ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2023 Управление, вычислительная техника и информатика № 62
Tomsk State University Journal of Control and Computer Science

Научная статья
УДК 681.51:519.6
doi: 10.17223/19988605/62/10

Вычисление коэффициента корреляции случайных величин с использованием дополнительной информации о форме их взаимосвязи

Виктор Анатольевич Удод

Томский государственный университет, Томск, Россия, pr.udod@mail.ru

Аннотация. Задача вычисления коэффициента корреляции двух случайных величин рассматривается в предположении, что математическое ожидание их произведения неизвестно, но известны первые два момента распределения каждой из них. Решение предлагается для случая, когда набор возможных значений одной величины определенным образом зависит от фиксированного значения другой величины. Показано, что с использованием дополнительной информации о форме взаимосвязи исследуемых случайных величин можно существенно упростить расчет коэффициента корреляции.

Ключевые слова: случайные величины; коэффициент корреляции; условное математическое ожидание.

***Для цитирования*:** Удод В.А. Вычисление коэффициента корреляции случайных величин с использованием дополнительной информации о форме их взаимосвязи // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 62. С. 92–100. doi: 10.17223/19988605/62/10

Original article
doi: 10.17223/19988605/62/10

Calculation of the correlation coefficient of random variables using additional information about the form of their relationship

Victor A. Udod

Tomsk State University, Tomsk, Russian Federation, pr.udod@mail.ru

Abstract. The problem of calculating the correlation coefficient of two random variables is considered under the assumption that the mathematical expectation of their product is unknown, but the first two moments of the distribution of each of them are known. The solution is proposed for the case when the set of possible values of one quantity depends in a certain way on a fixed value of another quantity. It is shown that with the use of additional information about the form of the relationship of the studied random variables, it is possible to significantly simplify the calculation of the correlation coefficient.

Keywords: random variables; correlation coefficient; conditional mathematical expectation.

***For citation*:** Udod, V.A. (2023) Calculation of the correlation coefficient of random variables using additional information about the form of their relationship. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitelnaja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 62. pp. 92–100. doi: 10.17223/19988605/62/10

Введение

В настоящее время корреляционный анализ широко используется практически во всех сферах науки и техники: в физике [1], химии [2], биологии [3], медицине [4], экономике [5], строительстве [6], а также в образовании [7] и т.д. При этом значительная часть исследований при решении той или иной задачи зачастую сводится к проведению парного корреляционного анализа, а именно к вычислению коэффициента корреляции Пирсона r_{xy} случайных величин X и Y:

$$r_{xy} = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{M(XY) - M(X)M(Y)}{\sqrt{D(X)D(Y)}}\,. \tag{1}$$

Здесь $\operatorname{cov}(X,Y) = M(XY) - M(X)M(Y)$ – ковариация случайных величин X и Y; $M(XY)$ – математическое ожидание произведения случайных величин X и Y; $M(X)$ и $M(Y)$ – математические ожидания, $\sigma_x = \sqrt{D(X)}$ и $\sigma_y = \sqrt{D(Y)}$ – средние квадратические отклонения, а $D(X)$ и $D(Y)$ – дисперсии случайных величин X и Y соответственно.

При вычислении коэффициента корреляции возможна ситуация, когда известны лишь числовые характеристики $M(X)$, $M(Y)$, $D(X)$ и $D(Y)$ исследуемых случайных величин X и Y, а $M(XY)$ неизвестно. В этом случае применение формулы (1) становится затруднительным либо вовсе невозможным. Определенным выходом из такой ситуации может служить получение статистической оценки параметра r_{xy}, т.е. вычисление выборочного коэффициента корреляции r^*_{xy}. Подобный подход часто применяется при решении как общетеоретических, так и прикладных задач. Между тем если для вычисления выборочного коэффициента корреляции r^*_{xy} необходимо проведение большого объема физических (химических, биологических и т.п.) экспериментов или имитационного моделирования, то эффективность такого подхода заметно снижается.

