COLLOQUIUM MATHEMATICUM

VOL. VI DÉDIÉ À M. CASIMIR KURATOWSKI 1958

REMARKS ON RANDOM, STRATIFIED AND SYSTEMATIC SAMPLING IN A PLANE

BY

S. ZUBRZYCKI (WROCŁAW)

§ 1. Introduction. In this note I will present some theorems on sampling in a plane. They are a contribution to the study of problems considered first for the linear case by Cochran. In the linear case Cochran [1] has proved that the systematic sampling is more efficient than both stratified and random sampling provided that the correlation function of the process in question is convex. Another result has been recently obtained by Hájek [3], who has shown that systematic sampling is the most efficient in a class of samplings where the expected number of sampling points falling in a fixed interval is proportional to the length of the interval. The plane case has been considered by several authors (see e. g. Quenouille [6], Williams [9]). Nevertheless it seems that this note brings in some new points to the set of problems under consideration. I have proved my theorems in connection with the problem of estimating gangue parameters (see [10], [11]) and I shall use this problem to illustrate the theory. The theorems of this note, although formulated for a plane, admit a straightforward generalization to the n-dimensional case.

§ 2. Probabilistic description of a gangue. We can distinguish two ways of treating probabilistically a problem of estimating a certain parameter, say a volume, of a gangue lying on a given domain, say on an exploitation block. The first consists in restricting the considerations to the already investigated domain and in regarding the measurements of a parameter, say thickness, as statistically independent observations. The thicknesses $y_1, \ldots, y_n$ measured at certain points of the investigated domain D are then regarded as realizations of n independent random variables with the same distribution function $F(x)$ defined for a given x as the ratio of the area of a part of D on which the thickness is less than x to the area of the whole domain D. Disregarding the shape of the distribution function $F(x)$, we can characterize the distribution of the random variables $y_1, \ldots, y_n$ by two numerical constants, their common expected value m and their common variance s^2. Our problem is then to estimate m on the basis of the values $y_1, \ldots, y_n$. Current statistical methods based on the concept of statistical independence of observations can be applied. Such were the very first applications of statistics to the problems of estimating parameters of geological gangues (see e. g. [7]) and they are thus presented in monographic geological literature (see e. g. Kreiter [4] or Smirnov [8]).

However, in order to justify the statistical independence of observations one must measure thickness at points chosen on D at random and independently. Usually we do not do so but distribute the points of measurements more or less systematically over D, believing rightly that greater exactness of estimation can be reached in this way. Therefore the use of current statistical methods in evaluating geological measurements is not satisfactorily justified. Moreover, the question arises how great an improvement of the exactness of estimation (if any) can be reached if a random distribution of measurement points is replaced by another, more systematical one. This question cannot be answered on the basis of the already described method of treating the problem.

The second method of treatment (adopted by me in [10] in connection with the estimation of gangue parameters and considered also by other authors, for the most part in connection with forestry and agricultural surveys, e. g. in [2] and [5]) consists in referring the postulates of goodness of estimation to the totality of blocks in a large gangue. Usually the size of an exploitation block can be considered very small as compared with the whole gangue, being often many kilometers long and broad. Not only the position of a measurement point on a block but also the position of the block on the gangue will be considered as random. In a more picturesque manner, we seek the volume of the gangue covered by a small carpet placed at random on a large gangue. This carpet plays the role of an exploitation block chosen at random. In this interpretation one describes the probabilistic structure of a gangue with the aid of a family of random variables $y(p)$, where the random variable $y(p)$ is assigned to the point p of a block D and represents the value of the parameter at that point. The distribution function $F(x;p)$ of a random variable $y(p)$ is conceived for a given x as the ratio of the measure of those positions of the block D for which the value of the parameter at the point p is less than x to the measure of all possible positions of that block. A similar meaning is given to the joint distribution function $F(x, \ldots, z; p, \ldots, q)$ of a system $y(p), \ldots, y(q)$ of random variables assigned to the points $p, \ldots, q$ of D.

