© В.А. СИМАХИН
sva_full@mail.ru

УДК 519.234.2

УСЛОВНЫЕ МЕРЫ ЗАВИСИМОСТИ

АННОТАЦИЯ. В работе рассмотрены условные меры зависимости случайных величин.

SUMMARY. The article considers the conditional dependence measures of random variables.

КЛЮЧЕВЫЕ СЛОВА. Условные меры зависимости, частная корреляция.
KEY WORDS. Conditional dependence measures, private correlation.

Введение. Практика статистических исследований задач многомерного анализа показывает, что для адекватного содержательного объяснения парных корреляционных связей факторов необходимо учитывать опосредованное влияние третьих связанных с ними факторов [1-5]. Это обстоятельство делает необходимым введение таких измерителей статистической связи между факторами, которые были бы «очищены» от опосредованного влияния других факторов. Такие меры зависимости должны давать оценку степени тесноты интересующей нас статистической связи между парами факторов при условии, что остальные факторы зафиксированы на некотором постоянном уровне (условные меры зависимости). В этом случае говорят о статистическом анализе *частных* связей и используют соответственно частные коэффициенты корреляции или другие корреляционные характеристики [3]. Частный коэффициент корреляции был введен Пирсоном [6] свыше 100 лет назад для нормального распределения, а формула для его вычисления по аналогии, используется в общем случае для любого многомерного распределения [1-2]. Частный коэффициент корреляции не зависит от фиксированных уровней мешающих факторов и вычисляется через парные коэффициенты корреляции [1-3]. В то же время условные меры статистической связи должны зависеть от заданных уровней мешающих факторов. Этот недостаток при вычислении частного коэффициента корреляции можно ликвидировать путем организации выборки специальной структуры, обеспечивающей наличие хотя бы нескольких наблюдений при каждом фиксированном значении мешающего фактора. Необходимо также отметить, что частный коэффициент корреляции является мерой линейной связи между факторами. Последнее обстоятельство в случае оценки парных связей привело к тому, что в практику статистических исследований было эвристически введено большое количество мер зависимости, позволяющих учитывать нелинейный характер стохастической зависимости между случайными величинами [3-12]. В итоге были сформулированы требования, которым должна удовлетворять мера зависимости в общем случае [9]. Меры парной зависимости определяются в виде функционалов от функций распределения, а их непараметрические оценки получались в классе эмпирических функционалов Мизеса и U-статистик [12]. В связи с развитием теории непараметрических оценок условных функционалов непараметрическая оценка частного коэффициента корреляции была получена сравнительно недавно [13], а алгоритм ее вычисления выглядит совсем иначе, чем его классический вариант [1-3].

В данной работе вводятся условные меры зависимости, представленные в виде функционалов от условных распределений для любых распределений, обобщающие классический частный коэффициент корреляции.

Постановка задачи. Обозначения. Пусть (X, Y, T) — случайная векторная величина с функцией распределения (ф.р.) $F(x, y, t)$ и плотностью $f(x, y, t)$. Пусть также $F_1(x)$; $F_2(y)$; $F_1(x/t)$; $F_2(y/t)$; $F(x, y/t)$; $f_1(x)$; $f_2(y)$; $f_1(x/t)$; $f_2(y/t)$; $f(x, y/t)$ — маргинальные и условные ф.р. и плотности X и Y и условии $T=t$, соответственно.

Требуется определить условную меру стохастической зависимости между случайными величинами X и Y при условии $T=t$. Случайная величина T может быть вектором.

В общем случае условная мера зависимости R(t) может быть определена в виде функционала $J(t)$ от неизвестных ф.р. $F_1(x/t)$; $F_2(y/t)$; $F_1(x, y/t)$ и их производных в виде $R(t)=J[F(x, y/t)]$. В непараметрической постановке задачи условные ф. р. считаются неизвестными.

Введем и рассмотрим некоторые условные меры зависимости между случайными величинами X и Y.

Классические меры зависимости

Условный коэффициент корреляции. Условный коэффициент корреляции определим следующим образом

$$R_1(t) = \frac{\operatorname{cov}(X, Y/t))}{\sqrt{D(X/t)D(Y/t)}}, \tag{1}$$

где $cov(X, Y/t)=\iint xydF(x, y/t) - \int xdF_1(x/t)\int ydF_2(y/t)$ и $D(X/t)$, $D(Y/t)$ — условные дисперсии случайных величин X и Y, соответственно.