Таким образом, актуальной является задача вычисления коэффициента корреляции r_{xy} на основе первых двух моментов распределения случайных величин X и Y и некоторой дополнительной (априорной) информации, характеризующей взаимосвязь между этими величинами. Решению данной задачи применительно к некоторым случайным величинам с определенной формой их взаимосвязи и посвящена настоящая работа.

1. Постановка задачи

Пусть X и Y есть случайные величины, удовлетворяющие определенным предположениям, описание которых приведено ниже.

Если X и Y – непрерывные случайные величины, то предполагается, что величина X может принимать только неотрицательные значения, а при заданном значении x величины X множество возможных значений величины Y представляет собой отрезок $[0, x]$.

Если же X и Y – дискретные случайные величины, то предполагается, что величина X может принимать только неотрицательные целочисленные значения, а при заданном значении x_i величины X величина Y может принимать значения 0, 1, 2, …, x_i (неотрицательные целочисленные значения, не превосходящие значения x_i).

Если X и Y – непрерывные случайные величины, то формализованное представление описанной взаимосвязи между величинами X и Y имеет вид:

$$A(X) \subset [0, \infty); \;\; A(Y|x) = [0, x]\,, \tag{2}$$

если X и Y – дискретные случайные величины, то

$$A(X) \subset Z_0; \; A(Y|x_i) = \{0, 1, 2, ..., x_i\}\,. \tag{3}$$

Здесь $A(X)$ – множество всех возможных значений случайной величины X; $A(Y|x)$ и $A(Y|x_i)$ – множество всех возможных значений случайной величины Y при $X = x$ и $X = x_i$ соответственно; Z_0 – множество неотрицательных целых чисел.

Требуется вычислить коэффициент корреляции случайных величин X и Y, удовлетворяющих ограничению (2) или (3) в предположении, что математическое ожидание их произведения неизвестно, но известны первые два момента распределения каждой из них.

Следует заметить, что взаимосвязь между случайными величинами, описываемая соотношениями (2) и (3), является весьма распространенной. В подтверждение этому приведем два примера из физики:

1. Случайная величина X – энергия кванта тормозного рентгеновского излучения, а случайная величина Y – энергия, переданная этим квантом поглотителю в результате взаимодействия с ним (соответствует ограничению (2)).

2. Случайная величина X – число квантов рентгеновского излучения, упавших на детектор за определенный промежуток времени, а случайная величина Y – число квантов, зарегистрированных детектором за тот же самый промежуток времени (соответствует ограничению (3)).

2. Подход к решению задачи и полученные результаты

При ограничении (2) или (3) случайная величина Y представляет собой некую случайную долю случайной величины X. Вследствие этого для решения поставленной задачи представляется целесообразным применить подход, основанный на предположении следующей формальной взаимосвязи данных величин:

$$Y = \alpha X\,, \tag{4}$$

где α– независимая от X случайная величина, распределенная на промежутке [0, 1].

Взяв математическое ожидание от обеих частей равенства (4) и учитывая независимость величин α и X, получим

$$M(Y) = M(\alpha X) = M(\alpha)M(X)\,,$$

откуда

$$M(\alpha) = \frac{M(Y)}{M(X)}\,. \tag{5}$$

Из (4) и независимости случайных величин α и X вытекает равенство

$$M(XY) = M(X\alpha X) = M(\alpha X^2) = M(\alpha)M(X^2)\,.$$

Отсюда, учитывая (5), получаем

$$M(XY) = \frac{M(Y)}{M(X)} M(X^2)\,. \tag{6}$$

При подстановке (6) в (1) будем иметь

$$r_{xy} = \frac{\frac{M(Y)}{M(X)} M(X^2) - M(X)M(Y)}{\sqrt{D(X)D(Y)}} = \frac{\frac{M(Y)}{M(X)}\left(M(X^2) - M^2(X)\right)}{\sqrt{D(X)D(Y)}} =$$

$$= \frac{\frac{M(Y)}{M(X)} D(X)}{\sqrt{D(X)D(Y)}} = \frac{M(Y)}{M(X)}\sqrt{\frac{D(X)}{D(Y)}}\,. \tag{7}$$