The following assumptions on the random variables $y(p)$ suggest themselves naturally:

(a) all random variables $y(p)$ have the same distribution; in particular, for each p the expected value $\mathrm{E}y(p) = m$ and the variance $\mathrm{D}^2 y(p) = s^2$;

(b) the correlation coefficient $\mathrm{R}(y(p), y(q))$ between random variables $y(p)$ and $y(q)$ depends only on the vector $\overrightarrow{pq}$ joining the points p and q, i. e.

$$\mathrm{R}(y(p), y(q)) = f(\overrightarrow{pq});$$

obviously $f(\overrightarrow{pq}) = f(\overrightarrow{qp})$;

(c) $\lim\limits_{|pq|\to 0} f(\overrightarrow{pq}) = 1$, where $|pq|$ denotes the length of a vector $\overrightarrow{pq}$.

Sometimes it is possible to assume also that

(d) $f(\overrightarrow{pq}) = g(|\overrightarrow{pq}|)$, i. e. that the correlation function $f(\overrightarrow{pq})$ depends only on the length $|\overrightarrow{pq}|$ of the vector $\overrightarrow{pq}$ and not on its direction.

The assumptions (a), (b) and (c) characterize the family $y(p)$, $p \in D$, of random variables as a plane stationary continuous stochastic process. If in addition the condition (d) is fulfilled, the process $y(p)$ is called *isotropic*. Disregarding the shape of the distribution function $F(x;p)$ we describe the correlation properties of this process by numerical constants m and s and by a correlation function $f(\overrightarrow{pq})$. The probability distribution of the variables $y(p)$ plays the role of a theoretical counterpart of the possible positions of an exploitation block D.

The problem now is to estimate the mean value

$$\eta = \eta(D) = \frac{1}{|D|} \iint_D y(p)\,dp$$

(dp stands here for a differential of area, and $|D|$ for the area of D), if the values of some of the variables $y(p)$ are known.

§ 3. Three methods of sampling. In the sequel we will compare standard errors of the estimation of η on the basis of n observations when the n measurement points $p_1, \ldots, p_n$ are chosen in three different ways. The first method — random sampling — consists in choosing each measurement point independently from the others and with uniform probability distribution over D. The second method — stratified sampling — consists in choosing each measurement point independently of the others and so that the i-th point has uniform probability distribution over D_i, where $D_1, \ldots, D_n$ are n disjoint parts of D. Finally the third method — systematic sampling — consists in choosing the first point from D_1 with uniform probability distribution over D_1 and to take for the i-th measurement point, $i = 2, 3, \ldots, n$, a point of D_i corresponding to a point chosen on D_1 by a translation establishing the congruence of D_1 and D_i, the symbols $D_1, \ldots, D_n$ standing now for n parts of D which are disjoint and congruent by translation.

For the estimate of η we always take the arithmetic mean of random variables $y(p_1), \ldots, y(p_n)$, where $p_1, \ldots, p_n$ are the chosen points. Let us denote this mean by $\bar{\eta}_r$ in the case of random sampling, by $\bar{\eta}_{st}$ in the case of stratified sampling and by $\bar{\eta}_{sy}$ in the case of systematic sampling. Let us further denote by $s_r^2, s_{st}^2, s_{sy}^2$ the squares of standard errors of estimation for the three methods of sampling, i. e. let us put

$$s_r^2 = \mathrm{E}(\eta-\bar{\eta}_r)^2, \quad s_{st}^2 = \mathrm{E}(\eta-\bar{\eta}_{st})^2, \quad s_{sy}^2 = \mathrm{E}(\eta-\bar{\eta}_{sy})^2.$$

Our aim is to find explicit formulae for those errors (§ 4) and to prove some theorems concerning their comparison (§§ 5 and 6).

