DOI: 10.26896/1028-6861-2018-84-7-76-82 УДК (UDC) 519.237.5

СКАЛЯРНАЯ МЕРА ВЗАИМОЗАВИСИМОСТИ МЕЖДУ СЛУЧАЙНЫМИ ВЕКТОРАМИ[1]

© **Александр Николаевич Тырсин**

Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, г. Екатеринбург, Россия; e-mail: at2001@yandex.ru

Статья поступила 14 февраля 2018 г.

Рассмотрена проблема оценивания тесноты взаимозависимости между случайными векторами разной размерности. Эти случайные векторы могут иметь произвольные многомерные непрерывные законы распределения. Получено аналитическое выражение для коэффициента тесноты взаимозависимости между случайными векторами. Он выражается через коэффициенты детерминации условных регрессий между компонентами случайных векторов. Для случая гауссовских случайных векторов получена более простая формула, выраженная через определители каждого из случайных векторов и определитель их объединения. Показано, что введенный коэффициент соответствует всем основным требованиям, предъявляемым к мере тесноты взаимозависимости между случайными векторами. Данный подход имеет преимущества по сравнению с методом канонических корреляций. Он позволяет определить фактическую тесноту взаимозависимости между случайными векторами. Кроме того, он может использоваться и при нелинейных корреляционных зависимостях между компонентами случайных векторов. Введенная мера достаточно просто интерпретируема и может практически применяться на реальных выборках данных. Приведены примеры расчета тесноты взаимозависимости между гауссовскими случайными векторами разной размерности.

Ключевые слова: случайный вектор; взаимозависимость; мера; дифференциальная энтропия; индекс детерминации; корреляционная матрица.

SCALAR MEASURE OF THE INTERDEPENDENCE BETWEEN RANDOM VECTORS

© **Alexander N. Tyrsin**

First President of Russia B. N. Yeltsin Ural Federal University, Yekaterinburg, Russia; e-mail: at2001@yandex.ru

Submitted February 14, 2018.

The problem of assessing tightness of the interdependence between random vectors of different dimensionality is considered. These random vectors can obey arbitrary multidimensional continuous distribution laws. An analytical expression is derived for the coefficient of tightness of the interdependence between random vectors. It is expressed in terms of the coefficients of determination of conditional regressions between the components of random vectors. For the case of Gaussian random vectors, a simpler formula is obtained, expressed through the determinants of each of the random vectors and determinant of their association. It is shown that the introduced coefficient meets all the basic requirements imposed on the degree of tightness of the interdependence between random vectors. This approach is more preferable compared to the method of canonical correlations providing determination of the actual tightness of the interdependence between random vectors. Moreover, it can also be used in case of non-linear correlation dependence between the components of random vectors. The measure thus introduced is rather simply interpretable and can be applied in practice to real data samplings. Examples of calculating the tightness of the interdependence between Gaussian random vectors of different dimensionality are given.

Keywords: random vector; interdependence; measure; differential entropy; index of determination; correlation matrix.

Природа многомерна, и одномерные методы статистического анализа часто оказываются малоэффективными при изучении сложных явлений [1]. Основная цель многомерного статистического анализа — выявление характера и структуры взаимосвязей между компонентами исследуемого многомерного признака [2 – 4]. В рамках этой цели представляет интерес проблема количественного оценивания тесноты совместной взаимосвязи (корреляционной зависимости) между многомерными случайными величинами.

В настоящее время данная проблема решена лишь частично — для случая тесноты взаимосвязи между компонентами случайного вектора. Для гауссовских случайных векторов предложена [5] скалярная мера — коэффициент тесноты совместной линейной корреляционной связи, равный

$$D_{e,\,m}(\mathbf{X}) = 1 - |\mathbf{R_X}|^{1/m}, \tag{1}$$

где $\mathbf{X} = (X_1, X_2, ..., X_m)$ — многомерная случайная величина, имеющая совместное нормальное распределение и корреляционную матрицу $\mathbf{R_X}$.

Для оценки как линейной, так и нелинейной взаимозависимости предложен [6] коэффициент

$$d_{e,\,m}(\mathbf{X}) = 1 - \exp[-2I(\mathbf{X})/m], \tag{2}$$

где $I(\mathbf{X}) = H(\widetilde{\mathbf{X}}) - H(\mathbf{X})$ — разность дифференциальных энтропий многомерных случайных величин $\mathbf{X}$ и $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_m)$, компоненты $\widetilde{X}_i$ которой являются взаимно независимыми и имеют те же распределения, что и X_i.

Основным недостатком формулы (2) является необходимость вычисления оценок дифференциальных энтропий, поскольку требуется оценивать по ограниченным выборкам одномерные и многомерные плотности распределений непрерывных случайных величин. Результат — низкая точность оценивания коэффициента $d_e(\mathbf{X})$. Это позволяет использовать соотношение (2) лишь для модельных случаев, когда законы распределений известны заранее. В целях практической реализации вместо (2) предложена [7] эквивалентная ей формула

$$d_{e,m}(\mathbf{X}) = 1 - \left[\prod_{k=2}^{m}(1 - R^2_{X_k/X_1X_2\ldots X_{k-1}})\right]^{1/m}, \tag{3}$$

где $R^2_{X_k/X_1X_2\ldots X_{k-1}}$ — индексы детерминации соответствующих регрессионных зависимостей, $k = 2, 3, ..., m$. Вопросы непараметрического оценивания индексов детерминации по многомерным выборочным данным рассмотрены в [8].