Таким образом, на основе предположения (4) согласно (7) нами получена следующая формула для вычисления коэффициента корреляции случайных величин X и Y, удовлетворяющих ограничению (2) или (3):

$$r_{xy} = \frac{M(Y)}{M(X)}\sqrt{\frac{D(X)}{D(Y)}}\,.$$

Между тем предположение (4) для случайных величин X и Y, удовлетворяющих ограничению (2) или (3), в действительности может и не выполняться. Тогда формулу (6) следует воспринимать как приближенную, а значит, и полученную выше формулу для вычисления коэффициента корреляции этих величин также следует воспринимать как приближенную, т.е.

$$r_{xy} \approx \frac{M(Y)}{M(X)}\sqrt{\frac{D(X)}{D(Y)}}\,. \tag{8}$$

При этом вполне закономерно возникает задача оценки погрешности формулы (8). Однако ее решение весьма затруднительно вследствие неограниченного многообразия законов распределения. В то же время можно указать условие (более общее по сравнению с (4)), при соблюдении которого формула (8) является точной, т.е. коэффициенты корреляции, вычисляемые по формуле (8) и по «истинной» формуле (1), будут одинаковыми.

Формула (8) является логическим следствием из (1) и (6), а значит, учитывая справедливость формулы (1), она будет точной, если формула (6) будет верна (будет точной).

Нетрудно убедиться, что формула (6) будет верна (будет точной), если условное математическое ожидание (УМО) $M(Y|x)$ имеет вид:

$$M(Y|x) = \theta x \tag{9}$$

для непрерывных случайных величин *X* и *Y* и, соответственно, вид:

$$M(Y|x_i) = \theta x_i \tag{10}$$

для дискретных случайных величин *X* и *Y*, где θ – постоянный множитель.

Покажем это на примере непрерывных случайных величин *X* и *Y* (для дискретных случайных величин ход рассуждений будет аналогичным). С этой целью введем обозначения: $f(x, y)$ – плотность вероятности непрерывной двумерной случайной величины (*X*, *Y*); $\varphi(x)$ – плотность вероятности случайной величины *X*; $\psi(y|x)$ – условная плотность вероятности случайной величины *Y* при *X* = *x*.

С учетом принятых обозначений формула (6) примет вид:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)dxdy = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)dxdy} \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f(x, y)dxdy\,. \tag{11}$$

Здесь и всюду в дальнейшем предполагается, что интегралы сходятся и при этом знаменатель в правой части (11) отличен от нуля.

Воспользуемся разложением

$$f(x, y) = \varphi(x)\psi(y|x)\,. \tag{12}$$

При подстановке (12) в (11) будем иметь

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\varphi(x)\psi(y|x)\,dxdy = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\varphi(x)\psi(y|x)\,dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\varphi(x)\psi(y|x)\,dxdy} \cdot \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2\varphi(x)\psi(y|x)\,dxdy\,. \tag{13}$$

Переходя в (13) от двойных интегралов к повторным интегралам получим

$$\int_{-\infty}^{\infty} x\varphi(x)\left(\int_{-\infty}^{\infty} y\psi(y|x)\,dy\right)dx = \frac{\int_{-\infty}^{\infty} \varphi(x)\left(\int_{-\infty}^{\infty} y\psi(y|x)\,dy\right)dx}{\int_{-\infty}^{\infty} x\varphi(x)\left(\int_{-\infty}^{\infty} \psi(y|x)\,dy\right)dx} \cdot \int_{-\infty}^{\infty} x^2\varphi(x)\left(\int_{-\infty}^{\infty} \psi(y|x)\,dy\right)dx\,. \tag{14}$$

Условная плотность вероятности (так же как и безусловная) обладает свойством нормировки, т.е.

$$\int_{-\infty}^{\infty} \psi(y|x)\,dy = 1\,. \tag{15}$$

С учетом (15) равенство (14) примет вид:

$$\int_{-\infty}^{\infty} x\varphi(x)\left(\int_{-\infty}^{\infty} y\psi(y|x)\,dy\right)dx = \frac{\int_{-\infty}^{\infty} \varphi(x)\left(\int_{-\infty}^{\infty} y\psi(y|x)\,dy\right)dx}{\int_{-\infty}^{\infty} x\varphi(x)dx} \cdot \int_{-\infty}^{\infty} x^2\varphi(x)dx\,. \tag{16}$$

