xi_p). \quad (1)$$

Легко доказать, что все $\alpha_1, \dots, \alpha_p$ равномерно распределены на отрезке $U^1 = [0; 1]$, так что их совместная функция распределения $C_p(t)$ является *копулой*. Общее определение этого термина таково: p -мерное распределение $C_p(t)$ на p -мерном единичном кубе U^p , у которого все маргинальные распределения являются *равномерными* на отрезке U^1 , называется *копулой* или, что точнее, *p -копулой*.

Теорема Склара (Sklar's theorem) утверждает, что какими бы ни были p -мерное распределение $H(x) = H(x_1, \dots, x_p)$ и его маргинальные распределения $F_k(x_k)$, $k = 1, \dots, p$, существует такая копула $C_p(t) = C_p(t_1, \dots, t_p)$, что для всех $x \in R^p$ выполняется равенство:

¹ Для случайного вектора $\xi = \xi(\omega)$, $\omega \in \Omega$ функция распределения $H(x)$ определяется как вероятность множества $\Omega_{\xi, x} = \{\omega : \xi_j(\omega) \leq x_j, 1 \leq j \leq p\}$, в отличие от принятого в русскоязычной литературе определения со строгим неравенством.

$$H(x_1, \dots, x_p) = C_p(F_1(x_1), \dots, F_p(x_p)). \quad (2)$$

Для $H \in \mathcal{H}$ эта теорема, как уже отмечалось, почти очевидна, теоретическая «возня» начинается в случае, когда на распределение $H(x)$ не накладывается никаких ограничений, и его маргиналы могут быть любыми одномерными распределениями, даже такими, когда проекции на ось Y скачков в точках разрыва образуют множество Кантора. Трудности «прячутся» в точках разрыва маргинальных функций распределения, т. к. величины $\alpha_1, \dots, \alpha_p$ из (1) с равномерным распределением на U^1 теперь не определены: если $F_1(u)$ разрывна в точке u_0 , то имеется интервал $(a; b) \subset U^1$, в который значения $F_1(\xi_1)$ не попадают, т. е. $\alpha_1 = F_1(\xi_1)$ не является *равномерно распределенной* на U^1 случайной величиной.

Ниже в разделе 5 будет приведено доказательство этой теоремы, которое практически точно повторяет доказательство из статьи (Rüschendorf, 2009). Вызывает удивление, что во многих исследованиях копулы авторы заново доказывают классическую теорему Колмогорова об условиях, которым должна удовлетворять p -мерная функция $H(x)$, чтобы быть функцией распределения, см., например, (Nelsen, 2010). Такое ощущение, что они не читали европейские учебники по теории вероятностей середины прошлого века, где она была приведена. А эта теорема необходима, чтобы проверить, является ли некая функция $C_p(t)$ копулой, т. е. *функцией распределения* с равномерными в U^1 маргиналами.

В настоящее время копулы используются в разных задачах, которые «подкидывает» финансовый рынок, страховое дело, эконометрика и многие другие области статистических приложений. С другой стороны, копулы привлекают к себе внимание специалистов по теории вероятностных метрических пространств (probabilistic metric space). На копулы сейчас мода, о чем свидетельствуют регулярные конференции² и недвусмысленно говорит следующая цитата из (Durante, Sempì, 2010):

«В последнее время усилились критические высказывания о копулах и их применениях, а отдельные люди заговорили о «копуломании»... Мы полагаем, что это естественная реакция на широкое распространение работ, в которых использования копул не всегда хоть чем-нибудь мотивированы. Следует сказать, что некоторые люди ошибочно думают, что копулы «решают все проблемы стохастической зависимости». Это заведомо не так! Копулы — это хороший инструмент для понимания некоторых проблем стохастической зависимости, но никак не «панацея» от всех проблем со стохастической зависимостью»³.

Конечно, сама идея создавать разнообразные модели многомерных распределений в пространстве p -мерного куба с равномерным распределением для отдельных компонент за-

² Конференции по теории копул *Probability distributions with given marginals*: Рим (1990), Сиэтл (1993), Прага (1996), Барселона (2000), Квебек (2004), Тарту (2007), Сан-Паулу (2010).

³ В оригинале: «However, a word of caution is in order here. Several criticisms have been recently raised about copulas and their applications, and several people started to speak about «copula craze» ... From our point of view, these criticisms were a quite natural reaction to such a wide diffusion of applications of copulas, not always in a well motivated way. It should be said that several people have wrongly interpreted copulas as the solution to «all problems of stochastic dependence». This is definitely not the case! Copulas are indispensable tool for understanding several problems about stochastic dependence, but they are not the «panacea» for all stochastic models».

служивает самого высокого одобрения. Когда-то идею равнометризации компонент мы с Л. Д. Мешалкиным эксплуатировали для визуализации многомерных данных⁴.

Примерно те же идеи использовались и при работе над Т-нормальными распределениями⁵. Для этого семейства копула, называемая *гауссовой*, имеет следующий вид:

$$C_{TN}(t_1, \dots, t_p | \Sigma) = \Phi_p \left(\Phi^{-1}(t_1), \dots, \Phi^{-1}(t_p) | 0; \Sigma \right), \quad (3)$$

где $\Phi_p(x | \mu; \Sigma)$ — p -мерное нормальное распределение, $\Phi^{-1}(t)$ — обратная к стандартной гауссовской функции распределения, Σ — матрица ковариаций, $\mu \in R^p$. Теперь, много лет спустя, Т-нормальное распределение может быть определено через гауссову копулу как

$$F(x_1, \dots, x_p) = C_{TN}(F_1(x_1), \dots, F_p(x_p) | \Sigma), \quad (4)$$

где $F_1(x), \dots, F_p(x)$ — любой набор непрерывных одномерных функций распределений.

Для использования этих распределений самыми важными являются два факта. Во-первых, выборочные функции $\hat{F}_1(x), \dots, \hat{F}_p(x)$ распределения отдельных компонент и матрица $\hat{\Sigma}$ коэффициентов корреляций Ван-дер-Вардена являются *достаточными и независимыми статистиками*. А во-вторых, оцененная копула $C_{TN}(t | \hat{\Sigma})$ по выборке размера n является оценкой для искомой копулы $C_{TN}(t | \Sigma)$ с дисперсией порядка n^{-1} независимо от размерности p .

2. Чем же копулы замечательны?

Во-первых, представление многомерного распределения в виде

$$F(x_1, \dots, x_p) = C(F_1(x_1), \dots, F_p(x_p)), \quad (5)$$

где $C(t_1, \dots, t_p)$ — любая копула, «делит» особенности исходного распределения на *маргинальные* и *коррелятивные*, на совокупность распределений $F_1(x_1), \dots, F_p(x_p)$ и копулу $C(t_1, \dots, t_p)$, в которой упакованы все связи между компонентами исходного распределения, поскольку копула *инвариантна ко всем монотонным преобразованиям* осей координат в R^p , т. е. в ней сосредоточены *все модели зависимостей* между случайными величинами.

Во-вторых, любая копула $C(t_1, \dots, t_p)$ удовлетворяет условию Липшица

$$|C(s_1, \dots, s_p) - C(t_1, \dots, t_p)| \leq |s_1 - t_1| + \dots + |s_p - t_p| \quad (6)$$

с *абсолютной* константой Липшица — коэффициентом при $d(s, t) = |s_1 - t_1| + \dots + |s_p - t_p|$. Доказательство неравенства (6) будет приведено в следующем разделе. Этим свойством гладкости семейство копул принципиально отличается от семейства всех распределений.

⁴ См. «Алгоритмы классификации и визуализации, основанные на квантильном анализе», Программное обеспечение ЭВМ. Библиотека прикладных программ БИМ-М, выпуск 20, Минск, 1989, 60–76.