Предложенная в [5, 6] нормировка (возведение в степень $1/m$) формально не обоснована даже для совместного нормального распределения. Действительно, для гауссовского случайного вектора $\mathbf{X}$ имеем [9]

$$I(\mathbf{X}) = \sum_{i=1}^{m} H(X_i) + \frac{1}{2}\ln|\mathbf{R}_{\widetilde{\mathbf{X}}}| -$$

$$-\sum_{i=1}^{m} H(X_i) - \frac{1}{2}\ln|\mathbf{R_X}| = \frac{1}{2}\ln|\mathbf{R_X}|,$$

следовательно, величина $I(\mathbf{X})$ не зависит в явной форме от размерности m вектора $\mathbf{X}$. Поэтому наряду с (1) и (3) представляются оправданными формулы

$$D_e(\mathbf{X}) = 1 - |\mathbf{R_X}|, \tag{4}$$

$$d_e(X) = 1 - \prod_{k=2}^{m}(1 - R^2_{X_k/X_1X_2\ldots X_{k-1}}). \tag{5}$$

Кроме оценки тесноты взаимосвязи между компонентами случайного вектора, необходимо также оценивать тесноту взаимозависимости между случайными векторами. Во многих приложениях выходных переменных бывает несколько. И возникает задача оценки связи между набором измеряемых (входных) переменных $\mathbf{X}$ и интересующими нас свойствами (выходными переменными) $\mathbf{Y}$ [1, 10]. Если число выходных переменных более одной, то тесноту корреляционной связи между группами входных и выходных переменных оценивают с помощью метода канонических корреляций [11, 12]. Этот метод является обобщением парной линейной корреляции и позволяет находить максимальные корреляционные связи между двумя группами случайных величин. Зависимость оценивают с помощью канонических переменных, вычисленных как линейные комбинации исходных признаков по каждой из групп. Эти канонические величины должны максимально коррелировать между собой, а их число определяется по числу переменных в меньшем множестве (если число переменных в них не одинаково). Данный метод имеет несколько существенных недостатков. Во-первых, он рассчитан только на случай, когда все исследуемые признаки связаны друг с другом линейно, что фактически предполагает наличие совместного нормального распределения у каждого из случайных векторов. Во-вторых, находится максимальная величина коэффициента корреляции между каноническими переменными, в то время как требуется оценить тесноту фактической взаимосвязи, которая может значительно отличаться от максимально возможной. Вместе с наличием множества (их число равно размерности вектора выходных переменных) определяемых коэффициентов корреляции это весьма затрудняет интерпретацию результатов, делая ее слишком общей и малоинформативной для практических задач. В-третьих, представление канонических переменных в виде только линейных комбинаций каждой из групп переменных также существенно ограничивает результаты максимизации. В-четвертых, требуется чтобы $m \geq l$, что также является еще одним ограничением.

Цель данной работы — попытка на основе энтропийного подхода обобщить меры (1) – (5) на случай оценки тесноты взаимозависимости между двумя группами переменных, устранив недостатки и ограничения, присущие методу канонических корреляций.

Пусть заданы два произвольно распределенных непрерывных случайных вектора $\mathbf{X} = (X_1, X_2, ..., X_m)$ и $\mathbf{Y} = (Y_1, Y_2, ..., Y_l)$, причем $m > 1$ и $l > 1$. Введем для них векторы $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_m)$ и $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_l)$ такие, что все компоненты $\widetilde{X}_i$ являются взаимно независимыми и имеют те же распределения, что и X_i, а все $\widetilde{Y}_j$ — взаимно независимы и имеют те же распределения, что и Y_j. Зададим также два случайных вектора размера $m + l$: $\mathbf{Z} = \mathbf{X} \cup \mathbf{Y} = (X_1, ..., X_m, Y_1, ..., Y_l)$ и $\widetilde{\mathbf{Z}} = \widetilde{\mathbf{X}} \cup \widetilde{\mathbf{Y}} = (\widetilde{X}_1, ..., \widetilde{X}_m, \widetilde{Y}_1, ..., \widetilde{Y}_l)$.

Скалярная мера $d_e(\mathbf{X}, \mathbf{Y})$ тесноты взаимозависимости между случайными векторами $\mathbf{X}$ и $\mathbf{Y}$ должна удовлетворять следующим очевидным требованиям:

1) $0 \leq d_e(\mathbf{X}, \mathbf{Y}) \leq 1$;

2) $d_e(\mathbf{X}, \mathbf{Y}) = 0$ — соответствует независимости между $\mathbf{X}$ и $\mathbf{Y}$;

3) $d_e(\mathbf{X}, \mathbf{Y}) = 1$ — означает наличие функциональной зависимости между случайными векторами $\mathbf{X}$ и $\mathbf{Y}$, т.е. хотя бы одна из компонент вектора $\mathbf{Y}$ функционально (не случайным образом) связана с компонентами вектора $\mathbf{X}$;

4) $d_e(\mathbf{X}, \mathbf{Y})$ должна быть непрерывной.

Введем величину $I(\mathbf{X}, \mathbf{Y}) = I(\mathbf{Z}) - I(\mathbf{X}) - I(\mathrm{Y})$, где $I(\mathbf{Z}) = H(\widetilde{\mathbf{Z}}) - H(\mathbf{Z})$, $I(\mathbf{X}) = H(\widetilde{\mathbf{X}}) - H(\mathbf{X})$, $I(\mathbf{Y}) = H(\widetilde{\mathbf{Y}}) - H(\mathbf{Y})$.