Каждый из внутренних интегралов в (16) представляет собой УМО $M(Y|x)$. Следовательно, (16) можно переписать в виде:

$$\int_{-\infty}^{\infty} x\varphi(x)M(Y|x)\,dx = \frac{\int_{-\infty}^{\infty} \varphi(x)M(Y|x)\,dx}{\int_{-\infty}^{\infty} x\varphi(x)dx}\cdot\int_{-\infty}^{\infty} x^2\varphi(x)dx\,. \quad (17)$$

Наконец, подставляя (9) в (17), получим тождество (учитывая, что $\theta = \text{const}$)

$$\int_{-\infty}^{\infty} \theta\, x^2\varphi(x)dx = \frac{\int_{-\infty}^{\infty} \theta\, x\varphi(x)dx}{\int_{-\infty}^{\infty} x\varphi(x)dx}\cdot\int_{-\infty}^{\infty} x^2\varphi(x)dx\,,$$

что и доказывает справедливость формулы (6) при условии (9).

3. Примеры условных законов распределения, когда полученная формула является точной

Приведем сначала два примера условных законов распределения непрерывной случайной величины Y на промежутке $[0, x]$, когда равенство (9) верно, а значит, формула (8) является точной. Предварительно заметим, что условие нормировки условной плотности вероятности $\psi(y|x)$ и выражение для УМО $M(Y|x)$ при этом соответственно будут иметь следующий вид:

$$\int_0^x \psi(y|x)\,dy = 1; \quad (18)$$

$$M(Y|x) = \int_0^x y\psi(y|x)\,dy\,. \quad (19)$$

В качестве первого примера рассмотрим распределение случайной величины Y на промежутке $[0, x]$, которое обладает симметрией относительно центра данного промежутка, т.е. когда условная плотность вероятности $\psi(y|x)$ характеризуется свойством

$$\psi(y|x) = \psi(x-y|x) \text{ для } \forall y \in \; [0,x] \text{ и } \psi(y|x) = 0, \text{если } y \notin [0,x]. \quad (20)$$

Заметим, что распространенным частным случаем симметричного распределения (20) является равномерное распределение

$$\psi(y|x) = \begin{cases} \dfrac{1}{x}, \text{если } y \in [0,x], \\ 0, \text{ если } y \notin [0,x]. \end{cases}$$

С учетом (20) и (18) будем иметь для УМО (19) (произведя в одном из интегралов, приведенных ниже, замену переменной y на переменную z по формуле $y = x - z$):

$$M(Y|x) = \int_0^x y\psi(y|x)\,dy = \frac{1}{2}\left[\int_0^x y\psi(y|x)\,dy + \int_0^x y\psi(y|x)\,dy\right] =$$

$$= \frac{1}{2}\left[\int_0^x y\psi(y|x)\,dy - \int_x^0 (x-z)\psi(x-z|x)\,dz\right] = \frac{1}{2}\left[\int_0^x y\psi(y|x)\,dy + \int_0^x (x-z)\psi(x-z|x)\,dz\right] =$$

$$= \frac{1}{2}\left[\int_0^x y\psi(y|x)\,dy + \int_0^x (x-y)\psi(x-y|x)\,dy\right] = \frac{1}{2}\left[\int_0^x y\psi(y|x)\,dy + \int_0^x (x-y)\psi(y|x)\,dy\right] =$$

$$= \frac{1}{2}\int_0^x (y+x-y)\psi(y|x)\,dy = \frac{1}{2}\int_0^x x\psi(y|x)\,dy = \frac{1}{2}x\int_0^x \psi(y|x)\,dy = \frac{1}{2}x\,.$$

Таким образом, в данном случае равенство (9) верно, а коэффициент $\theta = 1/2$.

