COLLOQUIUM MATHEMATICUM

VOL. VIII 1961 FASC. 1

CONCERNING RELATIVE ACCURACY OF STRATIFIED AND SYSTEMATIC SAMPLING IN A PLANE

BY

J. HÁJEK (PRAHA)

The conjecture P 254 formulated in [2] is disproved by the following theorem. We use the notation of [2].

THEOREM. *Let us have a plane stationary stochastic process* $y(p)$, $p \in E_2$, *with correlation function* $f(p, q)$ *given by*

$$f(p, q) = R[y(p), y(q)] = e^{-a|p-q|} \qquad (a > 0) \tag{1}$$

where $R[y(p), y(q)]$ *denotes the coefficient of correlation between* $y(p)$ *and* $y(q)$ *and* $|p-q|$ *denotes the distance of the points* p *and* q, *and consider two disjoint domains* D *and* D' *congruent by translation. Then for a sufficiently small positive number* a *the systematic sampling of two points from* $D \cup D'$ *is less efficient than the stratified sampling, i. e. we have*

$$s^2_{\text{sy}} > s^2_{\text{st}}. \tag{2}$$

Proof. In view of theorem 2 of [2], it suffices to show that for the sufficiently small a

$$e^{-a|p_0-p_0'|} > \frac{1}{|D_1|^2} \iint\limits_{D_1} \iint\limits_{D_2} e^{-a|p-p'|}\, dp\, dq, \tag{3}$$

where p_0 and p_0' are centres of gravity of D and D' and $|D|$ denotes the area of D (which is the same as that of D').

Now

$$e^{-a|p-q|} = 1 - a|p-q| + o(a|p-q|), \tag{4}$$

where $o(\cdot)$ denotes a quantity of smaller order; $|p-q|$ being bounded, the last term on the right side of (4) becomes negligible if a approaches zero. Therefore (3) will be implied by

$$|p_0 - p_0'| < \frac{1}{|D|^2} \iint\limits_{D} \iint\limits_{D'} |p-q|\, dp\, dq. \tag{5}$$

Putting $p_0 = (x_0, y_0)$, $p_0' = (x_0', y_0')$, $p = (x, y)$, $p' = (x', y')$, we may rewrite (5) as follows:

$$\sqrt{(x_0-x_0')^2+(y_0-y_0')^2} < \frac{1}{|D|^2}\iint_D \iint_{D'} \sqrt{(x-x')^2+(y-y')^2}\,dx\,dx'\,dy\,dy'. \tag{6}$$

In view of the Cauchy inequality

$$\begin{aligned} &\sqrt{(x_1-x_1')^2+(y_1-y_1')^2}+\sqrt{(x_2-x_2')^2+(y_2-y_2')^2} \\ &\quad\geqslant \sqrt{(x_1+x_2-x_1'-x_2')^2+(y_1+y_2-y_1'-y_2')^2} \\ &\quad= 2\sqrt{\left(\frac{x_1+x_2}{2}-\frac{x_1'+x_2'}{2}\right)^2+\left(\frac{y_1+y_2}{2}-\frac{y_1'+y_2'}{2}\right)^2} \end{aligned} \tag{7}$$

the function $\sqrt{(x-x')^2+(y-y')^2}$ of arguments x, x', y, y' is convex. Thus (6) is a special case of the Jensen inequality (see [1]). The theorem is thereby proved.

Comparing this theorem with theorem 4 of [2], we can see that there is no simple relation between the efficiencies of systematic and stratified sampling in a plane, even if we confine ourselves to exponential correlation functions and such regions as squares or regular hexagons.

REFERENCES

[1] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge 1934.

[2] S. Zubrzycki, *Remarks on random, stratified and systematic sampling in a plane*, Colloquium Mathematicum 6 (1958), p. 251-264.

Reçu par la Rédaction le 23. 3. 1960