§ 4. Standard errors of estimation. In the sequel $\eta(A)$ will denote, for an arbitrary domain A, a random variable defined by the formula

$$\eta(A) = \frac{1}{|A|}\iint\limits_A y(p)\,dp. \tag{4.1}$$

We are now going to prove

THEOREM 1. (a) *For an arbitrary domain D we have*

$$s_r^2 = \frac{1}{n}\big(s^2-\mathrm{D}^2\eta(D)\big); \tag{4.2}$$

(b) *if the domain D is the sum of disjoint domains $D_1, \ldots, D_n$, congruent by translation, or if the process $y(p)$ is isotropic and the domains $D_1, \ldots, D_n$ are congruent, then*

$$s_{st}^2 = \frac{1}{n}\big(s^2-\mathrm{D}^2\eta(D_1)\big); \tag{4.3}$$

(c) *if the domain D is the sum of domains $D_1, \ldots, D_n$ which are disjoint and congruent by translation, then*

$$s_{sy}^2 = \frac{s^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} f(\overrightarrow{p_i p_j})-\mathrm{D}^2\eta(D), \tag{4.4}$$

where $p_i \in D_i$ is the centre of gravity of D.

Proof. We shall use the following identity, which holds for an arbitrary domain A:

$$\mathrm{D}^2\eta(A) = \frac{1}{|A|^2}\iint\limits_A\Big\{\iint\limits_A f(\overrightarrow{pq})\,dp\Big\}\,dq. \tag{4.5}$$

We assume also, without loss of generality, that $m = 0$.

Let us begin with (a). We obviously have

$$\bar{\eta}_r = \frac{1}{n}\big(y(P_1)+\ldots+y(P_n)\big),$$

where $P_1, \ldots, P_n$ are n mutually independent random variables, each of which has uniform probability distribution over D, and which are independent of the stochastic process $y(p)$. We have further, in view of $m = 0$,

$$\begin{aligned} s_r^2 &= \mathrm{E}(\eta(D)-\bar{\eta}_r)^2 \\ &= \mathrm{E}\eta^2(D)-2\mathrm{E}\eta(D)\bar{\eta}_r+\mathrm{E}\bar{\eta}_r^2 \\ &= \mathrm{D}^2\eta(D)-2\mathrm{E}\eta(D)\cdot\frac{1}{n}\big(y(P_1)+\ldots+y(P_n)\big)+\mathrm{E}\bar{\eta}_r^2 \\ &= \mathrm{D}^2\eta(D)-\frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\eta(D)y(P_i)+\mathrm{E}\bar{\eta}_r^2. \end{aligned} \tag{4.6}$$

In virtue of the definition of $\eta(D)$ we have for each particular point $p \in D$

$$\begin{aligned} \mathrm{E}\eta(D)y(p) &= \mathrm{E}y(p)\cdot\frac{1}{|D|}\iint\limits_D y(p)\,dp \\ &= \frac{1}{|D|}s^2\iint\limits_D f(\overrightarrow{pq})\,dq. \end{aligned} \tag{4.7}$$

Since, moreover, for each $i = 1, 2, \ldots, n$, the random variable P_i is independent of the process $y(p)$ and therefore also independent of $\eta(D)$, we have for $i = 1, 2, \ldots, n$

$$\begin{aligned} \mathrm{E}\eta(D)y(P_i) &= \frac{1}{|D|}\iint\limits_D \mathrm{E}\eta(D)y(p)\,dp \\ &= \frac{1}{|D|^2}s^2\iint\limits_D\Big\{\iint\limits_D f(\overrightarrow{pq})\,dq\Big\}\,dp = \mathrm{D}^2\eta(D), \end{aligned} \tag{4.8}$$

the last equality being a consequence of (4.5).

We have further

$$(4.9)\qquad \mathrm{E}\bar{\eta}_r^2 = \mathrm{E}\left[\frac{1}{n}\big(y(P_1)+\ldots+y(P_n)\big)\right]^2$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}y(P_i)y(P_j).$$

Since for every two points $p, q \in D$ we have $\mathrm{E}y(p)y(q) = s^2 f(\overrightarrow{pq})$, and for $i \neq j$ the random variables P_i and P_j are independent and uniformly distributed over D, we have for $i \neq j$

$$(4.10)\qquad \mathrm{E}y(P_i)y(P_j) = \frac{1}{|D|^2}\iint_D\left\{\iint_D f(\overrightarrow{pq})\,dq\right\}dp = \mathrm{D}^2\eta(D).$$