В качестве второго примера рассмотрим условное распределение случайной величины Y на промежутке $[0, x]$ по степенному закону, т.е.

$$\psi(y|x) = \begin{cases} \dfrac{a+1}{x^{a+1}} y^a, \text{ если } y \in [0, x], \\ 0, \quad \text{если } y \notin [0, x], \end{cases}$$

где параметр $a > 0$. При подстановке данной плотности вероятности в (19) получим

$$M(Y|x) = \frac{a+1}{a+2} x .$$

Таким образом, в данном случае равенство (9) также верно, и при этом коэффициент $\theta = {a+1}/{a+2}$.

Приведем теперь три примера условных законов распределения дискретной случайной величины Y, удовлетворяющей ограничению (3), когда (10) верно, а значит, формула (8) является точной. Предварительно заметим, что согласно ограничению (3) случайная величина X принимает только целочисленные неотрицательные значения, т.е. все x_i (i = 1, 2, …, n) – целые неотрицательные числа, а при $X = x_i$ случайная величина Y распределена на множестве целых чисел $\{0, 1, 2, …, x_i\}$, т.е. имеет следующие возможные значения:

$$y_j = j, \ (j = 0,1, 2,..., x_i) . \tag{21}$$

Введем обозначение вероятности того, что случайная величина Y примет значение $y_j = j$ (j = 0, 1, 2, …, x_i) при условии, что случайная величина X приняла значение x_i (i = 1, 2, …, n):

$$S(j|i) = P\{Y = y_j | X = x_i\} \tag{22}$$

Заметим при этом, что свойство нормировки условной вероятности $S(j|i)$, учитывая (21), будет иметь вид:

$$\sum_{j=0}^{x_i} S(j|i) = 1. \tag{23}$$

Рассмотрим теперь в качестве первого примера распределение случайной величины Y на множестве $\{0, 1, 2, …, x_i\}$, которое обладает симметрией относительно медианы (серединного элемента) данного множества, т.е. когда условная вероятность $S(j|i)$ характеризуется свойством

$$S(j|i) = S(x_i - j|i), \ \forall j = 0, 1, 2,..., x_i . \tag{24}$$

Отметим, что распространенным частным случаем указанного условного распределения является дискретное равномерное распределение случайной величины Y на множестве $\{0, 1, 2, …, x_i\}$:

$$S(j|i) = \frac{1}{x_i + 1}, \ \forall j = 0, 1, 2,..., x_i.$$

С учетом (21), (23) и (24) будем иметь для УМО $M(Y|x_i)$ (произведя в одной из сумм, приведенных ниже, замену переменной j на переменную k по формуле $j = x_i – k$):

$$M(Y|x_i) = \sum_{j=0}^{x_i} y_j S(j|i) = \sum_{j=0}^{x_i} jS(j|i) = \frac{1}{2}\left[\sum_{j=0}^{x_i} jS(j|i) + \sum_{j=0}^{x_i} jS(j|i)\right] =$$

$$= \frac{1}{2}\left[\sum_{j=0}^{x_i} jS(j|i) + \sum_{k=x_i}^{0} (x_i - k)S(x_i - k|i)\right] = \frac{1}{2}\left[\sum_{j=0}^{x_i} jS(j|i) + \sum_{k=0}^{x_i} (x_i - k)S(x_i - k|i)\right] =$$

$$= \frac{1}{2}\left[\sum_{j=0}^{x_i} jS(j|i) + \sum_{j=0}^{x_i} (x_i - j)S(x_i - j|i)\right] = \frac{1}{2}\left[\sum_{j=0}^{x_i} jS(j|i) + \sum_{j=0}^{x_i} (x_i - j)S(j|i)\right] = \frac{1}{2}\sum_{j=0}^{x_i} x_i S(j|i) =$$

$$= \frac{1}{2} x_i \sum_{j=0}^{x_i} S(j|i) = \frac{1}{2} x_i .$$

Таким образом, в данном случае равенство (10) верно, а коэффициент θ = 1/2.

