DOI: 10.34020/2073-6495-2021-1-161-167

УДК 311, 519.2

О КОЭФФИЦИЕНТАХ ЭКСТРЕМАЛЬНОЙ ЗАВИСИМОСТИ ГАУССОВСКИХ РАСПРЕДЕЛЕНИЙ

Хрущев С.Е.

Новосибирский государственный университет
экономики и управления «НИНХ»
E-mail: s.e.hrushchev@edu.nsuem.ru

В работе рассматривается способ представления зависимости между показателями в виде копул. Копулы являются популярным математическим инструментарием. Это обусловлено тем, что, с одной стороны, в копулах выделены маргинальные распределения показателей, а с другой стороны, выделена структура зависимости между данными маргинальными распределениями, что позволяет весьма эффективно изучать связи, возникающие в реальных совокупностях. Отдельное внимание в работе уделяется коэффициентам экстремальной зависимости – важным числовым характеристикам связи в условиях экстремальных малых или экстремально больших значениях показателей. Показано, что даже в условиях тесной корреляции между показателями у двумерного гауссовского распределения нижний и верхний коэффициенты экстремальной зависимости принимают нулевые значения. Это говорит о невозможности прогнозирования значений одного показателя при фиксации слишком малых или слишком больших значений другого показателя. В настоящей работе показано, что связь между количеством заражений коронавирусной инфекцией COVID-19 на 100 000 человек и количество смертей от коронавирусной инфекции COVID-19 на 100 000 человек по регионам Российской Федерации может быть представлена в виде гауссовской копулы.

Ключевые слова: копула, гауссовское распределение, коэффициенты экстремальной зависимости, коронавирусная инфекция COVID-19.

ON THE EXTREMAL DEPENDENCE COEFFICIENTS OF GAUSSIAN DISTRIBUTIONS

Khrushchev S.E.

Novosibirsk State University of Economics and Management
E-mail: s.e.hrushchev@edu.nsuem.ru

The paper considers a way to represent the relationship between indicators in the form of copulas. Copulas are popular mathematical tools. This is due to the fact that, on the one hand, the marginal distributions of indicators are divided in the copulas, and on the other hand, the structure of the relationship between these marginal distributions is divided, which makes it possible to very effectively study the connections that arise in real populations. Special attention in the work is paid to extremal dependence coefficients - important numerical characteristics of the connection in conditions of extreme small or extremely large values of indicators. It is shown that even under conditions of close correlation between the indices for a two-dimensional Gaussian distribution, the lower and upper coefficients of the extreme dependence take zero values. This indicates the impossibility of predicting the values of one indicator when fixing too small or too large values of another indicator. This work shows that the relationship between the number of COVID-19 coronavirus infections per 100,000 people and the number of deaths from COVID-19 coronavirus infection per 100,000 people in the regions of the Russian Federation can be represented in the form of a Gaussian copula.

Keywords: copula, Gaussian distribution, extremal dependence coefficients, COVID-19 coronavirus infection.

1. ВВЕДЕНИЕ

Для описания зависимостей, возникающих между двумя показателями, используется совместная функция распределения. Хорошо известно, что совместное распределение содержит всю необходимую информацию для характеризации взаимосвязи между рассматриваемыми показателями, но оно не всегда удобно для описания численной меры связи. В такой ситуации удобно использовать числовые коэффициенты (коэффициент корреляции Пирсона, коэффициент ранговой корреляции Спирмена, коэффициент корреляции Кенделла), которые в том или ином виде отражают числовую меру зависимости. Отметим, что данные коэффициенты описывают некую среднюю зависимость (линейную или нелинейную) между изучаемыми показателями, но не всегда отражают связь, которая возникает при экстремальных (очень больших или очень малых) значениях каждого из них. Анализ экстремальных значений показателей привел к появлению теории экстремальных значений.

Теория экстремальных значений является важным инструментом при моделировании различных явлений (см, например, [6, 9, 12]), в которых достигаются большие или малые значения показателей (при моделировании катастрофических событий, землетрясений, цунами, экономических кризисов и др.). Как правило, в теории экстремальных значений связи между показателями описываются с помощью копул, а мерой числовой зависимости служат коэффициенты экстремальной зависимости.

В настоящей работе мы будем рассматривать случаи, когда парная связь задается с помощью совместного гауссовского распределения или гауссовской копулы – наиболее распространенной ситуации, возникающей в прикладных задачах. Для данного гауссовского распределения изучим поведение коэффициентов экстремальной зависимости.

