ՀԱՅԿԱԿԱՆ ՍՍՀ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԿԱԴԵՄԻԱՅԻ ԶԵԿՈՒՅՑՆԵՐ
ДОКЛАДЫ АКАДЕМИИ НАУК АРМЯНСКОЙ ССР
LXXIV 1982 5

УДК 53. 05/08

ФИЗИКА

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

Критерий неоднородности одномерных статистических систем

(Представлено чл.-корр. АН Армянской ССР Г. С. Саакяном 2/XII 1981)

Реальные статистические системы обычно наделены несколькими признаками, что усложняет их исследование. В связи с этим при работе с такими системами в качестве первого шага сосредотачивают внимание на одном из признаков, а по остальным ведут суммирование. Примерами систем, которые при таком инклюзивном подходе [1] приводятся к одномерным, могут служить смесь полезных компонентов с пустой породой, промышленные предприятия данной отрасли с распределенными между ними средствами, участки звездного неба с содержащимися в них светящимися объектами, пленки с фиксированными на них треками элементарных частиц и многое другое.

При экспериментальной обработке данных, снимаемых с одномерных статистических систем, в первую очередь устанавливается спектр возможных значений исследуемого признака x. В ряде случаев вместо самого признака используется его концентрация $\beta = x/\lambda$, где λ—максимальное количество признака, которое может приходиться на один элемент системы. Определенное таким образом β заключено в интервале $0 \leqslant \beta \leqslant 1$. Следующий шаг в обработке данных — это группировка элементов системы в подсистемы с данным β и построение плотности распределения $\gamma(\beta)$ элементов системы по концентрации признака. Каждая подсистема с данным β несет на себе определенное количество признака, и поэтому признак также оказывается распределенным с некоторой плотностью $\varepsilon(\beta)$ по собственной концентрации.

В настоящей статье рассматриваются следствия, к которым приводит сравнение этих двух плотностей распределений.

Из самого определения γ- и ε-распределений следует, что между ними имеется связь

$$\varepsilon(\beta) = \frac{\beta}{\bar{\beta}} \gamma(\beta), \tag{1}$$

где $\bar{\beta}$—средняя концентрация. Если $\gamma = \varepsilon$, то из уравнения связи имеем

$$\gamma(\beta) = \delta(\beta - \bar{\beta}), \tag{2}$$

где справа стоит дельта-функция Дирака. В системах с γ-распределением (2) все элементы с точки зрения признака являются тождественными, и потому такие системы целесообразно называть однородными. В неоднородных системах $(\gamma \neq \varepsilon)$ спектр концентраций признака обязательно содержит более одной точки. Будем считать систему тем неоднороднее, чем больше различие между γ- и ε-распределениями. При таком подходе в качестве меры неоднородности естественно принять критерий

$$L = \int_0^{\bar\beta} \left[1 - \frac{\beta}{\bar\beta} \right] \gamma(\beta) d\beta, \tag{3}$$

интегральным образом учитывающий различие между функциями γ и ε. Легко показать, что

$$0 \leq L \leq 1 - \bar\beta, \tag{4}$$

причем нижний предел достигается в однородных системах. Из неравенства (4) следует существование крайне неоднородных систем, для которых $L = 1 - \bar\beta$. Выясним, какими γ-распределениями они описываются. Формула (3) говорит о том, что чем сильнее функция $\gamma(\beta)$ сосредоточена вблизи точек $\beta = 0$ и $\beta = 1$, тем больше L. Поэтому естественно ожидать, что крайне неоднородным системам должны соответствовать γ-распределения, имеющие вид

$$\gamma(\beta) = 2(1 - \bar\beta)\delta(\beta) + 2\bar\beta\delta(\beta - 1). \tag{5}$$

Здесь учтено условие нормировки γ-распределения и определение $\bar\beta$. Из формул (3) и (5) следует, что для этих γ-распределений $L = 1 - \bar\beta$ и потому такие плотности распределений действительно описывают крайне неоднородные системы.

Неравенство (4) выделяет в плоскости $(L, \bar\beta)$ треугольник, в котором катету $(0, \bar\beta)$ соответствуют однородные, гипотенузе $L = 1 - \bar\beta$ — крайне неоднородные системы, а внутренним точкам — системы с промежуточной степенью неоднородности. Для описания того, насколько при данном $\bar\beta$ система далека от состояния крайней неоднородности, введем стадию неоднородности

$$E = \frac{L}{1 - \bar\beta}.$$

Системы, находящиеся на равной стадии неоднородности, группируются на отрезке, соединяющем соответствующую точку катета $(L, 0)$ с вершиной треугольника $(0, 1)$.