В качестве второго рассмотрим следующий пример. Случайная величина Y при $X = x_i$ распределена по биномиальному закону с параметрами x_i (число испытаний Бернулли) и p (вероятность успеха (некоторого события) в отдельном испытании Бернулли). Тогда согласно [8]

$$M(Y|x_i) = x_i p\,,$$

а значит, в данном случае равенство (10) также верно, а коэффициент $\theta = p$.

Рассмотрим теперь третий пример. Пусть имеется некоторая генеральная совокупность объема N, в которой M ($0 < M < N$) объектов обладают заданным свойством. Предположим, что объем выборки, подлежащей извлечению из этой генеральной совокупности, есть случайная величина X, которая может принимать значения 1, 2, …, n. Тогда случайная величина Y – число объектов в выборке, обладающих заданным свойством, при $X = x_i = i$ (i = 1, 2, …, n) будет иметь гипергеометрическое распределение с параметрами x_i, M, N. Следовательно, согласно [8]

$$M(Y|x_i) = x_i \frac{M}{N}\,,$$

а значит, в данном случае равенство (10) также верно, а коэффициент $\theta = \frac{M}{N}$. Заметим, что условие $A(Y|x_i) = \{0, 1, 2, ..., x_i\}$ из ограничения (3) при этом будет выполнено, если положить $M \ge n$; $N - M \ge n$.

Применим теперь формулу (6) для приближенного вычисления коэффициента корреляции двух случайных слагаемых случайной величины.

4. Вычисление коэффициента корреляции двух случайных слагаемых случайной величины

Пусть случайная величина X разбивается случайным образом на сумму двух случайных величин Y и Z, т.е.

$$X = Y + Z. \tag{25}$$

Требуется вычислить коэффициент корреляции случайных величин Y и Z.

Из (25) следует

$$\operatorname{cov}(Y, Z) = M(XY) - M(X)M(Y) - D(Y)\,.$$

Отсюда, принимая во внимание приближенную (в общем случае, т.е. без предположения (4)) формулу (6), будем иметь

$$\operatorname{cov}(Y, Z) \approx \frac{M(Y)}{M(X)} D(X) - D(Y)\,. \tag{26}$$

Из (26) окончательно получаем искомую приближенную формулу для вычисления коэффициента корреляции двух случайных слагаемых случайной величины

$$r_{yz} \approx \frac{\frac{M(Y)}{M(X)} D(X) - D(Y)}{\sqrt{D(Y)D(Z)}}\,. \tag{27}$$

Заметим, что полученная формула (27) является точной при соблюдении условия (9) для непрерывных величин или условия (10) для дискретных величин.

5. Пример использования полученных результатов

Пусть случайная величина X распределена по закону Пуассона с параметром λ. Случайные величины Y и Z удовлетворяют уравнению (25) и при $X = n$ распределены по биномиальному закону соответственно с параметрами n, p и n, $1 - p$. Тогда (в чем несложно убедиться) безусловные законы распределения случайных величин Y и Z будут представлять собой законы распределения Пуассона с параметрами λp и $\lambda(1-p)$ соответственно. Очевидно, что пары случайных величин X и Y , а также X и Z удовлетворяют ограничению (3).

Отметим как один из возможных вариантов следующий физический смысл рассматриваемых случайных величин: случайная величина X – число квантов моноэнергетического излучения, упавших на поглотитель за фиксированный промежуток времени, случайные величины Y и Z – число квантов среди тех, которые упали на поглотитель и при этом соответственно испытали / не испытали взаимодействие с поглотителем.

Применяя формулы (8) и (27) к рассматриваемым величинам, будем иметь

$$r_{xy} \approx \sqrt{p};\, r_{xz} \approx \sqrt{1-p};\, r_{yz} \approx 0\,. \tag{28}$$

С целью проверки полученных результатов (28) было проведено компьютерное моделирование этих случайных величин с последующим расчетом выборочных коэффициентов корреляции r^* при разных сочетаниях параметров λ и p. При этом для каждой пары значений данных параметров разыгрывалось 10 000 реализаций исследуемых случайных величин. Результаты моделирования приведены в таблице. Там же для наглядности приведены теоретические коэффициенты корреляции r, вычисленные по формулам (28).