В [7] получено, что

$$I(\mathbf{X}) = -\frac{1}{2}\sum_{k=2}^{m}\ln(1 - R^2_{X_k/X_1X_2\ldots X_{k-1}}),$$

$$I(\mathbf{Y}) = -\frac{1}{2}\sum_{k=2}^{l}\ln(1 - R^2_{Y_k/Y_1Y_2\ldots Y_{k-1}}).$$

Очевидно, что

$$I(\mathbf{Z}) = -\frac{1}{2}\left[\sum_{k=2}^{m}\ln(1 - R^2_{X_k/X_1\ldots X_{k-1}}) + \right.$$

$$\left. + \ln(1 - R^2_{Y_1/X_1\ldots X_m}) + \sum_{k=2}^{l}\ln(1 - R^2_{Y_k/X_1\ldots X_mY_1\ldots Y_{k-1}})\right].$$

Тогда имеем

$$I(\mathbf{X}, \mathbf{Y}) =$$

$$= -\frac{1}{2}\left[\ln(1 - R^2_{Y_1/X_1\ldots X_m}) + \sum_{k=2}^{l}\ln\frac{1 - R^2_{Y_k/X_1\ldots X_mY_1\ldots Y_{k-1}}}{1 - R^2_{Y_k/Y_1\ldots Y_{k-1}}}\right]. \quad (6)$$

Введем коэффициент $d_e(\mathbf{X}, \mathbf{Y})$ тесноты взаимозависимости между случайными векторами $\mathbf{X}$ и $\mathbf{Y}$:

$$d_e(\mathbf{X}, \mathbf{Y}) = 1 - e^{-2I(\mathbf{X}, \mathbf{Y})}. \quad (7)$$

С учетом (6) формула (7) примет вид

$$d_e(\mathbf{X}, \mathbf{Y}) = 1 - (1 - R^2_{Y_1/X_1\ldots X_m}) \times$$

$$\times \prod_{k=2}^{l}\frac{1 - R^2_{Y_k/X_1\ldots X_mY_1\ldots Y_{k-1}}}{1 - R^2_{Y_k/Y_1\ldots Y_{k-1}}}. \quad (8)$$

Формула (8) не имеет ясной интерпретации. Поэтому преобразуем ее с помощью эквивалентных преобразований. Умножив и разделив произведение справа в (8) на $\prod_{i=2}^{m}(1 - R^2_{X_i/X_1X_2\ldots X_{i-1}})$ и учитывая (5), получим

$$(1 - R^2_{Y_1/X_1X_2\ldots X_m})\prod_{k=2}^{l}\frac{1 - R^2_{Y_k/X_1X_2\ldots X_mY_1Y_2\ldots Y_{k-1}}}{1 - R^2_{Y_k/Y_1Y_2\ldots Y_{k-1}}} =$$

$$= \frac{\prod_{i=2}^{m}(1 - R^2_{X_i/X_1\ldots X_{i-1}})(1 - R^2_{Y_i/X_1\ldots X_m})}{\prod_{i=2}^{m}(1 - R^2_{X_i/X_1\ldots X_{i-1}})\prod_{k=2}^{l}(1 - R^2_{Y_k/Y_1\ldots Y_{k-1}})} \times$$

$$\times \prod_{k=2}^{l}(1 - R^2_{Y_k/X_1\ldots X_mY_1\ldots Y_{k-1}}) =$$

$$= \frac{1 - d_e(\mathbf{X} \cup \mathbf{Y})}{(1 - d_e(\mathbf{X}))(1 - d_e(\mathbf{Y}))},$$

следовательно,

$$d_e(\mathbf{X}, \mathbf{Y}) = 1 - \frac{1 - d_e(\mathbf{X} \cup \mathbf{Y})}{(1 - d_e(\mathbf{X}))(1 - d_e(\mathbf{Y}))}. \quad (9)$$

В частности, для гауссовских векторов **X** и **Y** запишем $d_e(\mathbf{X}) = D_e(\mathbf{X}) = 1 - |\mathbf{R_X}|$, $d_e(\mathbf{Y}) = D_e(\mathbf{Y}) = 1 - |\mathbf{R_Y}|$, $d_e(\mathbf{X} \cup \mathbf{Y}) = D_e(\mathbf{X} \cup \mathbf{Y}) = 1 - |\mathbf{R}_{\mathbf{X} \cup \mathbf{Y}}|$, т.е. формула (9) примет вид

$$D_e(\mathbf{X}, \mathbf{Y}) = 1 - \frac{|R_{\mathbf{X} \cup \mathbf{Y}}|}{|R_{\mathbf{X}}||R_{\mathbf{Y}}|}. \quad (10)$$

Проверим соответствие коэффициента $d_e(\mathbf{X}, \mathbf{Y})$ требованиям 1) – 4).

Теорема. Скалярная мера $d_e(\mathbf{X}, \mathbf{Y})$ тесноты взаимозависимости между случайными векторами **X** и **Y** удовлетворяет условиям 1 – 4.

Доказательство. Известно [13], что $1 \geq R^2_{Y_k/X_1X_2...X_mY_1Y_2...Y_{k-1}} \geq R^2_{Y_k/Y_1Y_2...Y_{k-1}} \geq 0$, или

$$0 \leq 1 - R^2_{Y_k/X_1X_2...X_mY_1Y_2...Y_{k-1}} \leq 1 - R^2_{Y_k/Y_1Y_2...Y_{k-1}} \leq 1. \quad (11)$$

Пусть между компонентами векторов **X** и **Y** отсутствует функциональная связь, т.е. все индексы детерминации, включая $R^2_{Y_1/X_1X_2...X_m}$, меньше 1. Это означает, что $\forall k \geq 2$,

$$0 < \frac{1 - R^2_{Y_k/X_1X_2...X_mY_1Y_2...Y_{k-1}}}{1 - R^2_{Y_k/Y_1Y_2...Y_{k-1}}} \leq 1,$$

т.е.

$$0 < (1 - R^2_{Y_1/X_1X_2...X_m}) \prod_{k=2}^{l} \frac{1 - R^2_{Y_k/X_1X_2...X_mY_1Y_2...Y_{k-1}}}{1 - R^2_{Y_k/Y_1Y_2...Y_{k-1}}} \leq 1,$$

следовательно, $0 \leq d_e(\mathbf{X}, \mathbf{Y}) < 1$.