2. КОПУЛЫ

Изучение копул и их применение в статистике набирает большую популярность. В настоящей работе отметим лишь необходимые нам результаты из теории копул. Достаточно полную информацию можно найти в [3, 4, 7, 8]. Мы будем рассматривать двумерные копулы.

Копулой двух случайных величин U и V, имеющих равномерное распределение на отрезке от 0 до 1, называется их совместная функция распределения, т.е.

$$C(u,v)=P(U<u,V<v),\text{ где } u,v\in[0,1].$$

Пусть изучаются два показателя X и Y. Будем предполагать, что их распределение непрерывно. Тогда по теореме Склара [10, 11] существует такая копула $C(u, v)$, что совместная функция распределения показателей X и Y представляется в виде:

$$F(x,y)=P(X<x,Y<y)=C(F(x),F(y)),$$

где $F_X(x)$, $F_Y(y)$ – маргинальные функции распределения случайных величин X и Y соответственно.

Получили, что совместная функция распределения представима в виде копулы, т.е. имеет место быть альтернативная форма представления зависимостей, возникающих между рассматриваемыми показателями X и Y. Копулы дают возможность разделить описание распределения случайного вектора на две части: частные распределения компонент и структура их зависимостей [1].

Если совместное распределение показателей X и Y является гауссовским с коэффициентом корреляции $-1<\rho<1$, то соответстувующая ей копула из теоремы Склара называется гауссовской и, как хорошо известно [8], имеет вид:

$$C(u,v)=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\exp\left\{\frac{-(x^2-2\rho xy+y^2)}{2(1-\rho^2)}\right\}dxdy,$$

где $\Phi^{-1}(u)$ – квантиль стандартного нормального распределения уровня u.

Получили, что гауссовская копула выражается через коэффициент корреляции, следовательно, вся информация о зависимости между элементами совместного гауссовского распределения хранится в этом коэффициенте, что еще раз объясняет его популярность при изучении гауссовских связей.

3. КОЭФФИЦИЕНТЫ ЭКСТРЕМАЛЬНОЙ ЗАВИСИМОСТИ

Описание зависимостей между экстремальными данными является полезным инструментом (см., например, [2, 5]). Коэффициенты экстремальной зависимости становятся популярными в управлении рисками. Выделяют нижний и верхний коэффициенты экстремальной зависимости для соответственно малых и больших значений показателей.

Нижний коэффициент экстремальной зависимости для двух случайных величин X и Y с непрерывными функциями распределения $F_X(x)$, $F_Y(y)$ соответственно определяется соотношением

$$\lambda_{low}=\lim_{t\to 0}P(Y<F_Y^{-1}(t)\,|\,X<F_X^{-1}(t))=\lim_{t\to 0}\frac{C(t,t)}{t},$$

где $C(t, t)$ – копула случайных величин X и Y при условии, что данный предел существует. Если $\lambda>0$, то будем говорить, что случайные величины X

и Y зависят при экстремально малых значениях, в противном случае будем говорить, что указанной зависимости нет.

Верхний коэффициент экстремальной зависимости для двух случайных величин X и Y с непрерывными функциями распределения $F_X(x)$, $F_Y(y)$ соответственно определяется соотношением

$$\lambda_{up} = \lim_{t\to 1} P(Y \geq F_Y^{-1}(t) \mid X \geq F_X^{-1}(t)) = \lim_{t\to 1} \frac{1-2t+C(t,t)}{t},$$

где $C(t, t)$ – копула случайных величин X и Y при условии, что данный предел существует. Если $\lambda_{up} > 0$, то будем говорить, что случайные величины X и Y зависят при экстремально больших значениях, в противном случае будем говорить, что указанной зависимости нет.

Докажем, что нижний и верхний коэффициенты экстремальной зависимости для случайных величин X и Y, имеющих совместное гауссовское распределение с коэффициентом корреляции $-1 < \rho < 1$, равны нулю.