⁵ Это такое распределение $H(x_1, \dots, x_p)$, которое трансформируется к p -мерному нормальному посредством монотонных преобразований $u_k = g_k(x_k)$, $k = 1, \dots, p$, компонент вектора $x^T = (x_1, \dots, x_p)$.

Далее, представление (5) дает сравнительно простые подходы к моделированию многомерных распределений. Особенно ярко это проявляется для случая $p = 2$. Действительно, если копула $C(u, v) = P\{\alpha \leq u, \beta \leq v\}$ имеет плотность $c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v} \in (0, +\infty)$ при всех $(u, v) \in U^2$, то плотность условного распределения β при условии, что $\alpha = u$, равна $f(v|u) = c(u, v)$, поскольку плотность распределения α равна 1 (плотность равномерного распределения на U^1). А это означает, что μ -квантиль $v(\mu|u)$ условного распределения β находится из уравнения $c_\alpha(u, v) - c_\alpha(u, 0) = \mu$, где $c_\alpha(u, v) = \frac{\partial}{\partial u} C(u, v)$, т. к. $\int_0^v f(z|u) dz = \frac{\partial}{\partial u} C(u, v) \Big|_0^v$. Отсюда следует, что две независимые и равномерно распределенные на U^1 случайные величины ξ_1 и ξ_2 , определяемые по формулам $\alpha = \xi_1$ и $\beta = v(\xi_2 | \xi_1)$, моделируют пару, совместное распределение которой в точности совпадает с копулой $C(u, v) = P\{\alpha \leq u, \beta \leq v\}$.

Следует отметить, что имеется весьма богатая коллекция семейств, состоящих из копул с разнообразными формами связей между компонентами. Об этом будет достаточно подробно рассказано в разделе 4.

Наконец, нельзя не упомянуть о том, что совокупность всех копул C^p размерности p является неким математическим объектом, который активно исследуется, особенно множество C^2 двумерных копул. Оказывается, что в этом «пространстве» можно ввести умножение (правда, не аддитивное), 0-элемент и 1-элемент. Интерес к таким проблемам привлекает специалистов по теории полугрупп, теории марковских процессов, теории надежности, финансовой математике. Подробное изложение свойств двумерных копул с весьма аккуратными доказательствами содержится в статье (Eryilmaz, 2011). Кое-какие из этих свойств обобщаются в разделе 3.

3. Особые свойства копул

3.1. Гладкость копул. Начнем с доказательства неравенства (6) для произвольной копулы $C(t_1, t_2) \in C^2$, поскольку случай копул любой размерности практически ничем не отличается от случая $p = 2$, но становится менее наглядным из-за более громоздких формул. Пусть $0 \leq u < v \leq 1$, $f(x) = C(x, v) - C(x, u)$ и $g(y) = C(v, y) - C(u, y)$. Очевидно, что $f(x)$ и $g(y)$ — неубывающие функции от $x \in [0; 1]$ и $y \in [0; 1]$ соответственно. Действительно, т. к. $C(t_1, t_2)$ является распределением неких случайных величин ξ и η , то

$$f(x) = P\{\xi \leq x; \eta \leq v\} - P\{\xi \leq x; \eta \leq u\} = P\{\xi \leq x; u < \eta \leq v\}$$

не убывает с ростом x . Совершенно такие же рассуждения пригодны для функции $g(y)$.

Отсюда следуют неравенства

$$0 = f(0) \leq f(x) = C(x, v) - C(x, u) \leq f(1) = v - u, \quad (7)$$

$$0 = g(0) \leq g(y) = C(v, y) - C(u, y) \leq g(1) = v - u. \quad (8)$$

Но

$$C(s_1, s_2) - C(t_1, t_2) = [C(s_1, s_2) - C(s_1, t_2)] + [C(s_1, t_2) - C(t_1, t_2)],$$

где $|C(s_1, s_2) - C(s_1, t_2)| \leq |s_2 - t_2|$ и $|C(s_1, t_2) - C(t_1, t_2)| \leq |s_1 - t_1|$, так что

$$|C(s_1, s_2) - C(t_1, t_2)| \leq |s_1 - t_1| + |s_2 - t_2|,$$

что и требовалось доказать.

Из (6) и общих теорем о монотонных функциях можно утверждать, что всякая копула $C(t_1, \dots, t_p)$ почти всюду дифференцируема по каждой из переменных, т. е. существуют интегрируемые функции $C'_k(t_1, \dots, t_p) = \frac{\partial}{\partial t_k} C(t_1, \dots, t_p)$, $k = 1, \dots, p$, такие, что

$$0 \leq C'_k(t_1, \dots, t_p) \leq 1, \quad k = 1, \dots, p.$$

Эти неравенства выполняются почти всюду на отрезке варьирования $t_k \in [0; 1]$ при условии, что остальные переменные фиксированы.

3.2. Копулы и марковские процессы. Возьмем две копулы $A(s_1, \dots, s_m)$ и $B(t_1, \dots, t_n)$ произвольных размерностей m и n . Пусть

$$C(u_1, \dots, u_{m+n-1}) = \int_{0 \leq x \leq u_m} A'_m(u_1, \dots, u_{m-1}, x) B'_1(x, u_{m+1}, \dots, u_{m+n-1}) dx,$$

где $A'_m(u_1, \dots, u_{m-1}, x) = \frac{\partial}{\partial x} A(u_1, \dots, u_{m-1}, x)$ и $B'_1(x, u_{m+1}, \dots, u_{m+n-1}) = \frac{\partial}{\partial x} B(x, u_{m+1}, \dots, u_{m+n-1})$,

что можно кратко записать как результат произведения копул: $C_{m+n-1}(\cdot) = A_m(\cdot) \circ B_n(\cdot)$. Эта операция над копулами используется, во-первых, для создания новых семейств копул, а во-вторых, связывает копулы с марковскими процессами. А именно, пусть $A(u | t_1, \dots, t_n)$, $u^T = (u_1, \dots, u_p)$ — копула для совместного распределения $X(t_1), \dots, X(t_n)$. Оказывается, что $X(t)$ будет марковским процессом тогда и только тогда, когда для любого $n \geq 2$ и любых $t_1 < \dots < t_n$

$$A(u | t_1, \dots, t_n) = A(\cdot | t_1, t_2) \circ A(\cdot | t_2, t_3) \circ \dots \circ A(\cdot | t_{n-1}, t_n).$$

Этот результат и многие другие, его сопровождающие, можно найти в (Darsow et al., 1999).

3.3. Алгебра двумерных копул. Семейство двумерных копул C^2 обладает своим собственным произведением, не выводящим за пределы семейства. Возьмем две двумерные копулы $A(u, v)$, $B(u, v)$ и положим $C = A * B$, если функция $C(x, y)$ определена равенством:

$$C(x, y) = A * B(x, y) = \int_0^1 A'_2(x, t) B'_1(t, y) dt,$$

где нижний индекс у производной — номер переменной, по которой проводилось дифференцирование копул. Известно (строго доказано), что $C = A * B \in C^2$, т. е. C также является копулой.

Две пограничные копулы, $\Pi_2(t) = t_1 \cdot t_2$ и $M_2(t) = \min(t_1, t_2)$, определяют для этой операции «нуль» $\mathbf{0} = \Pi_2(t)$ и «единицу» $\mathbf{1} = M_2(t)$ соответственно, поскольку прямыми вычис-

лениями устанавливается, что $C * \mathbf{0} = \mathbf{0} * C = \mathbf{0}$ и $C * \mathbf{1} = \mathbf{1} * C = C$ для любой копулы C . Эта операция с копулами подчиняется ассоциативному закону, хотя в общем случае не коммутативна. Для $p > 2$ множество \mathbf{C}^p (как математический объект) устроено гораздо сложнее, и пока не видно связей этих особых свойств со статистическими проблемами, поэтому они не будут обсуждаться в данной статье.

