ՀԱՅԿԱԿԱՆ ՍՍՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԿԱԴԵՄԻԱՅԻ ԶԵԿՈՒՅՑՆԵՐ
ДОКЛАДЫ АКАДЕМИИ НАУК АРМЯНСКОЙ ССР

LXXVI 1983 3

УДК 53.50/08

ФИЗИКА

В. И. Луценко, И. В. Луценко, В. М. Тер-Антонян

Критерий относительной корреляции признаков в многомерных статистических системах

(Представлено чл.-корр. АН Армянской ССР Г. С. Саакяном 11/VI 1982)

В [1] нами было показано, что для определения меры неоднородности одномерных статистических систем необходимо наряду с плотностью распределения $\gamma(\beta)$ элементов системы по концентрации признака β учитывать также инклюзивную функцию извлечения $\varepsilon(\beta)$, описывающую распределение признака по собственной концентрации. Покажем, что распространение такого подхода на многомерные системы позволяет внести ясность в вопрос о зависимости между признаками и единым образом описать как неоднородность этих систем, так и скоррелированность признаков в ней.

Для количественного определения меры скоррелированности признаков в статистической системе обычно пытаются тем или иным способом выявить отход эксклюзивной плотности распределения $\gamma(\beta_1, \beta_2)$ от произведения инклюзивных плотностей $\gamma^{(1)}(\beta_1)$ и $\gamma^{(2)}(\beta_2)$ [2]. В результате получаются критерии, характеризующие меру „абсолютной" взаимосвязи между признаками. При таком подходе остается открытым вопрос об относительной зависимости одного признака от другого. Чтобы убедиться в состоятельности такого вопроса, рассмотрим следующий пример. Пусть в элементах системы с фиксированным значением концентрации признака β_1 встречается только одно определенное значение концентрации β_2, а элементы с фиксированным значением β_2 содержат разные значения концентрации β_1 первого признака. Очевидно, в такой системе второй признак максимально скоррелирован с первым, тогда как зависимость первого признака от второго слабее. Итак, признаки в системе могут зависеть друг от друга в разной степени.

Исследуем этот вопрос подробнее. Введем эксклюзивные по двум исследуемым признакам функции извлечения $\varepsilon^{(1)}(\beta_1, \beta_2)$ и $\varepsilon^{(2)}(\beta_1, \beta_2)$, представляющие собой плотности распределения первого и второго признака по концентрациям β_1 и β_2. Легко показать, что

$$\varepsilon^{(i)}(\beta_1, \beta_2) = \frac{\beta_i}{\bar{\beta}_i} \gamma(\beta_1, \beta_2), \tag{1}$$

где $i = 1, 2$, $\bar{\beta}_1$ и $\bar{\beta}_2$—средние концентрации, а $\gamma(\beta_1, \beta_2)$—эксклюзивная плотность распределения элементов системы по концентрациям β_1 и β_2. Интегрируя обе части соотношения (1) по β_1 и β_2, приходим к инклюзивной матрице извлечения $\varepsilon^{(i)}(\beta_k)$, в которой диагональные элементы совпадают с введенными ранее в работе ([1]) функциями извлечения

$$\varepsilon^{(i)}(\beta_i)=\frac{\beta_i}{\bar{\beta}_i}\gamma^{(i)}(\beta_i),$$

а недиагональные элементы $(i\neq k)$

$$\varepsilon^{(i)}(\beta_k)=\int_0^1\frac{\beta_i}{\bar{\beta}_i}\gamma(\beta_i,\ \beta_k)d\beta_i$$

указывают, какая часть i-го признака сосредоточена в элементах системы с фиксированным значением k-го признака, т. е. имеют смысл функций сопутствующих извлечений.

Выразим функции сопутствующего извлечения через соответствующие им инклюзивные плотности $\gamma^{(k)}(\beta_k)$:

$$\varepsilon^{(i)}(\beta_k)=\frac{\bar{\beta}_i(\beta_k)}{\bar{\beta}_i}\gamma^{(k)}(\beta_k).$$

В последней формуле

$$\bar{\beta}_i(\beta_k)=\int_0^1\beta_i\gamma(\beta_i/\beta_k)d\beta_i$$

и означает среднюю концентрацию i-го признака в элементах системы с фиксированной концентрацией k-го, а $\gamma(\beta_i/\beta_k)$—соответствующая плотность условной вероятности.

Из смысла функции $\bar{\beta}_i(\beta_k)$ ясно, что i-ый признак независим от k-го, если

$$\bar{\beta}_i(\beta_k)=\bar{\beta}_i, \tag{2}$$

или на языке γ- и ε-распределений $(i\neq k)$

$$\varepsilon^{(i)}(\beta_k)=\gamma^{(k)}(\beta_k). \tag{3}$$

Условия (2) и (3) автоматически выполняются, если исследуемые признаки независимы в обычном смысле, т. е. если

$$\gamma(\beta_1,\ \beta_2)=\gamma^{(1)}(\beta_1)\gamma^{(2)}(\beta_2). \tag{4}$$