Выборочные и теоретические коэффициенты корреляции рассматриваемых случайных величин при разных сочетаниях параметров λ и *p*

λ	p	r^*_{xy}	r_{xy}	r^*_{xz}	r_{xz}	r^*_{yz}	r_{yz}
1 000	0,1	0,3230	0,3162	0,9479	0,9487	0,0048	0
1 000	0,5	0,7164	0,7071	0,6964	0,7071	–0,0018	0
1 000	0,9	0,9496	0,9487	0,3072	0,3162	–0,0065	0
100	0,1	0,3089	0,3162	0,9504	0,9487	–0,0023	0
100	0,5	0,7100	0,7071	0,7056	0,7071	0,0020	0
100	0,9	0,9487	0,9487	0,3416	0,3162	0,0270	0
10	0,1	0,3189	0,3162	0,9475	0,9487	–0,0009	0
10	0,5	0,7054	0,7071	0,7065	0,7071	–0,0032	0
10	0,9	0,9492	0,9487	0,3184	0,3162	0,0038	0

Как видно из таблицы, значения теоретических коэффициентов корреляции хорошо согласуются с соответствующими значениями выборочных коэффициентов корреляции, что делает целесообразным использование полученных в настоящей работе результатов при исследовании тесноты связи между случайными величинами, удовлетворяющими ограничению (2) или (3). Эти результаты могут быть использованы, в частности, при исследовании взаимосвязи выходных сигналов сэндвич-детекторов излучения, применяемых в рентгеновских досмотровых системах с целью обнаружения несанкционированных вложений в контролируемых объектах [9, 10].

Заключение

Получена формула (8) для вычисления коэффициента корреляции двух случайных величин при неизвестном математическом ожидании их произведения для случая, когда эти величины удовлетворяют ограничению (2) или (3) и известны первые два момента распределения каждой из них. Показано, что данная формула является точной при соблюдении условия (9) или (10). Приведены примеры законов распределения, для которых эти условия выполняются.

Получена формула (27) для вычисления коэффициента корреляции случайных слагаемых случайной величины, которая, как и формула (8), является точной при соблюдении тех же условий (9) или (10).

На примерах показано, как с использованием дополнительной информации о форме взаимосвязи исследуемых случайных величин можно существенно упростить расчет коэффициента корреляции.

Список источников

1. Базулин Е.Г. Два подхода к решению задач ультразвуковой дефектометрии: анализ высококачественного изображения отражателей и корреляционный анализ измеренных эхосигналов // Дефектоскопия. 2016. № 2. С. 11–32.
2. Жданов Ю.А., Минкин В.И. Корреляционный анализ в органической химии. Ростов н/Д : Изд-во Ростов. унта, 1966. 471 с.
3. Дунаев А.А. Адаптивный корреляционный анализ биоэлектрических сигналов // Российский медико-биологический вестник имени академика И.П. Павлова. 2003. № 1-2. С. 157–161.
4. Окатов Д.А., Минеева Т.А. Парный корреляционный анализ статистической обработки данных в медицине // Тенденции развития науки и образования. 2021. № 72-1. С. 87–91.
5. Постников В.П., Трубинова К.А. Корреляционный анализ влияния венчурного инвестирования на инновационное развитие экономики // Финансы и кредит. 2020. Т. 26, № 8 (800). С. 1767–1784.
6. Кургузов К.В., Фоменко И.К. Пространственно-корреляционный анализ инженерно-геологических данных на примере строительства логистического комплекса // Вестник МГСУ. 2019. Т. 14, № 8. С. 976–990.
7. Богомолова Е.В. Корреляционный анализ в современном педагогическом эксперименте // Информатика и прикладная математика. 2020. № 26. С. 20–25.
8. Кремер Н.Ш. Теория вероятностей и математическая статистика : учебник для вузов. М. : ЮНИТИ-ДАНА, 2001. 543 с.
9. Удод В.А., Воробейчиков С.Э., Назаренко С.Ю. Математические модели радиационных прозрачностей объекта контроля при использовании сэндвич-детекторов рентгеновского излучения // Дефектоскопия. 2020. № 2. С. 31–41.
10. Khan S.U., Khan I.U., Ullah I., Saif N., Ullah I. A review of airport dual energy X-ray baggage inspection techniques: Image enhancement and noise reduction // Journal of X-Ray Science and Technology. 2020. V. 28 (3). P. 481–505.