3.4. Границы Фреше-Хефдинга. Вообще говоря, это границы для вероятности одновременного осуществления событий A_1, \dots, A_p , т. е. для вероятности $Q_p = P\{A_1 \cdot \dots \cdot A_p\}$ в зависимости от вероятностей этих событий $f_1 = P\{A_1\}, \dots, f_p = P\{A_p\}$ и независимо от любых взаимосвязей между этими событиями. Точные (достижимые) границы даются неравенствами:

$$\max(0, f_1 + \dots + f_p - p + 1) \leq Q_p \leq \min(f_1, \dots, f_p). \quad (9)$$

Пусть закон распределения вектора $\xi = (\xi_1, \dots, \xi_p)$ задается копулой $C(t_1, \dots, t_p)$, и пусть даны события⁶ $A_k = \{\xi_k \leq t_k\}$, $k = 1, \dots, p$. Поскольку маргинальные распределения копулы являются равномерными распределениями на отрезке $U^1 = [0; 1]$, то $P\{A_k\} = t_k$, $1 \leq k \leq p$, и $L_p(t) \equiv \max(0, t_1 + \dots + t_p - p + 1) \leq C(t_1, \dots, t_p) \leq \min(t_1, \dots, t_p) \equiv M_p(t)$, т. е. при любых t_1, \dots, t_p значение любой копулы $C_p(t) \in \mathbf{C}^p$ лежит в диапазоне от $L_p(t)$ до $M_p(t)$. Из (9) можно выписать границы для функции распределения $H(x) = H(x_1, \dots, x_p)$ произвольного вектора $\xi = (\xi_1, \dots, \xi_p)$ через его маргинальные распределения $F_1(x_1), \dots, F_p(x_p)$:

$$\max(0, F_1(x_1) + \dots + F_p(x_p) - p + 1) \leq H(x) \leq \min(F_1(x_1), \dots, F_p(x_p)).$$

Доказательство неравенств (9). Правая часть (9) тривиальна, т. к. вероятность произведения событий меньше вероятности осуществления любого одного из них. Из очевидного неравенства

$$P\{A_1\} \leq P\{A_1 A_2\} + P\{\bar{A}_2\} = P\{A_1 A_2\} + 1 - P\{A_2\}$$

сразу же следует левая часть (9) для $p = 2$, а для $p > 2$ это неравенство доказывается по индукции. Поскольку всегда $H(x) \geq 0$, то достаточно доказать, что неравенство $Q_k \geq f_1 + \dots + f_k - k + 1$ выполняется при $k = p$, если оно выполняется при всех $k = 2, \dots, p - 1$. Во-первых, получим, что

$$P\{A_1 \dots A_{p-1} A_p\} + P\{A_1 \dots A_{p-1} \bar{A}_p\} = P\{A_1 \dots A_{p-1}\} \geq f_1 + \dots + f_{p-1} - (p - 1) + 1. \quad (10)$$

Далее, правую часть (10) легко преобразовать к виду:

$$f_1 + \dots + f_{p-1} - (p - 1) + 1 = (f_1 + \dots + f_{p-1} + f_p - p + 1) + (1 - f_p) = (f_1 + \dots + f_p - p + 1) + P\{\bar{A}_p\}.$$

⁶ Негласно предполагается, что вектор $\xi = \xi(\omega)$ определен на множестве Ω элементарных событий $\omega \in \Omega$.

Из этого равенства и (10) следует неравенство:

$$P\{A_1 \dots A_{p-1} A_p\} \geq (f_1 + \dots + f_p - p + 1) + (P\{\bar{A}_p\} - P\{A_1 \dots A_{p-1} \bar{A}_p\}). \quad (11)$$

А поскольку $P\{\bar{A}_p\} - P\{A_1 \dots A_{p-1} \bar{A}_p\} \geq 0$, то из (11) сразу же получим

$$P\{A_1 \dots A_{p-1} A_p\} \geq f_1 + \dots + f_p - p + 1,$$

что и требовалось доказать.

4. Конструктивные типы копул

4.1. Эллиптические копулы. В обследованной автором литературе наиболее общее определение эллиптической копулы дано в (Фантаццини, 2011). В чуть упрощенном изложении оно состоит в следующем.

Назовем p -мерный случайный вектор ξ эллиптическим, если $\xi = z + \vartheta \cdot A\xi$, где $z \in R^p$, $AA^T = \Sigma$, $\xi^T = (\xi_1, \dots, \xi_p)$ — случайный вектор, имеющий равномерное распределение на p -мерной сфере⁷ $S_p^1 = \{x : |x| = \sqrt{x_1^2 + \dots + x_p^2} = 1\}$, ϑ — случайный радиус, т.е. числовая случайная величина, распределенная на $R^+ = (0, +\infty)$ с плотностью $f_g(r) = K_{g,p} r^{p-1} g(r^2)$, где $K_{g,p}$ — константа, а $g(u)$ — неотрицательная интегрируемая функция (генератор). Введем обозначение $E(z, \Sigma, g)$ для семейства эллиптических векторов, задаваемых параметрами z , Σ и g .

Пусть теперь $\xi \in E(z, \Sigma, g)$. Оказывается, все нормированные компоненты $\xi_k^* = (\xi_k - z_k) / \sigma_k$, где σ_k^2 — дисперсия ξ_k , $k = 1, \dots, p$, одинаково распределены, причем их общая функция распределения $F_0(x)$ непрерывна. Распределение $C_p(t | \Sigma^*, g)$ случайного вектора $(F_0(\xi_1^*), \dots, F_0(\xi_p^*))$, где Σ^* — ковариационная матрица вектора ξ^* с компонентами ξ_k^* , $k = 1, \dots, p$, как раз и является эллиптической копулой.

Из таким образом определенного класса эллиптических копул на практике используются в основном лишь два типа: гауссовы копулы с генератором $g(u) = e^{-t/2}$ и T -копулы Стьюдента, которые порождаются генератором $g(u) = (1 + u/d)^{-(p+d)/2}$, где d интерпретируется как число степеней свободы.

Для $p = 2$ эти два типа копул имеют следующий вид:

$$C_\Phi(s, t) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\Psi(s)} \int_{-\infty}^{\Psi(t)} \exp\left(-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right) dx dy \quad \text{— гауссова копула,}$$

$$C_T(s, t) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{G(s/d)} \int_{-\infty}^{G(t/d)} \left(1 + \frac{x^2 - 2rxy + y^2}{2d(1-r^2)}\right)^{-(d+2)/2} dx dy \quad \text{— } T\text{-копула Стьюдента,}$$

⁷ Сконструировать такое распределение совсем просто: если $\gamma_1, \dots, \gamma_p$ — совокупность независимых случайных величин, имеющих стандартное нормальное распределение, то вектор с компонентами $\xi_k = \gamma_k / |\gamma|$, $1 \leq k \leq p$, будет равномерно распределен на p -мерной единичной сфере.

где функции $\Psi(t)$ и $G(t|d)$ — обратные к функциям распределения стандартного нормального закона и одномерного T -распределения Стьюдента с d степенями свободы соответственно.

4.2. **Архимедовы копулы.** Это семейство копул задается формулой

$$C_p(t|\psi) = \psi(\psi^{-1}(t_1) + \dots + \psi^{-1}(t_p)), \quad (12)$$

где генератором является функция $\psi(t)$, удовлетворяющая двум условиям:

а) $\psi(t)$ определена в R^+ , имеет $(p-2)$ производных и выполняется неравенство

$$(-1)^k \psi^{(k)}(t) \geq 0 \text{ для } k = 0, \dots, p-2 \text{ и всех } t \in R^+; \quad (13)$$

б) функция $(-1)^{p-2} \psi^{(p-2)}(t)$ является выпуклой и убывающей в R^+ .

