ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 25 МАТЕМАТИКИ Выпуск 1
2018

А. Н. Т ы р с и н (Екатеринбург, УрФУ). **Мера взаимозависимости между случайными векторами произвольных размерностей.**

В [1] предложен коэффициент оценки тесноты совместной корреляционной зависимости между компонентами случайного вектора $\mathbf{X} = (X_1, \ldots, X_m)$.

$$d_e(\mathbf{X}) = 1 - \prod_{k=2}^{m} \bigl(1 - R^2_{X_k/X_1X_2\ldots X_{k-1}}\bigr),$$

и нормированный вариант

$$d_{e,m}(\mathbf{X}) = 1 - \sqrt[m]{\prod\nolimits_{k=2}^{m} \bigl(1 - R^2_{X_k/X_1X_2\ldots X_{k-1}}\bigr)}$$

где $R^2_{X_k/X_1X_2\ldots X_{k-1}}$ — индексы детерминации соответствующих регрессионных зависимостей, $k = 2, 3, \ldots, m$.

Введем далее скалярную меру взаимозависимости между непрерывными случайными векторами произвольных размерностей $\mathbf{X} = (X_1, \ldots, X_m)$ и $\mathbf{Y} = (Y_1, \ldots, Y_l)$.

Зададим такие векторы $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_m)$ и $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_l)$, 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. Также зададим два случайных вектора размера $m + l$:

$$\mathbf{Z} = \mathbf{X} \bigcup \mathbf{Y} = (X_1, \ldots, X_m, Y_1, \ldots, Y_l) \quad \text{и} \quad \widetilde{\mathbf{Z}} = \widetilde{\mathbf{X}} \bigcup \widetilde{\mathbf{Y}} = (\widetilde{X}_1, \ldots, \widetilde{X}_m, \widetilde{Y}_1, \ldots, \widetilde{Y}_l).$$

Пусть $I(\mathbf{X},\mathbf{Y}) = I(\mathbf{Z}) - I(\mathbf{X}) - I(\mathbf{Y})$, где $I(\mathbf{Z}) = H(\widetilde{\mathbf{Z}}) - H(\mathbf{Z})$, $I(\mathbf{X}) = H(\widetilde{\mathbf{X}}) - H(\mathbf{X})$, $I(\mathbf{Y}) = H(\widetilde{\mathbf{Y}}) - H(\mathbf{Y})$ — разности дифференциальных энтропий случайных векторов. Известно [1], что

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

Отсюда можно получить

$$I(\mathbf{X},\mathbf{Y}) = -\frac{1}{2} \ln \bigl(1 - R^2_{Y_1/X_1X_2\ldots X_m}\bigr) - \frac{1}{2} \sum_{k=2}^{l} \ln \frac{1 - R^2_{Y_k/X_1\ldots X_mY_1\ldots Y_{k-1}}}{1 - R^2_{Y_k/Y_1\ldots Y_{k-1}}}. \qquad (1)$$

Введем коэффициент $d_e(\mathbf{X},\mathbf{Y})$ тесноты взаимозависимости между случайными векторами $\mathbf{X}$ и $\mathbf{Y}$ как $d_e(\mathbf{X},\mathbf{Y}) = 1 - e^{-2\,I(\mathbf{X},\mathbf{Y})}$. С учетом (1) имеем

$$d_e(\mathbf{X},\mathbf{Y}) = 1 - \bigl(1 - R^2_{Y_1/X_1X_2\ldots X_m}\bigr)\prod_{k=2}^{l}\frac{1 - R^2_{Y_k/X_1\ldots X_mY_1\ldots Y_{k-1}}}{1 - R^2_{Y_k/Y_1\ldots Y_{k-1}}}, \tag{2}$$

или

$$d_e(\mathbf{X},\mathbf{Y}) = 1 - \frac{1 - d_e(\mathbf{X}\bigcup\mathbf{Y})}{(1 - d_e(\mathbf{X}))\,(1 - d_e(\mathbf{Y}))}. \tag{3}$$