Рассмотрим теперь наличие функциональной связи. Пусть $\exists j \geq 2$, для которого $R^2_{Y_j/Y_1Y_2...Y_{j-1}} \to 1$, т.е. $1 - R^2_{Y_j/Y_1Y_2...Y_{j-1}} \to 0$ — бесконечно малая величина. С учетом (11) получим, что $1 - R^2_{Y_j/X_1X_2...X_mY_1Y_2...Y_{j-1}}$ — бесконечно малая величина порядка не ниже, чем $1 - R^2_{Y_j/Y_1Y_2,...Y_{j-1}}$, т.е.

$$0 < \frac{1 - R^2_{Y_j/X_1X_2...X_mY_1Y_2...Y_{j-1}}}{1 - R^2_{Y_j/Y_1Y_2...Y_{j-1}}} \leq 1.$$

Если $R^2_{Y_j/X_1X_2...X_mY_1Y_2...Y_{j-1}} \to 1$, а $R^2_{Y_j/Y_1Y_2...Y_{j-1}} < 1$, то

$$\frac{1 - R^2_{Y_j/X_1X_2...X_mY_1Y_2...Y_{j-1}}}{1 - R^2_{Y_j/Y_1Y_2...Y_{j-1}}} = 0$$

и $d_e(\mathbf{X}, \mathbf{Y}) = 1$ — случай функциональной зависимости между одной или несколькими компонентами вектора **Y** и компонентами вектора X.

Если $R^2_{Y_1/X_1X_2...X_m} \to 1$, то также $d_e(\mathbf{X}, \mathbf{Y}) = 1$ и компонента Y_1 функционально зависит от случайного вектора **X**.

Из формулы (8) видим, что случай $d_e(\mathbf{X}, \mathbf{Y}) = 0$ возможен тогда и только тогда, когда $R^2_{Y_1/X_1X_2...X_m} = 0$ и $R^2_{Y_k/X_1X_2...X_mY_1Y_2...Y_{k-1}} = 0$, $k = 2, 3, ..., l$. Из $R^2_{Y_k/X_1X_2...X_mY_1Y_2...Y_{k-1}} = 0$ следует, что $R^2_{Y_k/X_1X_2,...X_m} = 0$, $k = 2,3, ..., l$. Это означает, что каждая из компонент вектора **Y** не зависит от вектора **X**, т.е. имеем случай независимости между векторами **X** и **Y**.

Мера $d_e(\mathbf{X}, \mathbf{Y})$ является функцией от переменных $R^2_{Y_1/X_1...X_m}$, $R^2_{Y_k/X_1...X_mY_1...Y_{k-1}}$, $R^2_{Y_k/Y_1...Y_{k-1}}$, $k = 2, 3, ..., l$, которые заданы на гиперкубе $\prod_{i=1}^{2l-1}[0; 1]$. Проверим непрерывность на его внутренней области $U = \prod_{i=1}^{2l-1}(0; 1)$. Очевидно, что в этом случае все знаменатели $(1 - R^2_{Y_k/Y_1...Y_{k-1}})$ в (8) будут больше нуля, поэтому функция $d_e(\mathbf{X}, \mathbf{Y})$ определена всюду на множестве U, следовательно, является непрерывной на U.

Таким образом, введенная величина $d_e(\mathbf{X}, \mathbf{Y})$ удовлетворяет всем условиям 1 – 4.

Рассмотрим некоторые свойства коэффициента $d_e(\mathbf{X}, \mathbf{Y})$.

1. Введем случайный вектор

$$\hat{\mathbf{Y}} = (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_l).$$

Пусть

$$\forall k\ R^2_{Y_k/Y_1Y_2...Y_{k-1}} = R^2_{\hat{Y}_k/\hat{Y}_1\hat{Y}_2,...\hat{Y}_{k-1}},$$

$$\forall k \neq j\ R^2_{Y_k/X_1X_2...X_m} = R^2_{\hat{Y}_k/X_1X_2...X_m}$$

и

$$R^2_{Y_j/X_1X_2...X_m} < R^2_{\hat{Y}_j/X_1X_2...X_m}.$$

Поскольку номер j компоненты не влияет на величины $d_e(\mathbf{X}, \mathbf{Y})$ и $d_e(\mathbf{X}, \hat{\mathbf{Y}})$, то зададим без потери общности $j = 1$. Тогда

$$\frac{1-d_e(\mathbf{X},\hat{\mathbf{Y}})}{1-d_e(X,Y)} = \frac{1-R^2_{\hat{Y}_1/X_1\ldots X_m}}{1-R^2_{Y_1/X_1\ldots X_m}} < 1,$$

отсюда $1 - d_e(\mathbf{X}, \hat{\mathbf{Y}}) < 1 - d_e(\mathbf{X}, \mathbf{Y})$, т.е. $d_e(\mathbf{X}, \hat{\mathbf{Y}}) >$ $> d_e(\mathbf{X}, \mathbf{Y})$.

2. Пусть имеем случайный вектор $\hat{\mathbf{Y}} = (\hat{Y}_1, \hat{Y}_2, ..., \hat{Y}_l)$, для которого $d_e(\mathbf{X} \cup \hat{\mathbf{Y}}) = d_e(\mathbf{X} \cup \mathbf{Y})$ и $d_e(\hat{\mathbf{Y}}) > d_e(\mathbf{Y})$. Тогда $d_e(\mathbf{X}, \hat{\mathbf{Y}}) < d_e(\mathbf{X}, \mathbf{Y})$.