Описанная картина позволяет ввести понятие о состоянии неоднородности одномерной статистической системы. Будем считать, что такое состояние задано, если фиксированы L и $\bar\beta$. Состояние неоднородности вообще говоря может включать в себя множество „микросостояний", каждое из которых полностью определено своим γ-распределением. В результате каждой точке треугольника неоднородности соответствует ансамбль микросостояний. Мерой близости двух состояний неоднородности служит расстояние между точками треугольника, изображающими эти состояния, т. е.

$$r_{ij}=[(\overline{\beta}_i-\overline{\beta}_j)^2+(L_i-L_j)^2]^{1/2}. \tag{6}$$

Совокупность состояний неоднородности с данными $\overline{\beta}$, L или E образуют кластеры (2). Мера близости $\overline{\beta}$- и L-кластеров определяется формулой (6), если в нее формально подставить $L_i=L_j$ или $\overline{\beta}_i=\overline{\beta}_j$ соответственно, а расстояние между E-кластерами равно $|E_i-E_j|$.

В математической статистике изучение одномерных систем сводится к эксплуатации γ-распределения без какого-либо упоминания об ε-распределении (3). В то же время эта характеристика прочно вошла в обиход инженеров-обогатителей, и поэтому мы сохраняем за ней название „функция извлечения". Легко убедиться, что многие формулы из математической статистики могут быть переписаны в терминах разности между γ- и ε-распределениями. Например, дисперсия γ-распределения равна

$$D=\overline{\beta}\int_0^1\beta(\varepsilon-\gamma)d\beta.$$

Аналогичные формулы можно получить и для центральных моментов более высокого порядка. Это значит, что такие характеристики, как дисперсия D, коэффициент вариации v и другие являются „продуктами" неоднородности и описывают лишь частичные проявления последней (разброс, нормированный разброс и т. д.). Поэтому не удивительно, что в некоторых простых случаях по этим характеристикам можно весьма точно судить о мере неоднородности одномерных статистических систем. Так, для γ-распределения

$$\gamma(\beta)=\frac{\Theta(\beta-a)-\Theta(\beta-b)}{b-a},$$

где $b>a$ и Θ—функция Хевисайда, L-критерий выражается через коэффициент вариации следующим образом

$$L=\frac{\sqrt{3}}{4}v^2.$$

Неоднородные системы обязательно наделены некоторым „беспорядком", и в этом смысле „продуктом" неоднородности является также информационная энтропия S. Поэтому выясним, в какой степени можно по энтропии судить о неоднородности системы. Рассмотрим множество систем со средней концентрацией $\overline{\beta}=1/2$. Среди этих систем максимумом S обладает та, элементы которой распределены по концентрации β равномерно, т. е. $\gamma(\beta)=1$. Для таких систем $L=1/4$ и $E=1/2$, что означает, что система с максимумом S может быть далека от крайне неоднородного состояния (5) и, следовательно, информационная энтропия может заведомо ошибочно оценивать степень неоднородности систем со значительной стадией неоднородности.

Нам приятно поблагодарить Г. С. Саакяна, А. Н. Сисакяна и А. Г. Худавердяна за интерес к работе и полезные обсуждения.

Ереванский государственный университет

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

Միաչափ վիճակագրական համակարգերի անհամասեռության չափանիշը

Աշխատանքում ուսումնասիրված են վիճակագրական սիստեմներ, որոնց տարրերը կամայական ձևով բաշխված են ըստ իրենց բնութագրող հայտանիշի կոնցենտրացիայի։ Տրված են համասեռ, անհամասեռ և ծայրահեղ անհամասեռ միաչափ վիճակագրական սիստեմների (Մ. Վ. Ս.) հասկացությունների սահմանումներ։ Ստացված են ՄՎՍ-ի անհամասեռության չափը և աստիճանը որոշող L և E հայտանիշներ։

Պարզաբանվել է L—հայտանիշի դերը մաթեմատիկական վիճակագրության ընդունված բնութագրիչների նկատմամբ։ Կատարված է ՄՎՍ-ի դասակարգում ըստ կլաստերների և պարզաբանվել է կապը անհամասեռության և էնտրոպիայի միջև։

ЛИТЕРАТУРА—ԳՐԱԿԱՆՈՒԹՅՈՒՆ

[1] *R. P. Feynman*. Phys. Rev. Lett., vol. 23, 1415 (1969). [2] *Б. Дюран, Д. Оделл*. Кластерный анализ, Мир, М., 1977. [3] *Г. Крамер*, Математические методы статистики, Мир, М., 1975.