$R_1(t)$ является условной мерой линейной связи двух случайных величин.

В случае нормальных распределений $R_1(t)$ является коэффициентом корреляции условного двумерного нормального распределения и совпадает с частным коэффициентом корреляции [1-6].

Условный коэффициент угловой связи. Простейшей мерой зависимости между случайными величинами X и Y является коэффициент угловой связи R, который определяется как сумма вероятностей нахождения (X, Y) в первом и третьем квадранте декартовой системы координат [4-5]. Обычно R определяется по отношению к некоторой фиксированной точке (x_0, y_0), выбор которой достаточно произволен. Обычно в качестве (x_0, y_0) берут медианные значения M_1 и M_2 случайных величин X и Y.

Введем условную квадрантную меру зависимости в виде

$$R_2(t)=\iint Sign(x - x_0)Sign(y - y_0)dF(x, y/t), \tag{2}$$

где

$$Sign(x)=\begin{cases} 1, & x>0, \\ -1, & x<0. \end{cases}$$

и $F_1(M_1/t)=0,5$; $F_2(M_2/t)=0,5$ — условные медианы X и Y — соответственно.

Можно показать, что $R_2(t) = \frac{2}{\pi}\arcsin(R_1(t))$ для нормальных распределений.

Условный коэффициент согласованности (конкордации). Возможность естественным путем устранить произвол в выборе точки (x_0, y_0) в $R_2(t)$ — это считать точку (x_0, y_0) случайной и усреднить по всем (x_0, y_0). В итоге мы приходим к условной мере зависимости, которую по аналогии назовем условным коэффициентом согласованности (конкордации)

$$R_3(t)=\iint\iint Sign(x-x_0)Sign(y-y_0)dF(x,y/t)dF(x_0,y_0/t) \qquad (3)$$

Условные ранговые коэффициенты связи. Если для получения $R_3(t)$ мы усредняли $R_2(t)$ по всем значениям (x_0, y_0) с распределением $F(x, y/t)$, то не лишено смысла усреднить эту вероятность по маргинальным распределениям $F_1(x/t)$ и $F_2(y/t)$, взятым независимо.

Таким образом, мы приходим к следующей условной мере зависимости

$$Q(t)=\iint\iint Sign(x-x_0)Sign(y-y_0)dF(x,y/t)dF(x_0/t)dF_2(y_0/t) \qquad (4)$$

$Q(t)$ порождает аналоги условных ранговых коэффициентов связи

$$R_5(t)=3Q(t)-3 \qquad (5)$$

- условный коэффициент ранговой корреляции Спирмена и

$$R_6(t)=2Q(t)-1 \qquad (6)$$

- условный коэффициент ранговой корреляции Пирсона.

Можно рассмотреть и другие меры зависимости, основанные на различных расстояниях между распределениями $F(x, y/t)$ и $F_1(x/t)$ $F_2(y/t)$, например, на расстояниях Кульбака-Лейблера — $\rho_1(t)$, Хеллингера — $\rho_2(t)$, $\chi^2 - \rho_3(t)$, где:

$$\rho_1(t) \doteq \iint \ln\left[\frac{f(x,y/t)}{f_1(x/t)f_2(y/t)}\right]dF(x,y/t);$$

$$\rho_2(t)=\iint\left[\sqrt{f(x,y/t)}-\sqrt{f_1(x/t)f_2(y/t)}\right]^2 dxdy=\iint\left[1-\sqrt{\frac{f_1(x/t)f_2(y/t)}{f(x,y/t)}}\right]^2 dF(x,y/t);$$

$$\rho_3=\iint\left[\frac{f(x,y)-f_1(x)f_2(y)}{f(x,y)}\right]^2 dF(x,y).$$

Имеют место соотношения:

$$\rho_1 \approx \frac{1}{2}\rho_2;\ \rho_2 \approx 4\rho_3;\ \rho_2 \geq \rho_3;\ \rho_1 \geq \rho_3$$

Если с помощью этих расстояний характеризовать степень близости распределений, когда отношение $\dfrac{f(x,y/t)}{f_1(x/t)f_2(y/t)}$ близко к 1, то асимптотически все они ведут себя одинаково с точностью до постоянных множителей. Имеются и другие меры зависимости [12]. Поэтому ниже мы рассмотрим только меру зависимости, введенную Линфутом [7], основанную на расстоянии ρ_1.

