Доклады Академии Наук СССР
1940. Том XXVII, № 5

МАТЕМАТИКА

В. И. РОМАНОВСКИЙ

ОБ ИНДУКТИВНЫХ ВЫВОДАХ В СТАТИСТИКЕ

(Представлено академиком А. Н. Колмогоровым 21 III 1940)

К числу основных проблем статистики относится вопрос о разыскании характеристик генеральных совокупностей по соответственным характеристикам случайных выборок из них. Разрешение этого вопроса чаще всего основывается на теореме Байеса или на теоремах, подобных ей, связанных с введением в рассуждения априорных вероятностей, от элементов произвола и необоснованности в которых невозможно освободиться. Но можно тот же вопрос разрешить иначе, совершенно не прибегая к априорным вероятностям.

На эту возможность указал впервые Р. Фишер. Он ввел понятие о доверительной (фидуциальной) вероятности и изложил свои идеи в применении к выборкам из нормальных совокупностей в ряде работ, начиная с 1930 г. ([1–4]). Эти идеи получили дальнейшее развитие у ряда других, главным образом английских, ученых: у Ю. Неймана, Э. Пирсона, Бартлета, Вилкса и др. ([5–9]).

В настоящей заметке идеи Р. Фишера применяются к выборкам из произвольных совокупностей.

Будем для простоты рассматривать генеральную совокупность S с количественным аргументом x; пусть x имеет в S конечное число значений

$$x_1, x_2, \ldots, x_s \tag{1}$$

с долями

$$p_1, p_2, \ldots, p_s. \tag{2}$$

S может быть конечной или бесконечной совокупностью. Предположим, что из S берется случайная выборка S' по схеме возвращенного шара; пусть в ней значения (1) имеют частоты

$$n_1, n_2, \ldots, n_s$$

и пусть объем ее равен n.

Теперь можно доказать следующую, очень простую и очень важную теорему [([10]), стр. 108].

Теорема 1. *Каковы бы ни были* $n, p_1, p_2, \ldots, p_s$ *и* $\varepsilon > 0$,

$$P\left[\sum\left(\frac{n_i}{n} - p_i\right)^2 < \varepsilon^2\right] > 1 - \frac{1-\frac{1}{s}}{n\varepsilon^2}. \tag{3}$$

Так как правая часть неравенства (3) не зависит от p_i, то из этой теоремы тотчас следует, что если доли p_i значений (1) совокупности S нам неизвестны, то мы можем их заменить относительными частостями $\frac{n_i}{n}$, наблюденными в выборке S', причем с вероятностью, как угодно близкой к достоверности, мы можем ожидать, что совершенные при этой замене ошибки будут сколь угодно малы, если n будет достаточно большим числом.

Мы видим, что это заключение совершенно свободно от априорных вероятностей *.

Теперь мы покажем, какую роль играет полученный нами вывод в индуктивных суждениях статистики.

Пусть θ представляет некоторую характеристику генеральной совокупности S, зависящую от значений (1) и неизвестных нам их долей (2); пусть затем θ' представляет соответствующую характеристику выборки S', составленную из значений (1) и их частостей $\frac{n_i}{n}$ так же, как θ составлена из (1) и (2). Тогда нетрудно видеть справедливость следующей теоремы, имеющей обширное применение в статистике и свободной от априорных вероятностей.

Теорема 2. *Если функция* $\theta(z_1, z_2, \ldots, z_s)$ *непрерывна для значений* $z_1, z_2, \ldots, z_s$, *удовлетворяющих условиям*

$$z_i \geqslant 0,\ i = 1, 2, \ldots, s;\ \sum z_i \leqslant 1,$$

то вероятность неравенства

$$\left|\theta\left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_s}{n}\right) - \theta(p_1, p_2, \ldots, p_s)\right| > \varepsilon$$

стремится к нулю при $n \to \infty$ *равномерно относительно*

$$p_1, p_2, \ldots, p_s,$$

каково бы ни было заданное $\varepsilon > 0$.

В заключение заметим, что очень просто можно доказать следующую теорему, подобную теореме 1.