3. Пусть имеем случайный вектор $\hat{\mathbf{Y}} = (\hat{Y}_1, \hat{Y}_2, ..., \hat{Y}_l)$, для которого $d_e(\mathbf{X} \cup \hat{\mathbf{Y}}) > d_e(\mathbf{X} \cup \mathbf{Y})$ и $d_e(\hat{\mathbf{Y}}) = d_e(\mathbf{Y})$. Тогда $d_e(\mathbf{X}, \hat{\mathbf{Y}}) > d_e(\mathbf{X}, \mathbf{Y})$.

4. Пусть имеем случайный вектор $\hat{\mathbf{X}} = (\hat{X}_1, \hat{X}_2, ..., \hat{X}_m)$, для которого $d_e(\hat{\mathbf{X}} \cup \mathbf{Y}) = d_e(\mathbf{X} \cup \mathbf{Y})$ и $d_e(\hat{\mathbf{X}}) > d_e(\mathbf{X})$. Тогда $d_e(\hat{\mathbf{X}}, \mathbf{Y}) < d_e(\mathbf{X}, \mathbf{Y})$.

Свойства 2 – 4 непосредственно следуют из (9).

Для коэффициентов $d_e(\mathbf{X}, \mathbf{Y})$ и $D_e(\mathbf{X}, \mathbf{Y})$ также можно ввести нормировку. Поскольку размерности m и l могут меняться одновременно, но с сохранением при этом $(m + l)$ постоянной, то нормировку проведем относительно средней размерности векторов, равной $(m + l)/2$. Поэтому формулам (8) – (10) поставим в соответствие выражения:

$$d_{e,m,l}(\mathbf{X},\mathbf{Y}) = 1 -$$

$$-\left[(1-R^2_{Y_1/X_1\ldots X_m})\prod_{k=2}^{l}\frac{1-R^2_{Y_k/X_1\ldots X_m Y_1\ldots Y_{k-1}}}{1-R^2_{Y_k/Y_1\ldots Y_{k-1}}}\right]^{\frac{2}{m+l}},$$

$$d_{e,m,l}(\mathbf{X},\mathbf{Y}) = 1 - \left[\frac{1-d_e(\mathbf{X}\cup\mathbf{Y})}{(1-d_e(\mathbf{X}))(1-d_e(\mathbf{Y}))}\right]^{\frac{2}{m+l}},$$

$$D_{e,m,l}(\mathbf{X},\mathbf{Y}) = 1 - \left(\frac{|\mathbf{R}_{\mathbf{X}\cup\mathbf{Y}}|}{|\mathbf{R}_{\mathbf{X}}||\mathbf{R}_{\mathbf{Y}}|}\right)^{\frac{2}{m+l}}. \quad (12)$$

В целях лучшей интерпретируемости и наглядности результатов исследуем случай гауссовских случайных векторов $\mathbf{X}$ и $\mathbf{Y}$. Рассмотрим два примера: $m = l = 2$ и $m = l = 3$.

Пример 1. Пусть $\mathbf{X} = (X_1, X_2)$, $\mathbf{Y} = (Y_1, Y_2)$, $\mathbf{X} \cup \mathbf{Y} = (X_1, X_2, Y_1, Y_2)$. Для определенности зададим

$$\mathbf{R}^{(0)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & \rho_{X_1X_2} & \rho_{X_1Y_1} & \rho_{X_1Y_2} \\ \rho_{X_1X_2} & 1 & \rho_{X_2Y_1} & \rho_{X_2Y_2} \\ \rho_{X_1Y_1} & \rho_{X_2Y_1} & 1 & \rho_{Y_1Y_2} \\ \rho_{X_1Y_2} & \rho_{X_2Y_2} & \rho_{Y_1Y_2} & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & 0{,}5 & -0{,}7 & 0{,}6 \\ 0{,}5 & 1 & -0{,}5 & 0{,}3 \\ -0{,}7 & -0{,}5 & 1 & -0{,}5 \\ 0{,}6 & 0{,}3 & -0{,}5 & 1 \end{pmatrix},$$

причем определители $|\mathbf{R}^{(0)}_{\mathbf{X}}| = 0{,}75$, $|\mathbf{R}^{(0)}_{\mathbf{Y}}| = 0{,}75$, $|\mathbf{R}^{(0)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}226$, коэффициенты детерминации (имеем совместное нормальное распределение, т.е. линейный вид регрессионной зависимости) $\rho^2_{Y_1/X_1X_2} = 0{,}52$, $\rho^2_{Y_2/Y_1} = 0{,}25$, $\rho^2_{Y_2/X_1X_2Y_1} = 0{,}373$. Согласно (10), (12) получили $D^{(0)}_e(\mathbf{X}, \mathbf{Y}) = 0{,}599$, $D^{(0)}_{e,2,2}(\mathbf{X}, \mathbf{Y}) = 0{,}367$.

Увеличим тесноту взаимосвязи между компонентами X_2 и Y_2 (за счет увеличения $\rho_{X_2Y_2}$ с 0,3 до 0,9) при сохранении неизменными остальных значений корреляционной матрицы:

$$\mathbf{R}^{(1)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & 0{,}5 & -0{,}7 & 0{,}6 \\ 0{,}5 & 1 & -0{,}5 & 0{,}9 \\ -0{,}7 & -0{,}5 & 1 & -0{,}5 \\ 0{,}6 & 0{,}9 & -0{,}5 & 1 \end{pmatrix},$$

где $|\mathbf{R}^{(1)}_{\mathbf{X}}| = 0{,}75$, $|\mathbf{R}^{(1)}_{\mathbf{Y}}| = 0{,}75$, $|\mathbf{R}^{(1)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}056$, $\rho^2_{Y_1/X_1X_2} = 0{,}52$, $\rho^2_{Y_2/Y_1} = 0{,}25$, $\rho^2_{Y_2/X_1X_2Y_1} = 0{,}843$, $D^{(1)}_e(\mathbf{X}, \mathbf{Y}) = 0{,}900$, $D^{(1)}_{e,2,2}(\mathbf{X}, \mathbf{Y}) = 0{,}683$. Получили рост тесноты взаимозависимости между векторами $\mathbf{X}$ и $\mathbf{Y}$.