Рассмотрим таблицу, для которой верхнее число в каждой клетке означает концентрацию первого признака, нижнее — концентрацию второго. Элементами такой системы являются клетки. Простые вычисления показывают, что в приведенном примере условия (2) и (3) соблюдаются, а условие (4) нарушается. Итак, при общей скоррелированности признаков, т. е. нарушении условия (4), может оказаться, что i-ый признак не зависит от k-го и вместе с этим k-ый признак не зависит от i-го. Из сказанного следует, что условия (2) и (3) являются условиями относительной независимости двух признаков в многомерной статистической системе. При нарушении этих условий за меру относительной зависимости признаков естественно принять следующий критерий ($i \neq k$):

$$L_k^i = \int_{\Delta k} [\gamma^{(k)}(\beta_k) - \varepsilon^{(i)}(\beta_k)] d\beta_k. \qquad (5)$$

Здесь интегрирование ведется по области Δ_k, в которой подынтегральное выражение положительно. Если формально принять $i=k$, то (5) переходит во введенный в работе [1] критерий неоднородности

$$L^i = \int_0^{\bar{\beta}_i} [\gamma^{(i)}(\beta_i) - \varepsilon^{(i)}(\beta_i)] d\beta_i.$$

0,1 0,01	0,2 0,01	0,1 0,02	0,2 0,04	0,3 0,02	0,1 0,01	0,4 0,01	0,1 0,03
0,2 0,01	0,4 0,03	0,1 0,01	0,4 0,01	0,2 0,01	0,4 0,02	0,2 0,01	0,1 0,01
0,2 0,04	0,1 0,01	0,3 0,01	0,2 0,01	0,1 0,02	0,2 0,02	0,2 0,01	0,1 0,04
0,1 0,03	0,3 0,01	0,2 0,01	0,3 0,04	0,1 0,01	0,4 0,02	0,2 0,01	0,1 0,02
0,1 0,01	0,3 0,01	0,1 0,02	0,1 0,02	0,2 0,01	0,2 0,03	0,2 0,04	0,2 0,01

Легко показать, что при $i \neq k$

$$0 \leqslant L_k^i \leqslant L^i,$$

причем нижний предел соответствует случаям, когда выполняется хотя бы одно из условий (3), а верхний предел достигается в системах, i-ый признак которых является однозначной функцией k-го.

В заключение отметим следующее:

1. Условие абсолютной независимости признаков (4) весьма чувствительно к ограничениям, не влияющим на „динамическую" структуру исследуемой системы. Так, любое ограничение кинематического характера (например, сохранение энергии-импульса в процессах множественного рождения) приводит к нарушению условия (4), и если иметь в виду только это нарушение, то легко сделать ошибочный вывод о мере зависимости между признаками.

2. Выяснение меры относительной скоррелированности признаков может оказаться важным и в чисто прикладных задачах, например, в задаче об извлечении признака из системы. Так, если технически легче извлечь первый признак, то, имея критерий (5), легко

решить вопрос о целесообразности извлечения второго признака за счет извлечения первого.

Мы признательны Г. С. Саакяну и А. Г. Худавердяну за интересные обсуждения.

Ереванский государственный университет

Ի. Վ. ԼՈՒՑԵՆԿՈ, Վ. Ի. ԼՈՒՑԵՆԿՈ, Վ. Մ. ՏԵՐ-ԱՆՏՈՆՅԱՆ

Բազմաչափ վիճակագրական սիստեմների հատկանիշների միջև հարաբերական կոռելյացիայի հայտանիշը

Վիճակագրական սիստեմի հատկանիշների փոխկոռելյացման չափը որակական որոշման համար սովորաբար փորձում են այս կամ այն կերպ զանազանել էքսկլյուզիվ խորությունները ինկլյուզիվ խորությունների արտադրյալից։ Արդյունքում ստացվում է հատկանիշների միջև գործող «բացարձակ» փոխադարձ կապի «բացարձակ» հայտանիշ, և բաց է մնում մի հատկանիշի հարաբերական կապի հարցը մյուս հատկանիշի հետ, մի հարց, որն անկասկած հետաքրքիր է, քանի որ սիստեմում հատկանիշները կարող են տարբեր աստիճանի կախվածություն ունենալ միմյանցից։ Այս հոդվածի նպատակն է գտնել այն հայտանիշը, որը քանակապես կորոշի տարբեր հատկանիշների հարաբերական փոխկոռելյացման աստիճանը։ Ստացված են (2) և (3) հարաբերական անկախության պայմանները և պարզվել է, որ այդ պայմաններին բավարարելու դեպքում հատկանիշները սիստեմում հանդիսանում են դինամիկորեն անկախ։

Առաջարկված է L_k^i *հայտանիշը, որը միաժամանակ նկարագրում է վիճակագրական սիստեմի անհամասեռությունը և նրանում հատկանիշների փոխկոռելյացիան։ Պարզաբանվել են հարաբերական կախվածության հայտանիշի փոփոխության սահմանները և բերված է մի սիստեմի օրինակ, որում հատկանիշները հարաբերականորեն անկախ են նրանց ընդհանուր նկատելի փոխկոռելյացման դեպքում։*

ЛИТЕРАТУРА — ԳՐԱԿԱՆՈՒԹՅՈՒՆ

[1] *В. И. Луценко* и др., ДАН АрмССР, т. 74, № 5 (1982). [2] *Г. Крамер,* Математические методы статистики, Мир, М., 1975.