Теорема 3. *Если из генеральной совокупности* S *произведены две случайные выборки* S' *и* S'' *по схеме возвращенного шара, объемов* n' *и* n'' *и с частотами* n_i' *и* n_i'' *для значений* (1), *то для любого* $\varepsilon > 0$ *справедливо неравенство*

$$P\left[\sum\left(\frac{n_i'}{n'} - \frac{n_i''}{n''}\right)^2 < \varepsilon^2\right] > 1 - \frac{1 - \frac{1}{s}}{\varepsilon^2}\left(\frac{1}{n'} + \frac{1}{n''}\right). \qquad (4)$$

Рассматривая выборку S', как уже произведенную, а выборку S'', как предстоящую и неизвестную, мы выводим из написанного неравенства, что с вероятностью, как угодно близкой к достоверности, все частости $\frac{n_i''}{n''}$ будут отличаться от соответственных частостей $\frac{n_i'}{n'}$ как угодно мало, если числа n' и n'' будут достаточно велики. И этот вывод свободен от априорных вероятностей.

Все изложенное можно без труда распространить на случайные выборки, произведенные по схеме возвращенного шара из генеральной

* Следует заметить, что в доказательстве неравенства (3) они также не принимают никакого участия.

совокупности с любым конечным числом аргументов, как угодно распределенных в ней *.

Поступило
21 III 1940

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] R. A. Fisher, Proc. Camb. Phil. Soc., 26, 528—535 (1930). [2] R. A. Fisher, Proc. Roy. Soc., A, 139, 343—348 (1933). [3] R. A. Fisher, Proc. Roy. Soc., A, 144, 285—307 (1934). [4] R. A. Fisher, Ann. of Eugenics, VI, 391—398 (1935). [5] C. J. Clopper a. E. S. Pearson, Biometrica, 26, 404—413 (1934). [6] G. Neyman, Phil. Trans., A, 236, 333—380 (1937). [7] M. S. Bartlett, Proc. Roy. Soc., A, 160, 268—282 (1937). [8] M. S. Bartlett, Ann. of Math. Stat., X, 129—138 (1939). [9] S. S. Wilks, Ann. of Math. Stat., IX, 272—280 (1938). [10] В. Романовский, Математическая статистика (1938).

---

* Некоторые формулировки сообщения В. И. Романовского могут дать повод к недоразумениям. Поэтому я считаю необходимым отметить здесь следующее:

В теореме 1 вероятность

$$P\left[\sum\left(\frac{n_i}{n}-p_i\right)^2<\varepsilon^2\right]$$

берется при *заданных* n, $p_1, p_2, \ldots, p_s$ и *случайных* $n_1, n_2, \ldots, n_s$. Теорема утверждает, что эта вероятность удовлетворяет неравенству (3), каковы бы ни были $n, p_1, p_2, \ldots, p_s$. В этой форме теорема, безусловно, правильна. Было бы, однако, ошибочно понимать дело так, что отсюда получается какая-либо оценка вероятности неравенства

$$\sum\left(\frac{n_i}{n}-p_i\right)^2<\varepsilon^2$$

при заданных $n_1, n_2, \ldots, n_s$ и случайных $p_1, p_2, \ldots, p_s$. Получить такую оценку без введения закона распределения для самих $p_1, p_2, \ldots, p_s$ нельзя.

Точно так же в теореме 3 вероятность

$$P\left[\sum\left(\frac{n_i'}{n'}-\frac{n_i''}{n''}\right)^2<\varepsilon^2\right]$$

берется при заданных $n', n'', p_1, p_2, \ldots, p_s$ и случайных $n_1', n_2', \ldots, n_s', n_1'', n_2'', \ldots, n_s''$. Теорема утверждает, что эта вероятность удовлетворяет неравенству (4), каковы бы ни были $n', n'', p_1, p_2, \ldots, p_s$. Никаких заключений о вероятности неравенства

$$\sum\left(\frac{n_i'}{n'}-\frac{n_i''}{n''}\right)^2<\varepsilon^2$$

при заданных $n_1', n_2', \ldots, n_s'$ и случайных $n_1'', n_2'', \ldots, n_s''$ вывести из теоремы 3 нельзя. В явной форме такого рода ошибочные заключения можно найти в цитированной В. И. Романовским работе R. A. Fisher'a [Annals of Eugenics, 6, стр. 391—392 (1935)]. Тем не менее я считаю, что теоремы типа теорем 1—3 и аналогичных теорем R. A. Fisher'a и его последователей при их правильном понимании могут служить достаточным обоснованием практических приемов математической статистики. Подробнее свою точку зрения по этому важному вопросу я изложу в особой статье.

Академик **А. Н. Колмогоров**