Теперь для исходных векторов $\mathbf{X}$ и $\mathbf{Y}$ уменьшим тесноту корреляционной связи между компонентами $\mathbf{Y}$ (за счет изменения $\rho_{Y_1Y_2}$ с –0,5 до –0,1), сохранив остальные значения корреляционной матрицы теми же:

$$\mathbf{R}^{(2)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & 0{,}5 & -0{,}7 & 0{,}6 \\ 0{,}5 & 1 & -0{,}5 & 0{,}3 \\ -0{,}7 & -0{,}5 & 1 & -0{,}1 \\ 0{,}6 & 0{,}3 & -0{,}1 & 1 \end{pmatrix},$$

где $|\mathbf{R}^{(2)}_{\mathbf{X}}| = 0{,}75$, $|\mathbf{R}^{(2)}_{\mathbf{Y}}| = 0{,}99$, $|\mathbf{R}^{(2)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}154$, $\rho^2_{Y_1/X_1X_2} = 0{,}52$, $\rho^2_{Y_2/Y_1} = 0{,}01$, $\rho^2_{Y_2/X_1X_2Y_1} = 0{,}573$, $D^{(2)}_e(\mathbf{X}, \mathbf{Y}) = 0{,}793$, $D^{(0)}_{e,2,2}(\mathbf{X}, \mathbf{Y}) = 0{,}545$. Наблюдаем также увеличение тесноты взаимозависимости между $\mathbf{X}$ и $\mathbf{Y}$.

Уменьшим для предыдущего случая тесноту корреляционной связи между компонентами вектора $\mathbf{X}$ (за счет изменения $\rho_{X_1X_2}$ с 0,5 до 0,1), сохранив остальные значения корреляционной матрицы теми же:

$$\mathbf{R}^{(3)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & 0{,}1 & -0{,}7 & 0{,}6 \\ 0{,}1 & 1 & -0{,}5 & 0{,}3 \\ -0{,}7 & -0{,}5 & 1 & -0{,}1 \\ 0{,}6 & 0{,}3 & -0{,}1 & 1 \end{pmatrix},$$

причем $|\mathbf{R}^{(3)}_{\mathbf{X}}| = 0{,}99$, $|\mathbf{R}^{(3)}_{\mathbf{Y}}| = 0{,}99$, $|\mathbf{R}^{(3)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}008$, $\rho^2_{Y_1/X_1X_2} = 0{,}677$, $\rho^2_{Y_2/Y_1} = 0{,}01$, $\rho^2_{Y_2/X_1X_2Y_1} = 0{,}975$, $D^{(3)}_e(\mathbf{X},\mathbf{Y}) = 0{,}992$, $D^{(3)}_{e,2,2}(\mathbf{X},\mathbf{Y}) = 0{,}910$; теснота взаимозависимости между $\mathbf{X}$ и $\mathbf{Y}$ стала еще выше.

Пример 2. $\mathbf{X} = (X_1, X_2, X_3)$, $\mathbf{Y} = (Y_1, Y_2, Y_3)$, $\mathbf{X}\cup\mathbf{Y} = (X_1, X_2, X_3, Y_1, Y_2, Y_3)$. Зададим, например,

$$\mathbf{R}^{(0)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & 0{,}5 & 0{,}5 & 0{,}7 & 0{,}3 & 0{,}5 \\ 0{,}5 & 1 & 0{,}5 & 0{,}5 & 0{,}2 & -0{,}3 \\ 0{,}5 & 0{,}5 & 1 & 0{,}7 & 0{,}6 & 0{,}3 \\ 0{,}7 & 0{,}5 & 0{,}7 & 1 & 0{,}5 & 0{,}5 \\ 0{,}3 & 0{,}2 & 0{,}6 & 0{,}5 & 1 & 0{,}5 \\ 0{,}5 & -0{,}3 & 0{,}3 & 0{,}5 & 0{,}5 & 1 \end{pmatrix},$$

где $|\mathbf{R}^{(0)}_{\mathbf{X}}| = 0{,}5$, $|\mathbf{R}^{(0)}_{\mathbf{Y}}| = 0{,}5$, $|\mathbf{R}^{(0)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}0137$, $\rho^2_{Y_1/X_1X_2X_3} = 0{,}655$, $\rho^2_{Y_2/Y_1} = 0{,}25$, $\rho^2_{Y_2/X_1X_2X_3Y_1} = 0{,}396$, $\rho^2_{Y_3/Y_1Y_2} = 0{,}333$, $\rho^2_{Y_3/X_1X_2X_3Y_1Y_2} = 0{,}869$. Имеем $D^{(0)}_e(\mathbf{X},\mathbf{Y}) = 0{,}945$, $D^{(0)}_{e,3,3}(\mathbf{X},\mathbf{Y}) = 0{,}620$.