References

1. Bazulin, E.G. (2016) Two approaches to solving the problems of ultrasonic defectometry: an analysis of a high -quality image of reflectors and a correlation analysis of measured echosignals. *Defektoskopiya – Russian Journal of Non-Destructive Testing*. 2. pp. 11–32.
2. Zhdanov, Yu.A. & Minkin, V.I. (1966) *Korrelyatsionnyy analiz v organicheskoy khimii* [Correlation analysis in organic chemistry]. Rostov on Don: Rostov State University.
3. Dunaev, A.A. (2003) Adaptivnyy korrelyatsionnyy analiz bioelektricheskikh signalov [Adaptive correlation analysis of bioelectric signals]. *Rossiyskiy mediko-biologicheskiy vestnik imeni akademika I.P. Pavlova – I.P. Pavlov Russian Medical Biological Herald*. 1-2. pp. 157–161.
4. Okatov, D.A. & Mineeva, T.A. (2021) Parnyy korrelyatsionnyy analiz statisticheskoy obrabotki dannykh v meditsine [Pair correlation analysis of statistical data processing in medicine]. *Tendentsii razvitiya nauki i obrazovaniya*. 72-1. pp. 87–91. DOI: 10.18411/lj-04-2021-19
5. Postnikov, V.P. & Trubinova, K.A. (2020) A correlation analysis of the impact of venture investment on innovative economic development. *Finansy i kredit – Finance and Credit*. 26(8). pp. 1767–1784. DOI: 10.24891/fc.26.8.176
6. Kurguzov, K.V. & Fomenko, I.K. (2019) Spatial and correlation analysis of engineering-geological survey data for logistics center construction. *Vestnik MGSU – VESTNIK MGSU (Monthly Journal on Construction and Architecture).* 14(8). pp. 976–990. DOI: 10.22227/1997-0935.2019.8.976-990
7. Bogomolova, E.V. (2020) Korrelyatsionnyy analiz v sovremennom pedagogicheskom eksperimente [Correlation analysis in a modern pedagogical experiment]. *Informatika i prikladnaya matematika.* 26. pp. 20–25.
8. Kremer, N.Sh. (2001) *Teoriya veroyatnostey i matematicheskaya statistika* [Theory of Probability and Mathematical Statistics]. Moscow: Unity-Dana.
9. Udod, V.A., Vorobeychikov, S.E. & Nazarenko, S.Yu. (2020) Matematicheskie modeli radiatsionnykh prozrachnostey ob"ekta kontrolya pri ispol'zovanii sendvich-detektorov rentgenovskogo izlucheniya [Mathematical models of radiation transparency of the control object when using sandwich detectors of X-ray radiation]. *Defektoskopiya – Russian Journal of Non-Destructive Testing*. 2. pp. 31–41.
10. Khan, S.U., Khan, I.U., Ullah, I, Saif, N. & Ullah, I. (2020) A review of airport dual energy X-ray baggage inspection techniques: Image enhancement and noise reduction. *Journal of X-Ray Science and Technology*. 28(3). pp. 481–505. DOI: 10.3233/xst-200663

***Информация об авторе*:**
Удод Виктор Анатольевич – профессор, доктор технических наук, профессор кафедры информационных технологий и бизнес-аналитики Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: pr.udod@mail.ru

Автор заявляет об отсутствии конфликта интересов.

***Information about the author*:**
Udod Victor A. (Doctor of Technical Sciences, Professor, National Research Tomsk State University, Tomsk, Russian Federation). E-mail: pr.udod@mail.ru

The author declares no conflicts of interests.

Поступила в редакцию 30.08.2022; принята к публикации 01.03.2023

Received 30.08.2022; accepted for publication 01.03.2023