Действительно, как уже отмечалось выше, имеет место представление

$$C(t,t) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(t)} \int_{-\infty}^{\Phi^{-1}(t)} \exp\left\{\frac{-(x^2-2\rho xy+y^2)}{2(1-\rho^2)}\right\} dxdy.$$

Продифференцировав по формуле Лейбница данную функцию, получим

$$\frac{dC(t,t)}{dt} = \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(t)} \exp\left\{\frac{-(s-\rho\Phi^{-1}(t))^2}{2(1-\rho^2)}\right\} ds = 2\int_{-\infty}^{\Phi^{-1}(t)} \varphi(s)ds,$$

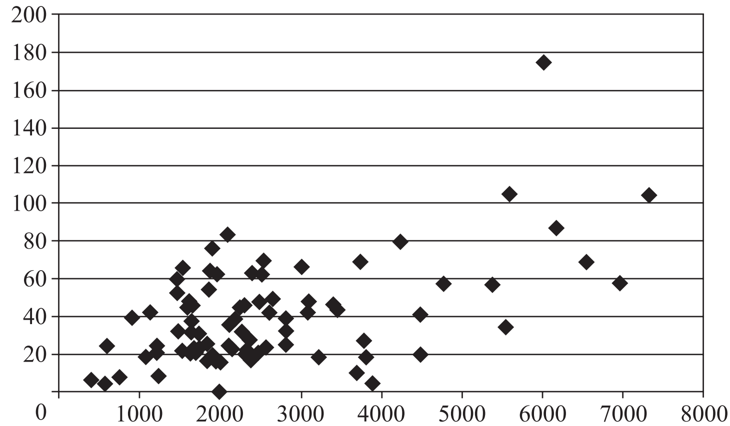
где подынтегральная функция $\varphi(s)$ является функцией плотности гауссовского распределения со средним значением $\rho\Phi^{-1}(t)$ и со стандартным отклонением $\sqrt{1-\rho^2}$. Значит, $\lim_{t\to 0}\frac{dC(t,t)}{dt} = 0$, а $\lim_{t\to 1}\frac{dC(t,t)}{dt} = 2$. Применяя правило Лопиталя для нахождения коэффициентов экстремальной зависимости, получаем

$$\lambda_{low} = 0, \quad \lambda_{up} = 0.$$

Это говорит о том, что в гауссовском случае зависимость между экстремально малыми или экстремально большими значениями показателей X и Y не наблюдается, даже в случаях тесной корреляции между данными случайными величиными. С одной стороны, если имеет место сильная корреляция между случайными величинами, то мы можем построить качественный линейный прогноз для одного из показателей, фиксируя некие допустимые значения другого показателя. А с другой стороны, если рассматриваются эксремально малые или экстремально большие значения, то такой прогноз строить весьма опасно. В этом случае теряется зависимость между показателями, что объясняется нулевыми значениями коэффициентов экстремальной зависимости.

4. ПРИМЕР

Пусть показатель X – количество заражений коронавирусной инфекцией COVID-19 на 100 000 человек, Y – количество смертей от коронавирусной инфекции COVID-19 на 100 000 человек. Изучается связь между этими



Корреляционное поле между показателями X и Y

показателями. Информационный массив данных по регионам Российской Федерации взят с сайта https://yandex.ru/covid19/stat на 25 января 2021 г. Корреляционное поле представлено на рисунке, где по осям абсцисс и ординат откладываются соответственно значения показателей X и Y.

Визуальный анализ корреляционного поля позволяет сделать вывод, что между показателями наблюдается слабая линейная связь. Об этом же свидетельствует значение коэффициента корреляции, равное 0,53 (с помощью критерия Стьюдента подтвердилась гипотеза о значимости этого коэффициента).

С помощью критерия Пирсона была проверена гипотеза о том, что показатели X и Y имеют двумерное гауссовское распределение. Реально достигнутый уровень значимости составил 0,073, что свидетельствует в пользу принятия нулевой гипотезы (о гауссовском распределении рассматриваемых показателей). А значит связь между рассматриваемыми показателями может быть описана с помощью гауссовской копулы.

С одной стороны, очевидно, что рост заражений приводит к росту смертей, но с другой стороны, следует с особой осторожностью строить какие-то прогнозы даже в рамках математической модели исследуемой взаимосвязи, в силу нулевых значений коэффициентов экстремальной зависимости для гауссовского случая.

5. ЗАКЛЮЧЕНИЕ

В работе рассматривается способ представления зависимости между показателями в виде копул. Данное представление оказывается удобнее, чем использование совместных функций распределений изучаемых показателей. Это обусловлено тем, что, с одной стороны, в копулах выделены маргинальные распределения показателей, что, конечно же, играет важную роль при исследовании реальных совокупностей, а с другой – выделена структура зависимости между найденными маргинальными распределениями.