Увеличим тесноту взаимосвязи между компонентами X_3 и Y_3 (за счет изменения $\rho_{X_3Y_3}$ с 0,3 до 0,45) при сохранении неизменными остальных значений корреляционной матрицы:

$$\mathbf{R}^{(1)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & 0{,}5 & 0{,}5 & 0{,}7 & 0{,}3 & 0{,}5 \\ 0{,}5 & 1 & 0{,}5 & 0{,}5 & 0{,}2 & -0{,}3 \\ 0{,}5 & 0{,}5 & 1 & 0{,}7 & 0{,}6 & 0{,}45 \\ 0{,}7 & 0{,}5 & 0{,}7 & 1 & 0{,}5 & 0{,}5 \\ 0{,}3 & 0{,}2 & 0{,}6 & 0{,}5 & 1 & 0{,}5 \\ 0{,}5 & -0{,}3 & 0{,}45 & 0{,}5 & 0{,}5 & 1 \end{pmatrix},$$

где $|\mathbf{R}^{(1)}_{\mathbf{X}}| = 0{,}5$, $|\mathbf{R}^{(1)}_{\mathbf{Y}}| = 0{,}5$, $|\mathbf{R}^{(1)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}00852$, $\rho^2_{Y_1/X_1X_2X_3} = 0{,}655$, $\rho^2_{Y_2/Y_1} = 0{,}25$, $\rho^2_{Y_2/X_1X_2X_3Y_1} = 0{,}396$, $\rho^2_{Y_3/Y_1Y_2} = 0{,}333$, $\rho^2_{Y_3/X_1X_2X_3Y_1Y_2} = 0{,}918$, $D^{(1)}_e(\mathbf{X},\mathbf{Y}) = 0{,}966$, $D^{(1)}_{e,3,3}(\mathbf{X},\mathbf{Y}) = 0{,}676$. Получили увеличение тесноты взаимозависимости между $\mathbf{X}$ и $\mathbf{Y}$.

Теперь для исходных векторов $\mathbf{X}$ и $\mathbf{Y}$ уменьшим тесноту корреляционной связи между компонентами вектора $\mathbf{Y}$ (за счет изменения $\rho_{Y_2Y_3}$ с 0,5 до 0,1), оставив без изменения остальные значения корреляционной матрицы:

$$\mathbf{R}^{(2)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & 0{,}5 & 0{,}5 & 0{,}7 & 0{,}3 & 0{,}5 \\ 0{,}5 & 1 & 0{,}5 & 0{,}5 & 0{,}2 & -0{,}3 \\ 0{,}5 & 0{,}5 & 1 & 0{,}7 & 0{,}6 & 0{,}3 \\ 0{,}7 & 0{,}5 & 0{,}7 & 1 & 0{,}5 & 0{,}5 \\ 0{,}3 & 0{,}2 & 0{,}6 & 0{,}5 & 1 & 0{,}1 \\ 0{,}5 & -0{,}3 & 0{,}3 & 0{,}5 & 0{,}1 & 1 \end{pmatrix},$$

где $|\mathbf{R}^{(2)}_{\mathbf{X}}| = 0{,}5$, $|\mathbf{R}^{(2)}_{\mathbf{Y}}| = 0{,}54$, $|\mathbf{R}^{(2)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}0132$, $\rho^2_{Y_1/X_1X_2X_3} = 0{,}655$, $\rho^2_{Y_2/Y_1} = 0{,}25$, $\rho^2_{Y_2/X_1X_2X_3Y_1} = 0{,}396$, $\rho^2_{Y_3/Y_1Y_2} = 0{,}280$, $\rho^2_{Y_3/X_1X_2X_3Y_1Y_2} = 0{,}873$, $D^{(2)}_e(\mathbf{X},\mathbf{Y}) = 0{,}951$, $D^{(2)}_{e,3,3}(\mathbf{X},\mathbf{Y}) = 0{,}634$; также наблюдаем рост тесноты взаимозависимости между $\mathbf{X}$ и $\mathbf{Y}$.

Уменьшим для предыдущего случая тесноту корреляционной связи между компонентами вектора $\mathbf{X}$ (за счет изменения $\rho_{X_2X_3}$ с 0,5 до 0,1), сохранив остальные значения корреляционной матрицы теми же:

$$\mathbf{R}^{(3)}_{\mathbf{X}\cup\mathbf{Y}} = \begin{pmatrix} 1 & 0{,}5 & 0{,}5 & 0{,}7 & 0{,}3 & 0{,}5 \\ 0{,}5 & 1 & 0{,}1 & 0{,}5 & 0{,}2 & -0{,}3 \\ 0{,}5 & 0{,}1 & 1 & 0{,}7 & 0{,}6 & 0{,}3 \\ 0{,}7 & 0{,}5 & 0{,}7 & 1 & 0{,}5 & 0{,}5 \\ 0{,}3 & 0{,}2 & 0{,}6 & 0{,}5 & 1 & 0{,}1 \\ 0{,}5 & -0{,}3 & 0{,}3 & 0{,}5 & 0{,}1 & 1 \end{pmatrix},$$

где $|\mathbf{R}^{(3)}_{\mathbf{X}}| = 0{,}54$, $|\mathbf{R}^{(3)}_{\mathbf{Y}}| = 0{,}54$, $|\mathbf{R}^{(3)}_{\mathbf{X}\cup\mathbf{Y}}| = 0{,}00339$, $\rho^2_{Y_1/X_1X_2X_3} = 0{,}721$, $\rho^2_{Y_2/Y_1} = 0{,}25$, $\rho^2_{Y_2/X_1X_2X_3Y_1} = 0{,}392$, $\rho^2_{Y_3/Y_1Y_2} = 0{,}280$, $\rho^2_{Y_3/X_1X_2X_3Y_1Y_2} = 0{,}963$, $D^{(3)}_e(\mathbf{X},\mathbf{Y}) = 0{,}988$, $D^{(3)}_{e,3,3}(\mathbf{X},\mathbf{Y}) = 0{,}774$; рост тесноты взаимозависимости между $\mathbf{X}$ и $\mathbf{Y}$ очевиден.