В ряде статей имеются ссылки на работу (Mikusinski et al., 1991), в которой доказывается, что (12) является копулой тогда и только тогда, когда выполняются требования а) и б). Такая ссылка имеется, например, в весьма аккуратной с математической точки зрения статье (McNeil, Nešlehová, 2009). Конкретные виды копул приводить не будем, их можно найти в (Durante, Sempì, 2010) или в (Фантаццини, 2011). Остановимся лишь на двух *пограничных* копулах, которые играют в теории особую роль. Это копулы *независимости* $\Pi_p(t) = t_1 \cdot \dots \cdot t_p$ и «*вырождения*» $M_p(t) = \min(t_1, \dots, t_p)$. Первая соответствует распределению вектора, компоненты которого независимы (в совокупности), а вторая — когда все компоненты равны друг другу. При этом в обоих случаях речь идет о равномерно распределенных на отрезке $[0; 1]$ величинах.

Приведем формулы генераторов трех семейств, которые чаще всего упоминаются:

$$\psi_{\text{ГН}}(t) = \exp(-t^{1/\theta}), \theta \geq 1 \text{ (Gumbel-Hougaard);}$$

$$\psi_{\text{МТС}}(t) = (\max(1 + \theta t, 0))^{-1/\theta}, \theta \geq -1/(p-1), \theta \neq 0 \text{ (Mardia-Takanasi-Clayton);}$$

$$\psi_{\text{Ф}}(t) = -\frac{1}{\theta} \cdot \ln(1 - (1 - e^{-\theta})e^{-t}), \theta \geq 1 \text{ (Frank).}$$

Маленькая математико-историческая справка. Еще в 1826 году великий норвежский математик Нильс Хенрик Абель (1802–1829) доказал: если функция $f(x, f(y, z))$ инвариантна к перестановкам переменных, а $f(x, y)$ — дифференцируема, то существует дифференцируемая и обратимая функция $\psi(t)$ такая, что $\psi(f(x, y)) = \psi(x) + \psi(y)$. Этот результат является, во-первых, предтечей архимедовых копул, а во-вторых, стал основой второй части пятой проблемы Гильберта, сформулированной в 1900 году, которая в английском варианте звучит следующим образом:

«*In how far are the assertions which we can make in the case of differentiable functions true under proper modifications without this assumption?*»⁸

Сам вопрос относится к *ассоциативным* уравнениям $f(x, f(y, z)) = f(f(x, y), z)$, без требований к равенству всех перестановок между собой. Пока эту проблему никто не ре-

⁸ Смысловой перевод: «Сколько сильно можно ослабить требования к дифференцируемым функциям, чтобы при их подходящей модификации результат остался бы справедливым?» (в сравнении с требованиями и результатом Абеля).

шил — самым выдающимся считается результат, который получил в 1949 году венгерский математик János Aczél, живший и работавший в Канаде. Он состоит в следующем.

Пусть J — открытый или полуоткрытый интервал. Функция $f: J \times J \rightarrow J$ ассоциативна, непрерывна и сократима (*cancellative*) тогда и только тогда, когда $f(x, y) = \psi^{-1}(\psi(x) + \psi(y))$, где функция $\psi: J \rightarrow \mathbb{R}^1$ непрерывна и строго монотонна.

В одну сторону доказательство совсем простое. Поскольку

$$f(x, f(y, z)) = \psi^{-1}(\psi(x) + \psi(f(y, z))) \text{ и } \psi(f(y, z)) = \psi(y) + \psi(z),$$

то $f(x, f(y, z)) = \psi^{-1}(\psi(x) + \psi(y) + \psi(z))$. Точно из тех же соображений получим, что $f(f(x, y), z) = \psi^{-1}(\psi(x) + \psi(y) + \psi(z))$, т. е. $f(x, f(y, z)) = f(f(x, y), z)$. Далее, из непрерывности и строгой монотонности $\psi(t)$ непрерывность $f(x, y)$ следует очевидным образом и, наверное, также легко доказывается свойство *cancellative* (сократимости), точный математический смысл которого я не знаю и не смог найти в доступной мне литературе.

Название «архимедовы копулы» этого семейства происходит из-за их одного свойства, очень похожего на аксиому Архимеда (Archimedean axiom):

- Для любых положительных чисел a и b существует такое целое $n < +\infty$, что $b < an$.

Пусть теперь $C(x, y) = \psi^{-1}(\psi(x) + \psi(y))$ — копула, т. е. $\psi(t)$ является, согласно (13), выпуклой и убывающей функцией на всем множестве $\{t \geq 0\}$. Оказывается, что аналог аксиомы Архимеда создает рекуррентное соотношение $\varphi_{n+1}(a) = C(a, \varphi_n(a))$ с $\varphi_1(a) \equiv a$:

- Для любых a и b из отрезка $[0; 1]$ существует такое целое $n < +\infty$, что $\varphi_n(a) < b$.

Приведем доказательство, тем более что оно очень простое. Несложно увидеть, что $\varphi_{n+1}(a) = \psi^{-1}(n\psi(a))$: для этого нужно подставить $x = a$ и $y = \varphi_1(a)$ в $\psi^{-1}(\psi(x) + \psi(y))$, а далее — по индукции. Поскольку $\psi(a)$ и $\psi(b)$ положительны, то по аксиоме Архимеда существует $n < +\infty$ такое, что $n\psi(a) > \psi(b)$, так что $\varphi_{n+1}(a) = \psi^{-1}(n\psi(a)) < \psi^{-1}(\psi(b)) = b$, что и требовалось доказать. Последнее неравенство оправдано тем, что ψ^{-1} — убывающая функция (как обратная к убывающей ψ).

4.3. Другие конструкции копул. Сначала сформулируем один результат о взвешивании копул. Пусть $C_0(t)$ и $C_1(t)$ — две произвольные копулы, являющиеся распределениями двух d -мерных векторов ξ_0 и ξ_1 , случайная величина τ принимает два значения, 0 и 1, с вероятностями α и $1 - \alpha$ соответственно, и все они (ξ_0 , ξ_1 и τ) независимы. Тогда (Mikusinski et al., 1991) случайный вектор

$$\xi = \xi_0 I\{\tau = 0\} + \xi_1 I\{\tau = 1\}, \quad (14)$$

где $I\{B\}$ — индикатор события B , распределен по закону $C_{0 \oplus 1}(t) = \alpha C_0(t) + (1 - \alpha)C_1(t)$. Другими словами, взвесь двух копул снова является копулой (множество копул выпукло).

Используя (14) и две пограничные копулы $\Pi_p(t)$, $M_p(t)$, можно построить новое семейство копул (копулы Fréchet–Mardia):

$$C_{FM}(t) = \alpha \Pi_p(t) + (1 - \alpha)M_p(t). \quad (15)$$

Вот еще одна конструкция копул (Cuadras, Augé, 1981). Пусть ξ_1, \dots, ξ_p и τ — независимые случайные величины, причем $P\{\xi_k \leq t\} = t^\alpha$, $\alpha \in (0; 1)$, $k = 1, \dots, p$, и $P\{\tau \leq t\} = t^{1-\alpha}$. Тогда вектор ξ с компонентами $\xi_k = \max(\xi_k, \tau)$ распределен по закону:

$$C_{CA}(t) = \left(\prod_p(t) \right)^\alpha \left(M_p(t) \right)^{1-\alpha}. \quad (16)$$

По иной схеме создается EFGM-семейство копул⁹. Пусть $J = \{j_1, \dots, j_k\}$ — любая последовательность таких целых чисел, что $1 \leq j_1 < \dots < j_k \leq p$, $k = 1, \dots, p$, и пусть $\mathbf{J}(p)$ — множество последовательностей, содержащих не менее двух элементов¹⁰. Далее, каждой последовательности $J \in \mathbf{J}(p)$ поставим в соответствие действительное число $\alpha(J)$ и определим

$$C_{\text{EFGM}}(t_1, \dots, t_p) = \prod_{k=1}^p t_k \cdot \left(1 + \sum_{J \in \mathbf{J}(p)} \alpha(J) \prod_{k \in J} (1 - t_k) \right), \quad (17)$$

где на параметры наложены следующие условия: $|\alpha(J)| \leq 1$ и $1 + \sum_{J \in \mathbf{J}(p)} \alpha(J) \prod_{k \in J} \varepsilon_k \geq 0$ при любых $\varepsilon_k \in \{-1, +1\}$, $k = 1, \dots, p$. В качестве примера можно привести копулу (17) для $p = 2$, равную

$$C_{\text{EFGM}}(t_1, t_2) = t_1 t_2 (1 + \alpha \cdot (1 - t_1)(1 - t_2)),$$

где $|\alpha| \leq 1$, поскольку $\mathbf{J}(p)$ для $p = 2$ состоит всего из одной последовательности $J = \{1, 2\}$.