Условный коэффициент Линфута. Реньи [9] сформулировал семь аксиом, которым должна удовлетворять парная мера зависимости $R(x, y)$ в общем случае. Аксиомы Реньи позволили не только прояснить вопрос и упорядочить хорошо известные меры зависимости, но и поставить вопрос поиска и введения в повседневную практику только таких мер зависимости, которые удовлетворяют всем аксиомам Реньи. Эти аксиомы обобщаются и на условные меры зависимости. Линфутом [7], на основе расстояния Кульбака-Лейблера, была введена информационная мера зависимости, удовлетворяющая всем аксиомам Реньи, которую можно обобщить на условные распределения

$$R_7(t)=[1-exp(-2I(x, y/t))]^{½}, \qquad (7)$$

где $I(x,y/t)=\iint \ln\left[\dfrac{f(x,y/t)}{f_1(x/t)f_2(y/t)}\right]dF(x,y/t)$.

Например, можно показать, что для нормального распределения $I(x,y/t)=-\dfrac{1}{2}\ln(1-R_1^2(t))$.

Теоретико-информационный подход к построению мер зависимости оказался достаточно плодотворным и позволил найти меры зависимости в рамках подмножества аксиом Реньи [5, 10].

Выводы

Интерес к задаче «условных мер зависимости» был вызван рядом моментов. Практические задачи многомерной статистики в непараметрической постановке (регрессионный анализ, редукция факторного пространства ...) требовали анализа степени тесноты связей в общем случае. Обращение к литературе [1-5] показывает, что в этом вопросе мы можем опираться только на частный коэффициент корреляции, который имеет ряд недостатков, не позволяющих его использование в общем случае.

В данной работе введен целый ряд условных мер зависимости, в виде функционалов от условных распределений ((3.1)-(3.6), (4.1)). Условный коэффициент Линфута (4.1) удовлетворяет всем аксиомам Реньи в общем случае. Запись мер зависимости в функциональном виде позволяет получать непараметрические оценки этих функционалов в виде условных функционалов Мизеса и условных U-статистик.

СПИСОК ЛИТЕРАТУРЫ

1. Кендалл М. Дж., Стюарт А. Статистические выводы и связи. М.: Наука, 1973. 899 с.
2. Андерсон Т. Введение в многомерный статистический анализ. М.: Наука, 1963. 500 с.
3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. М.: Финансы и статистика. 1985. 487 с.
4. Кендэл М. Ранговые корреляции. М.: Статистика, 1975. 214 с.
5. Елисеева И.И., Рукавишников В.О. Группировка, корреляция, распознавание образов. М.: Статистика, 1977. 144 с.
6. Pearson, K. Mathematical contributions to the theory of evolution — III. Regression, heredity and panmixia // Phil. Trans. 1986. Pp. 253-318
7. Linfoot, E.H. An information measure of correlation // Information and control. 1957. № 1. Pp. 85-89.
8. Kruskal, W.H. Ordinal measures of association // J. of the Amer. Statist. Assoc. 1958. № 53, Pp. 814-861.
9. Renyi, A. On measures of dependence // Acta Math. Acad. Scien. Hung. 1959. 10. № 3-4. Pp. 441-451.
10. Bell, C.B. Mutual information and maximal correlation as measures of dependence // Ann. Math. Stat. 1962. 33, Pp. 587-593.
11. Lehmann, E.L. Some consepts of dependence // Ann. Math. Stat. 1966. 37. Pp. 1137-1153.
12. Симахин В.А. Меры зависимости и их непараметрические оценки // Вестник Курган. гос. ун-та. 2010. №1 (17), серия «Технические науки», Вып. №5. С. 150-155.
13. Васильев В.А., Добровидов А.В., Кошкин Г.М. Непараметрическое оценивание функционалов от распределений стационарных последовательностей: М.: Наука, 2004, 508 с.