Отдельное внимание в работе уделяется коэффициентам экстремальной зависимости – важным числовым характеристикам связи в условиях экстремальных малых или экстремально больших значениях показателей.

Показано, что даже в условиях тесной корреляции между показателями у двумерного гауссовского распределения нижний и верхний коэффициенты экстремальной зависимости принимают нулевые значения. Это говорит о невозможности прогнозирования значений одного показателя при фиксации слишком малых или слишком больших значений другого показателя.

В настоящей работе показано, что связь между количеством заражений коронавирусной инфекцией COVID-19 на 100 000 человек и количеством смертей от коронавирусной инфекции COVID-19 на 100 000 человек по регионам Российской Федерации могут быть представлены в виде гауссовской копулы.

Литература

1. *Фантаццини Д*. Моделирование многомерных распределений с использованием копула-функций. I // Прикладная эконометрика. 2011. № 2 (22). С. 98–134.
2. *Ane T., Kharoubi C*. Dependence structure and risk measure. J. Business. 2003. 76:3. P. 411–438.
3. *Cherubini U., Vecchiato W., Luciano E*. Copula methods in finance. Wiley, 2004.
4. *Joe H*. Multivariate models and dependence concepts. London: Chapman Hall, 1997.
5. *Junker M., May A*. Measurement of aggregate risk with copulas. Econom. J. 2005. 8:3. P. 428–454.
6. *Lambert P., Vandenhende F*. A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. Statistics in Medicine. 2002. 21. P. 3197–3217.
7. *Malevergne Y., Sornette D*. Extreme financial risks (From dependence to risk management). Heidelberg: Springer, 2006.
8. *Nelsen R.B*. An introduction to copulas. Lecture Notes in Statistics, 2nd Ed. New York: Springer-Verlag, 2006. 276 p.
9. *Salvadori G., De Michele C*. On the use of copulas in hydrology: Theory and practice. Journal of Hydrologic Engineering. 2007. 12 (4). P. 369–380.
10. *Sklar A*. Fonctions de répartition á n dimensions et leurs marges. Publ. Inst. Statis. Univ. Paris, 1959. 8. P. 229–231.
11. *Sklar A*. Random variables, distribution functions, and copulas: Personal look backward and forward. Lecture notes. Monograph series. 1996. 28. P. 1–14.
12. *Zhang L., Singh V*. Bivariate flood frequency analysis using the copula method. Journal of Hydrologic Engineering. 2006. 11 (2). P. 150–164.

Bibliography

1. *Fantatstsini D*. Modelirovanie mnogomernykh raspredelenii s ispol'zovaniem kopula-funktsii. I // Prikladnaya ekonometrika. 2011. № 2 (22). P. 98–134.
2. *Ane T., Kharoubi C*. Dependence structure and risk measure. J. Business. 2003. 76:3. P. 411–438.
3. *Cherubini U., Vecchiato W., Luciano E*. Copula methods in finance. Wiley, 2004.
4. *Joe H*. Multivariate models and dependence concepts. London: Chapman Hall, 1997.
5. *Junker M., May A*. Measurement of aggregate risk with copulas. Econom. J. 2005. 8:3. P. 428–454.
6. *Lambert P., Vandenhende F*. A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. Statistics in Medicine. 2002. 21. P. 3197–3217.
7. *Malevergne Y., Sornette D*. Extreme financial risks (From dependence to risk management). Heidelberg: Springer, 2006.

8. *Nelsen R.B*. An introduction to copulas. Lecture Notes in Statistics, 2nd Ed. New York: Springer-Verlag, 2006. 276 p.
9. *Salvadori G., De Michele C*. On the use of copulas in hydrology: Theory and practice. Journal of Hydrologic Engineering. 2007. 12 (4). P. 369–380.
10. *Sklar A*. Fonctions de répartition á n dimensions et leurs marges. Publ. Inst. Statis. Univ. Paris, 1959. 8. P. 229–231.
11. *Sklar A*. Random variables, distribution functions, and copulas: Personal look backward and forward. Lecture notes. Monograph series. 1996. 28. P. 1–14.
12. *Zhang L., Singh V*. Bivariate flood frequency analysis using the copula method. Journal of Hydrologic Engineering. 2006. 11 (2). P. 150–164.