4.4. Пертурбации с копулами. По-английски это звучит «Distortions of copulas», *искажение* копул. Фактически за этим термином кроется один класс преобразований, которые копулу переводят в копулу. А именно, пусть $C_p(t) \in C^p$, а функция $\psi(z)$, преобразующая отрезок $[0; 1]$ в себя, удовлетворяет следующим требованиям: а) непрерывна, б) строго возрастает, в) вогнута, г) $\psi(0) = 0$, $\psi(1) = 1$. Тогда

$$C_p^{(\psi)}(t_1, \dots, t_p) = \psi^{-1} \left(C_p(\psi(t_1), \dots, \psi(t_p)) \right) \in C^p,$$

т. е. функция, полученная после преобразования копулы $C_p(t)$, снова является копулой.

Приведем несколько функций, которые наиболее часто используются для «пертурбаций» копул (табл. 1).

Таблица 1. Пять популярных функций, используемых для «пертурбаций» копул

Преобразование	Формула	Условие вогнутости
Пропорциональное изменение риска (Proportional hazard)	$z^{1/\lambda}$	$\lambda \geq 1$
Экспоненциальное	$(1 - e^{-\lambda z}) / (1 - e^{-\lambda})$	$\lambda > 0$
Логарифмическое	$(\ln[1 - z(1 - e^\lambda)]) / \lambda$	$\lambda > 0$
Преобразование Ванга	$\Phi(\Phi^{-1}(z) + \lambda)$	$\lambda \geq 0$
Двойное степенное	$1 - (1 - z)^\lambda$	$\lambda \geq 1$

⁹ Авторы статьи (Durante, Sempi, 2010) связывают это семейство с независимыми исследованиями Morgenstern и Gumbel, их развитием в работе Farlie в 1960-х гг., хотя первоначальная идея была еще в более ранней и долгое время забытой статье Euraud 1936 года.

¹⁰ $\mathbf{J}(p)$ содержит $2^p - p - 1$ последовательностей, удовлетворяющих этому условию.

5. Доказательство теоремы Склара

Теорема Склара. Если $H(x) = P\{\xi_1 \leq x_1, \dots, \xi_p \leq x_p\}$ — функция распределения произвольного случайного вектора ξ , принимающего значения в R^p , и $\{F_k(x_k), k=1, \dots, p\}$ — маргинальные функции распределения его компонент, то существует такая копула $C_p(t) = C_p(t_1, \dots, t_p)$, что

$$H(x) = H(x_1, \dots, x_p) = C_p(F_1(x_1), \dots, F_p(x_p)) \text{ для всех } x \in R^p. \quad (18)$$

Для $H \in \mathcal{H}$ доказательство тривиально, поскольку всем условиям теоремы удовлетворяет функция $C_p(t_1, \dots, t_p) = P\{\alpha_1 \leq t_1, \dots, \alpha_p \leq t_p\}$, где $\alpha_k = F_k(\xi_k)$, $k=1, \dots, p$. В общем случае эта идея не работает, т. к. при этом можно не получить равномерные на $[0; 1]$ случайные величины α_k , $1 \leq k \leq p$. Преодолеть это препятствие позволяет следующая лемма.

Лемма. Пусть ξ — произвольная случайная величина на числовой прямой R^1 с функцией распределения $F(x)$, τ — независимая от ξ случайная величина, равномерно распределенная на $[0; 1]$. Тогда, если ввести функции

$$F(x-) = \lim_{\varepsilon \downarrow 0} F(x - \varepsilon), \quad F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad F(x, \lambda) = F_{<}(x) + \lambda \cdot F_{=}(x), \quad (19)$$

где $F_{<}(x) = P\{\xi < x\}$ и $F_{=}(x) = P\{\xi = x\}$, то случайная величина $\alpha = F(\xi, \tau)$ равномерно распределена на отрезке $[0; 1]$ и $\xi = F^{-1}(\alpha)$ почти всюду.

Доказательство. Определим нижний b -квантиль $q_b^* = \sup\{x : F(x) < b\}$, $b \in (0; 1)$, и величины $r(b) = F_{=}(q_b^*)$, $q(b) = F_{<}(q_b^*)$. Докажем сначала, что осуществление события

$$\Omega_b = \{\xi < q_b^*\} \cup \{\xi = q_b^*, q(b) + r(b)\tau \leq b\} \quad (20)$$

эквивалентно неравенству $\alpha \leq b$. Поскольку $\alpha = F(\xi, \tau)$, то $\alpha \leq b$ означает, что $F_{<}(\xi) + \tau F_{=}(\xi) \leq b$. А это равносильно утверждению

$$(\xi, \tau) \in \{(x, \lambda) : F_{<}(x) + \lambda F_{=}(x) \leq b\}, \quad (21)$$

поэтому надо доказать, что из (20) следует (21), и наоборот.

Если верно (20), то либо $\xi = q_b^*$ и $q(b) + r(b)\tau \leq b$, либо $\xi < q_b^*$. В первом случае из определения $q(b)$ и $r(b)$ получим, что $F_{<}(\xi) + \tau F_{=}(\xi) \leq b$, т. е. (21) выполняется. А если $\xi < q_b^*$, то $F(\xi) < b$, что следует из определения q_b^* , так что из (20) всегда следует (21).

Если верно (21) и $\xi = q_b^*$, то осуществляется событие $B_b = \{\xi = q_b^*, q(b) + r(b)\tau \leq b\}$ и, тем самым, (20). Далее, (20) автоматически следует из осуществления события $\hat{B}_b = \{\xi < q_b^*\}$, так что надо лишь доказать, что (21) и $\xi > q_b^*$ несовместимы.

Итак, пусть $\xi > q_b^* = \sup\{x : F(x) < b\}$. Если ξ попадает в точку непрерывности функции распределения, тогда $F_{=}(\xi) = 0$ и $F(\xi) = F_{<}(\xi)$, так что $F(\xi) \leq b$, а это (в зоне непрерывности) означает, что $\xi \leq q_b^*$.

Если же ξ попадает в точку разрыва x_0 и $F_{<}(x_0) + \tau F_{=}(x_0) \leq b$ по (21), то либо $\xi = q_b^* = x_0$ для любого $b \in [F_{<}(x_0), F_{<}(x_0) + F_{=}(x_0)]$, либо $b > F(x_0) = F_{<}(x_0) + F_{=}(x_0) = F(\xi)$ и $\xi < q_b^*$. Этим полностью завершается доказательство эквивалентности (20) и (21).