Рассмотренные примеры наглядно иллюстрируют введенную меру взаимозависимости между случайными векторами.

Таким образом, введена скалярная мера взаимозависимости между непрерывными произвольно распределенными случайными векторами. Получен частный результат введенной меры для гауссовских случайных векторов. Введенная мера достаточно просто интерпретируема, позволяет однозначно оценивать тесноту взаимозависимости между случайными векторами произвольных размерностей. Простота полученных аналитических выражений позволяет практически применять предложенную меру на выборках достаточно малого объема.

ЛИТЕРАТУРА

1. **Эсбенсен К.** Анализ многомерных данных. Избранные главы / Пер. с англ. — Черноголовка: Изд-во ИПХФ РАН, 2005. — 160 с.

2. Вероятность и математическая статистика: Энциклопедия. — М.: Большая Российская энциклопедия, 1999. — 910 с.
3. **Орлов А. И.** Прикладная статистика. — М.: Экзамен, 2006. — 671 с.
4. **Тюрин Ю. Н.** Многомерная статистика: гауссовские линейные модели. — М.: Издательство Московского университета, 2011. — 136 с.
5. **Pena D., Rodriguez J.** Descriptive Measures of Multivariate Scatter and Linear Dependence / Journal of Multivariate Analysis. 2003. Vol. 85. Issue 2. P. 361 – 374. DOI: 10.1016/S0047-259X(02)00061-1.
6. **Pena D., Van der Linde A.** Dimensionless Measures of Variability and Dependence for Multivariate Continuous Distributions / Communications in Statistics: Theory and Methods. 2007. Vol. 36. Issue 10. P. 1845 – 1854. DOI: 10.1080/03610920601126449.
7. **Тырсин А. Н.** Мера совместной корреляционной зависимости многомерных случайных величин / Заводская лаборатория. Диагностика материалов. 2014. Т. 80. № 1. С. 76 – 80.
8. **Тырсин А. Н.** Энтропийное моделирование многомерных стохастических систем. — Воронеж: Научная книга, 2016. — 156 с. https://elibrary.ru/download/elibrary_25475510_92447945.pdf.
9. **Тырсин А. Н., Соколова И. С.** Энтропийно-вероятностное моделирование гауссовских стохастических систем / Математическое моделирование. 2012. Т. 24. № 1. С. 88 – 102. http://www.mathnet.ru/links/f5cf79ca6d17de8bb14a28fdad137d6a/mm3199.pdf.
10. **Биргер И. А.** Техническая диагностика. — М.: Машиностроение, 1978. — 240 с.
11. **Сошникова Л. А., Тамашевич В. Н., Уебе Г., Шефер М.** Многомерный статистический анализ в экономике. — М.: ЮНИТИ-ДАНА, 1999. — 598 с.
12. **Manly B. F. J., Navarro A. J. A.** Multivariate Statistical Methods. A Primer. 4th ed. — CRC Press, 2017. — 255 p.
13. **Елисеева И. И., Курышева С. В., Коростелева Т. В. и др.** Эконометрика. Изд. 2-е, перераб. и доп. — М.: Финансы и статистика, 2007. — 576 с.

REFERENCES

1. **Esbensen K. H.** Multivariate Data Analysis — In Practice. 5th Edition. — Oslo: CAMO Process AS, 2002.
2. Probability and mathematical statistics: Encyclopedia. — Moscow: Bol'shaya Rossiiskaya Éntsiklopediya, 1999. — 910 p. [in Russian].
3. **Orlov A. I.** Applied statistics. — Moscow: Ékzamen, 2006. — 671 p. [in Russian].
4. **Tyurin Yu. N.** Multidimensional statistics: Gaussian linear models. — Moscow: Izd. MGU, 2011. — 136 p. [in Russian].
5. **Pena D., Rodriguez J.** Descriptive Measures of Multivariate Scatter and Linear Dependence / Journal of Multivariate Analysis. 2003. Vol. 85. Issue 2. P. 361 – 374. DOI: 10.1016/S0047-259X(02)00061-1.
6. **Pena D., Van der Linde A.** Dimensionless Measures of Variability and Dependence for Multivariate Continuous Distributions / Communications in Statistics: Theory and Methods. 2007. Vol. 36. Issue 10. P. 1845 – 1854. DOI: 10.1080/03610920601126449.
7. **Tyrsin A. N.** Measure of joint correlation dependence of multidimensional random variables / Zavod. Lab. Diagn. Mater. 2014. Vol. 80. N 1. P. 76 – 80 [in Russian].
8. **Tyrsin A. N.** Entropy modeling of multidimensional stochastic systems. — Voronezh: Nauchnaya kniga, 2016. — 156 p. [in Russian].
9. **Tyrsin A. N., Sokolova I. S.** Entropy-probabilistic modeling of Gaussian stochastic systems / Matem. Modelir. 2012. Vol. 24. N 1. P. 88 – 102 [in Russian].
10. **Birger I. A.** Technical diagnostics. — Moscow: Mashinostroenie, 1978. — 240 p. [in Russian].
11. **Soshnikova L. A., Tamashevich V. N., Uebe G., Shefer M.** Multidimensional statistical analysis in Economics. — Moscow: YuNITI-DANA, 1999. — 598 p. [in Russian].
12. **Manly B. F. J., Navarro A. J. A.** Multivariate Statistical Methods. A Primer. 4th ed. — CRC Press, 2017. — 255 p.
13. **Eliseeva I. I., Kurysheva S. V., Korosteleva T. V., et al.** Econometrics. 2nd Edition. — Moscow: Finansy i statistika, 2007. — 576 p. [in Russian].