Для частного случая, когда векторы $\mathbf{X}$ и $\mathbf{Y}$ имеют совместные нормальные распределения вместо (3) имеем

$$D_e(\mathbf{X},\mathbf{Y}) = 1 - \frac{|\mathbf{R}_{\mathbf{X}\bigcup\mathbf{Y}}|}{|\mathbf{R}_{\mathbf{X}}|\cdot|\mathbf{R}_{\mathbf{Y}}|}, \tag{4}$$

где $|\mathbf{R}_{\mathbf{X}}|$, $|\mathbf{R}_{\mathbf{Y}}|$, $|\mathbf{R}_{\mathbf{X}\bigcup\mathbf{Y}}|$ — определители корреляционных матриц случайных векторов $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{X}\bigcup\mathbf{Y}$.

Справедлива теорема [2].

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

1. $0 \leqslant d_e(\mathbf{X},\mathbf{Y}) \leqslant 1$.
2. *Случай $d_e(\mathbf{X},\mathbf{Y}) = 0$ соответствует независимости между $\mathbf{X}$ и $\mathbf{Y}$.*
3. *Случай $d_e(\mathbf{X},\mathbf{Y}) = 1$ означает наличие функциональной зависимости между $\mathbf{X}$ и $\mathbf{Y}$, т.е. хотя бы одна из компонент вектора $\mathbf{Y}$ функционально (не случайным образом) связана с компонентами вектора $\mathbf{X}$.*
4. *Мера $d_e(\mathbf{X},\mathbf{Y})$ является непрерывной функцией.*

Следует заметить, что 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, сохраняя при этом $(m+l)$ постоянной величиной, то нормировку проведем относительно средней размерности векторов, равной $(m+l)/2$. Поэтому формулам (2)–(4) поставим в соответствие выражения:

$$\begin{aligned}
d_{e,m,l}(\mathbf{X},\mathbf{Y}) \;&=\; 1 - \Bigl((1 - R^2_{Y_1/X_1X_2\ldots X_m})\prod_{k=2}^{l}\frac{1 - R^2_{Y_k/X_1\ldots X_mY_1\ldots Y_{k-1}}}{1 - R^2_{Y_k/Y_1\ldots Y_{k-1}}}\Bigr)^{2/(m+l)},\\
d_e(\mathbf{X},\mathbf{Y}) \;&=\; 1 - \left(\frac{1 - d_e(\mathbf{X}\bigcup\mathbf{Y})}{(1 - d_e(\mathbf{X}))(1 - d_e(\mathbf{Y}))}\right)^{2/(m+l)},\\
D_{e,m,l}(\mathbf{X},\mathbf{Y}) \;&=\; 1 - \left(\frac{|\mathbf{R}_{\mathbf{X}\bigcup\mathbf{Y}}|}{|\mathbf{R}_{\mathbf{X}}|\cdot|\mathbf{R}_{\mathbf{Y}}|}\right)^{2/(m+l)}.
\end{aligned}$$

Выводы. 1. Введена скалярная мера взаимозависимости между непрерывными произвольно распределенными случайными векторами.

2. Получен частный результат введенной меры для гауссовских случайных векторов.

3. Простота полученных аналитических выражений позволяет практически применять предложенную меру на выборках достаточно малого объема.

Работа выполнена при финансовой поддержке гранта РФФИ № 17-01-00315а.

СПИСОК ЛИТЕРАТУРЫ

1. *Тырсин А. Н.* Мера совместной корреляционной зависимости многомерных случайных величин. — Заводская лаборатория. Диагностика материалов, 2014, т. 80, № 1, с. 76–80.
2. *Тырсин А. Н.* Скалярная мера взаимозависимости между случайными векторами. — Заводская лаборатория. Диагностика материалов, 2018, т. 84. В печати.