Проведенный анализ показывает, что $P\{\alpha \leq b\} = P\{\Omega_b\} = P\{\hat{B}_b \cup B_b\}$, так что, учитывая независимость ξ и τ и то, что $\hat{B}_b \cap B_b = \emptyset$, при $r(b) = F_-(q_b^*) > 0$ получим:

$$P\{\alpha \leq b\} = P\{\hat{B}_b\} + P\{B_b\} = q(b) + r(b)P\{q(b) + \tau r(b) \leq b\} = q(b) + r(b)P\left\{\tau \leq \frac{b - q(b)}{r(b)}\right\} = b,$$

т. к. $P\{\tau \leq u\} = u$ и значит $q(b) + r(b)P\{\tau \leq (b - q(b))/r(b)\} = q(b) + r(b)[(b - q(b))/r(b)] = b$.

Осталось доказать, что $\xi = F^{-1}(\alpha)$ почти всюду. Но поскольку для всех $b \in [F_-(x), F(x)]$ имеет место равенство $F^{-1}(b) = x$, и всегда $F_-(\xi) \leq \alpha \leq F(\xi)$, то $F^{-1}(\alpha) = \xi$, что и требовалось доказать.

Эта лемма является, по сути, остовом доказательства теоремы. Пусть для каждого маргинального распределения $F_k(x)$, $1 \leq k \leq p$, функция $F_k(x, \lambda)$ определена (при любых $x \in R^1$ и $\lambda \in [0; 1]$) по тем же правилам, как и функция $F(x, \lambda)$ в формуле (19). Если τ_1, \dots, τ_p — набор независимых (в совокупности и от вектора $\xi^T = (\xi_1, \dots, \xi_p)$) равномерно распределенных на $[0; 1]$ случайных величин, то все $\alpha_k = F_k(\xi_k, \tau_k)$, $1 \leq k \leq p$, также распределены равномерно на $[0; 1]$ и $C_p(t_1, \dots, t_p) = P\{\alpha_1 \leq t_1, \dots, \alpha_p \leq t_p\}$ является искомой копулой. Это следует из того, что

$$H(x) = P\{F_1^{-1}(\alpha_1) \leq x_1, \dots, F_p^{-1}(\alpha_p) \leq x_p\} = P\{\alpha_1 \leq F_1(x_1), \dots, \alpha_p \leq F_p(x_p)\} = C_p(F_1(x_1), \dots, F_p(x_p)).$$

6. Таблицы сопряженности (2×2) и связанные с ними копулы

В самых разных исследованиях возникают *таблицы сопряженности*, простейшими из них являются таблицы (2×2).

Таблица 2. Совместное распределение двух дихотомических признаков в одной выборке

	$\xi = 0$	$\xi = 1$	Сумма
$\eta = 0$	n_{00}	n_{01}	n_{0*}
$\eta = 1$	n_{10}	n_{11}	n_{1*}
Сумма	n_{*0}	n_{*1}	n_{**}

В ячейках этой таблицы указано, во-первых, число случаев каждого варианта из четырех возможных, (0; 0), (0; 1), (1; 0) и (1; 1), например, n_{01} представляет вариант (0; 1) — число наблюдений, в которых одновременно $\xi = 1$ и $\eta = 0$. Во-вторых, в ней даны суммы, например, $n_{0*} = n_{00} + n_{01}$, и общее число наблюдений $n_{**} = n_{00} + n_{01} + n_{10} + n_{11} = n_{0*} + n_{1*} = n_{*0} + n_{*1}$. Общая модель таких таблиц может быть представлена похожей таблицей с *вероятностями* (табл. 3).

Это две биномиальные величины ξ и η , принимающие значения 0 и 1, причем ξ принимает эти значения с вероятностями p и $1 - p$ соответственно, а η — с вероятностями q и $1 - q$ соответственно. Связь между ξ и η характеризуется числом r , которое необходимо удовлетворяет неравенствам Фреше–Хефдинга:

$$L(p, q) \equiv \max(0, p + q - 1) \leq r \leq \min(p, q) \equiv M(p, q). \quad (22)$$

Таблица 3. Общая модель совместного распределения двух биномиальных случайных величин

	$\xi = 0$	$\xi = 1$	$P\{\eta = \cdot\}$
$\eta = 0$	$1 + r - p - q$	$p - r$	$1 - q$
$\eta = 1$	$q - r$	r	q
$P\{\xi = \cdot\}$	$1 - p$	p	

Теорема Склара утверждает, что существует копула $C(s, t | p, q, r)$ такая, что совместное распределение *дискретных* величин ξ и η равно $P\{\xi \leq x, \eta \leq y\} = C(F(x), G(y) | p, q, r)$, где $P\{\xi \leq x\} = F(x)$ и $P\{\eta \leq y\} = G(y)$. Поиск в Интернете не дал никакой информации о копулах, представляющих распределение векторов с дискретной областью значений. Поэтому выпишем здесь такую копулу в простейшем варианте двумерного биномиального вектора, наблюдения над которым приводят к таблице сопряженности (2×2) общего вида.

В ходе доказательства теоремы Склара было установлено, что к искомой копуле ведет совместное распределение случайных величин $\alpha_1 = F(\xi, \tau_1)$ и $\alpha_2 = G(\eta, \tau_2)$, определяемых формулами (19). Легко проверить, что

$$\alpha_1 = F(\xi, \tau_1) = (1 - p)\tau_1(1 - \xi) + [(1 - p) + p\tau_1]\xi,$$

$$\alpha_2 = G(\eta, \tau_2) = (1 - q)\tau_2(1 - \eta) + [(1 - q) + q\tau_2]\eta,$$

и обе величины, α_1 и α_2 , действительно распределены равномерно на единичном отрезке.

Обратим внимание, что случайные величины ξ и η являются *индикаторами*, т. е. слабые в выражениях для α_1 и α_2 «работают» отдельно друг от друга. Так, например, $\alpha_1 = (1 - p)\tau_1$, если $\xi = 0$ и $\alpha_1 = (1 - p) + p\tau_1$, если $\xi = 1$. Эта связь *обратима*: диапазон значений от 0 до $(1 - p)$ для α_1 реализуется лишь при $\xi = 0$, а при $\xi = 1$ значение α_1 варьируется от $(1 - p)$ до 1. Это разбивает единичный квадрат (область значений α_1 и α_2) на четыре подобласти, которые задаются простыми парами неравенств:

$$\mathbf{D}_{00} = \{(s, t): 0 \leq s \leq 1 - p, 0 \leq t \leq 1 - q\}, \quad \mathbf{D}_{01} = \{(s, t): 0 \leq s \leq 1 - p, 1 - q < t \leq 1\},$$

$$\mathbf{D}_{10} = \{(s, t): 1 - p < s \leq 1, 0 \leq t \leq 1 - q\}, \quad \mathbf{D}_{11} = \{(s, t): 1 - p < s \leq 1, 1 - q < t \leq 1\}.$$

Очевидно, что в каждой из этих подобластей копула будет описываться *разными* аналитическими формулами. Подробные выкладки не столь сложны и не требуют какой-либо изобретательности, поэтому приведем сразу результат.

Копула $C(s, t | p, q, r) = P\{\alpha_1 \leq s, \alpha_2 \leq t\}$ в определенных выше подобластях равна:

$$\mathbf{D}_{00}: C(s, t | p, q, r) = \frac{s}{1 - p} \cdot \frac{t}{1 - q} \cdot (1 + r - p - q);$$

$$\mathbf{D}_{01}: C(s, t | p, q, r) = (1 + r - p - q) \cdot \frac{s}{1 - p} + \frac{s}{1 - p} \cdot \frac{t - (1 - q)}{q} \cdot (q - r);$$

$$\mathbf{D}_{10}: C(s, t | p, q, r) = (1 + r - p - q) \cdot \frac{t}{1 - q} + \frac{s - (1 - p)}{p} \cdot \frac{t}{1 - q} \cdot (p - r);$$

$$\mathbf{D}_{11}: C(s, t | p, q, r) = (1 + r - p - q) + \frac{t-1+q}{q} \cdot (q-r) + \frac{s-1+p}{p} \cdot (p-r) + \frac{s-1+p}{p} \cdot \frac{t-1+q}{q} \cdot r.$$

Несложно видеть, что линии уровня (постоянных значений) являются гиперболами. Например, для подобласти \mathbf{D}_{01} : если $C(s, t | p, q, r) = C_0 \in (0; 1)$, то $s(t - t_0) = h_0$, где $t_0 = (pq - r) / (q - r)$ и $h_0 = C_0 q(1 - p) / (q - r)$.

Несмотря на то что все эти выкладки и утверждения весьма просты, они требуют внимания и каких-нибудь иллюстративных рисунков, чтобы «увидеть результат». Для исправления этого «педагогического ущерба» разберем случай, когда $P\{\xi = 1\} = P\{\eta = 1\} = 0.5$, а совместная вероятность этих событий $r = P\{\xi = 1, \eta = 1\} \neq pq = 0.5 \cdot 0.5 = 0.25$, т. е. между двумя «бросаниями симметричных монеток», символизирующих пару случайных величин ξ и η , имеется стохастическая связь (точнее, всегда, кроме случая, когда $r = pq$). Совместное распределение этих величин, каждая из которых характеризует выпадение определенной стороны при бросании симметричной монеты (1 — «герб», 0 — «решетка»), приведено в табл. 4.

Таблица 4. Общая модель совместного распределения двух зависимых биномиальных случайных величин

	$\xi = 0$	$\xi = 1$	$P\{\eta = \cdot\}$
$\eta = 0$	r	$(1 - 2r)/2$	0.5
$\eta = 1$	$(1 - 2r)/2$	r	0.5
$P\{\xi = \cdot\}$	0.5	0.5	

Эта стохастическая связь между ξ и η характеризуется числом r , а точнее степенью отклонения r от 0.25 (если $r = 0.25$, то случайные величины ξ и η независимы, что проверяется непосредственно).

Заметим, что и в общей двумерной модели (табл. 2) равенство $r = pq$ равносильно независимости ξ и η , тогда как значения r в интервале от pq до $\min(p, q)$ указывают на *положительную* связь между ξ и η , а в интервале от $\max(0, p + q - 1)$ до pq — на *отрицательную*. Этот факт позволяет сразу же ввести *новый* коэффициент связи, который обусловлен границами Фреше–Хефдинга:

$$\rho_{FH} = \frac{r - pq}{(1 - p)(1 - q)} \cdot I\{r > pq\} + \frac{r - pq}{\min\{p(1 - q), q(1 - p)\}} \cdot I\{r \leq pq\},$$

где $I\{B\}$ — индикатор соотношения B . Легко проверить, что этот коэффициент изменяется в границах от -1 до $+1$, показывая степень связи и ее знак ($\rho_{FH} = 0$ тогда и только тогда, когда ξ и η независимы). Проверочные расчеты сравнения выборочных значений ρ_{FH} с широко используемым коэффициентом сопряженности признаков в табл. 1 $r_p = (n_{00}n_{11} - n_{01}n_{10}) / \sqrt{n_{*0}n_{0*}n_{*1}n_{1*}}$ показали их близость. Для модели из табл. 3 *чисто случайно* выбиралось значение r между 0 и 0.25, т. к. $\max(0, p + q - 1) = 0$ и $\min(p, q) = 0.5$ для $p = q = 0.5$, затем с полученным r моделировалась выборка из 100 парных значений (ξ, η) и вычислялись коэффициенты r_{FH} (оценка для ρ_{FH}) и r_p . Этот модельный эксперимент повторялся свыше трех тысяч раз, результаты представлены на рис. 1.

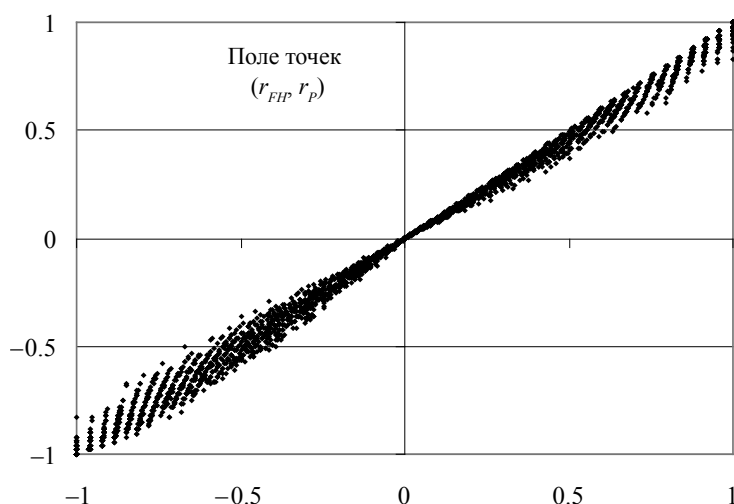


Рис. 1. Результаты моделирования коэффициентов r_{FH} и r_P между двумя случайными величинами, принимающими значения 0 и 1 с равными вероятностями (всего 3315 точек)

В этой упрощенной модели $P\{\xi = 1\} = P\{\eta = 1\} = 0.5$ намного проще выглядят формулы для соответствующей копулы, зависящей только от $r \in [0; 0.5]$:

- а) $C(s, t | r) = 4rst$ в $\mathbf{D}_{00} = \{0 \leq s \leq 0.5, 0 \leq t \leq 0.5\}$;
- б) $C(s, t | r) = 2rs + (1 - 2r)s(2t - 1)$ в $\mathbf{D}_{01} = \{0.5 < s \leq 1, 0 \leq t \leq 0.5\}$;
- в) $C(s, t | r) = 2rt + (1 - 2r)t(2s - 1)$ в $\mathbf{D}_{10} = \{0 \leq s \leq 0.5, 0.5 < t \leq 1\}$;
- г) $C(s, t | r) = r + (1 - 2r)(s + t - 1) + r(2s - 1)(2t - 1)$ в $\mathbf{D}_{11} = \{0.5 < s \leq 1, 0.5 < t \leq 1\}$.

Равенства а)–г) получить несложно, поскольку при $p = q = 0.5$ случайные величины, формирующие распределение $C(s, t | r)$, задаются совсем простыми формулами: $\alpha_1 = (\tau_1 + \xi) / 2$ и $\alpha_2 = (\tau_2 + \eta) / 2$.

В заключение приведем графические иллюстрации распределения $C(s, t | r)$ для трех значений r :

- а) $r = 0.25$ — отсутствие зависимости;
- б) $r = 0.05$ — сильная отрицательная зависимость;
- в) $r = 0.45$ — сильная положительная зависимость.

На рисунке 2 изображены квантованные «карты» значений соответствующих копул $C(s, t | 0.25)$, $C(s, t | 0.05)$ и $C(s, t | 0.45)$ на единичном квадрате с шагом квантования

$\Delta C = 0.05$, т. е. 20 градаций тона по интервалам значений $\left(\frac{k-1}{20}; \frac{k}{20}\right)$, $1 \leq k \leq 20$, от светлого к темному.

Какие же значения коэффициентов r_{FH} и r_P соответствуют словам о «сильной» положительной с $r = 0.45$ и «сильной» отрицательной с $r = 0.05$ связи? С каждым из этих r были смоделированы 375 выборок, по 100 двумерных наблюдений в каждой. В итоге получены две выборки с 750 значениями r_{FH} и r_P в каждой. Сводные результаты представлены в табл. 5.

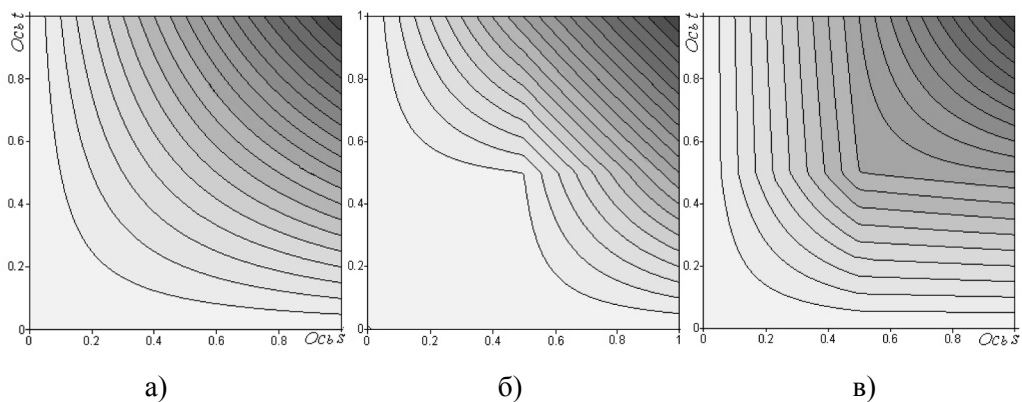


Рис. 2. Графическое представление копул в трех вариантах:

- а) независимые случайные величины ξ и η ($r = 0.25$);
- б) сильная отрицательная связь ($r = 0.05$);
- в) сильная положительная связь ($r = 0.45$)

Таблица 5. Статистические характеристики коэффициентов r_{FH} и r_P , оцениваемых по выборкам размера 100 (рассчитаны по 750 повторным моделированиям)

Статистика	$r = 0.45$		$r = 0.05$	
	r_{FH}	r_P	r_{FH}	r_P
Среднее	0.8438	0.8026	-0.802	-0.799
Сигма	0.0695	0.0628	0.0776	0.0618
Квартиль 0.25	0.7949	0.7607	-0.863	-0.841
Медиана	0.8389	0.8084	-0.801	-0.8
Квартиль 0.75	0.8929	0.8432	-0.756	-0.76

Рассматривая эту таблицу, можно увидеть, что в распределении этих коэффициентов есть некая «кособокость», и она у них несколько разная, хотя различия между ними невелики.

7. Заключение

Сейчас копулами увлечены многие, особенно в таких направлениях деятельности, как управление финансами, анализ страховых рисков, моделирование макроэкономических процессов. Анализ публикаций, информацию о которых можно найти в Интернете, показывает, что практически во всех областях использования статистических данных появляются работы с привлечением копул для моделирования зависимостей. Приведем только три примера: классификация дистанционных изображений (Крылов, 2010), внедрение сетей сотовой подвижной связи третьего поколения (Поскакухин, 2009) и долговечность машин и агрегатов, обусловленная когерентной надежностью их деталей (Eryilmaz, 2011). С другой стороны, знакомство с десятками публикаций самого разного сорта приводит к неутешительному выводу о том, что в настоящее время вся практика использования копул представляет «метод

тыка». А именно, никто пока не знает, как осмысливать те зависимости, которые «прячутся» в n -мерных копулах (уже при $n > 3$), хотя попытки осмысления существуют.

В настоящей статье нет, по существу, ничего, что было бы неизвестно, это — очень субъективное изложение «самых главных» сведений о копулах, которые необходимо знать всякому прикладному статистику. Вот краткий перечень того, что представляется «самым главным»:

1. Если многомерное распределение наблюдаемых признаков имеет плотность (неважно, известную или неизвестную), то его копула — это совместное распределение тех же признаков после их равнометризации (величина $\alpha = F(\xi)$, где $F(x)$ — функция распределения признака ξ , является его равнометризацией). Для дискретных распределений идея копулы имеет теоретический смысл, а в реальных исследованиях эксплуатировать копулу в такой ситуации едва ли разумно. Это легко понять из включенного в статью раздела 6 о таблицах сопряженности (2×2) и связанных с ними копулах.

2. Всякая копула является гладкой функцией, и можно считать, что все используемые на практике копулы имеют плотность.

3. Все конструктивное множество копул, из которого извлекаются копулы для анализа данных в конкретных исследованиях, можно рассматривать как состоящее из трех типов: эллиптические, архимедовы и креативные. К последнему типу можно отнести все копулы, которые были изобретены разными авторами в разное время как некое конкретное параметрическое семейство функций или сконструированы из других копул (с помощью умножения, преобразования, взвеси, пертурбации).

Следуя этому представлению о «главном», в статье всякий раз, когда это можно сделать без обращения к высшим материям математики, дается доказательство особенных свойств копулы. Единственным исключением стало доказательство теоремы Склара из статьи (Rüschendorf, 2009), идея которого сама по себе имеет содержательную составляющую. И надо сделать еще одно суммирующее замечание. Теория копул переживает сейчас тот же период, какой был лет пятьдесят тому назад в теории распределений: искали новые семейства, изобретали способы оценивания параметров, осмысливали разные особенности как отдельных типов распределений, так и их широких классов.

Из проведенного анализа литературы можно сделать один важный вывод: использование копул в разных статистических задачах только начинается, но их работоспособность всегда будет зависеть не от мощности технических средств, а от культуры анализа данных, от понимания предметной области, поставляющей эти данные и дающей ключ к интерпретации зависимостей в копуле.

Список литературы

Крылов В. А. (2010). Моделирование и классификация многоканальных дистанционных изображений с использованием копул. *Информатика и ее применения*, 4 (4), 33–37.

Пеникас Г. И. (2011). Модели «копула» в задачах хеджирования ценового риска. *Прикладная эконометрика*, 22 (2), 3–21.

Посакухин В. Н. (2009). Проблемы внедрения сетей сотовой подвижной связи третьего поколения в Московском регионе. *T-Сотт — Телекоммуникации и Транспорт*, спецвыпуск «Технологии информационного общества», Часть 1, Июнь 2009. 124–128.

Фантаццини Д. (2011). Моделирование многомерных распределений с использованием копула-функций. I. *Прикладная эконометрика*, 22 (2), 98–134.

Bouzebda S., Keziou A. (2010). New estimates and tests of independence in semiparametric copula models. *Kybernetika (Prague)*, 46 (1), 178–201.

Cuadras C. M., Augé J. (1981). A continuous general multivariate distribution and its properties. *Communication in Statistics — Theory and Methods*, 10 (4), 339–353.

Darsow W. F., Nguyen B., Olsen E. T. (1992). Copulas and Markov process. *Illinois Journal of Mathematics*, 36 (4), 600–642.

Durante F., Sempi C. (2010). Copula theory: An introduction. In: Jaworski P., Durante F., Härdle W., Rychlik T. (eds), *Copula Theory and Its Applications*, Proceedings of the Workshop, Warsaw, 25–26 September 2009, Springer.

Eryilmaz S. (2011). Estimation in coherent reliability system through copulas. *Reliability Engineering & System Safety*, 96 (5), 564–568.

Genest C., MacKay R. J. (1986). The joy of copulas: Bivariate distributions with uniform marginals. *American Statistician*, 40 (4), 280–283.

Genest C., Rivest L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88, 1034–1043.

McNeil A. J., Nešlehová J. (2009). Multivariate Archimedean copulas, d -monotone functions and ℓ_1 -norm symmetric distributions. *Ann. Statist.*, 37 (5B), 3059–3097.

Mikusinski P., Sherwood H., Taylor M. D. (1991). Probabilistic interpretations of copulas and their convex sums. In: *Advances in probability distributions with given marginals* (Rome, 1990), *Math. Appl.*, Kluwer Acad. Publ., Dordrecht, 67, 95–112.

Nelsen R. B. (1999). *An introduction to copulas*. Lecture Notes in Statistics, 139. NY: Springer-Verlag.

Rüschendorf L. (2009). On the distributional transform, Sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139 (11), 3921–3927.