For $i = j$ we obviously have

$$(4.11)\qquad \mathrm{E}y(P_i)y(P_j) = \frac{1}{|D|}\iint_D \mathrm{E}y^2(p)\,dp = s^2.$$

In view of (4.10) and (4.11) it follows from (4.9) that

$$(4.12)\qquad \mathrm{E}\bar{\eta}_r^2 = \frac{1}{n^2}\big(ns^2+n(n-1)\mathrm{D}^2\eta(D)\big) = \frac{s^2}{n}+\frac{n-1}{n}\mathrm{D}^2\eta(D).$$

From (4.6), (4.8) and (4.12) we get

$$s_r^2 = \mathrm{D}^2\eta(D) - \frac{2}{n}\sum_{i=1}^{n}\mathrm{D}^2\eta(D) + \frac{s^2}{n} + \frac{n-1}{n}\mathrm{D}^2\eta(D)$$

$$= \frac{s^2}{n} + \left(1-2+\frac{n-1}{n}\right)\mathrm{D}^2\eta(D) = \frac{1}{n}\big(s^2-\mathrm{D}^2\eta(D)\big),$$

which proves (a).

Subsequently, let us prove (b). We now have

$$\bar{\eta}_{st} = \frac{1}{n}\big(y(P_1)+\ldots+y(P_n)\big),$$

where $P_1, \ldots, P_n$ are n mutually independent random variables, the variable P_i having uniform probability distribution over D_i, which are moreover independent of the stochastic process $y(p)$. We have, as before,

$$(4.13)\qquad s_{st}^2 = \mathrm{E}\big(\eta(D)-\bar{\eta}_{st}\big)^2$$

$$= \mathrm{E}\eta^2(D) - 2\mathrm{E}\eta(D)\bar{\eta}_{st} + \mathrm{E}\bar{\eta}_{st}^2$$

$$= \mathrm{D}^2\eta(D) - \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\eta(D)y(P_i) + \mathrm{E}\bar{\eta}_{st}^2.$$

Owing to (4.7) and to the probability distribution of P_i we can conclude that

$$\mathrm{E}\eta(D)y(P_i) = \frac{s^2}{|D_i|\,|D|}\iint_{D_i}\left\{\iint_D f(\overrightarrow{pq})\,dq\right\}dp$$

$$= \frac{ns^2}{|D|^2}\iint_{D_i}\left\{\iint_D f(\overrightarrow{pq})\,dq\right\}dp.$$

It follows that

$$(4.14)\qquad \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\eta(D)y(P_i) = \frac{2}{n}\sum_{i=1}^{n}\frac{n}{|D|^2}\iint_{D_i}\left\{\iint_D f(\overrightarrow{pq})\,dq\right\}dp$$

$$= \frac{2}{|D|^2}\iint_D\left\{\iint_D f(\overrightarrow{pq})\,dq\right\}dp = 2\mathrm{D}^2\eta(D).$$

We have further, as before,

$$\mathrm{E}\bar{\eta}_{st}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}y(P_i)y(P_j).$$

Since for $i \neq j$ the random variables P_i and P_j are independent and have uniform probability distribution over D_i and D_j respectively, we get for $i \neq j$

$$\mathrm{E}y(P_i)y(P_j) = \frac{s^2}{|D_i|\,|D_j|}\iint_{D_i}\left\{\iint_{D_j} f(\overrightarrow{pq})\,dq\right\}dp.$$

For $i = j$ we obviously have $\mathrm{E}y(P_i)y(P_j) = \mathrm{E}y^2(P_i) = s^2$. Therefore we have

$$(4.15)\qquad \mathrm{E}\bar{\eta}_{st}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}y(P_i)y(P_j)$$

$$= \frac{1}{n^2}\Big[\sum_{\substack{i,j=1\\ i\neq j}}^{n}\frac{s^2}{|D_i|\,|D_j|}\iint_{D_i}\Big\{\iint_{D_j} f(\overrightarrow{pq})\,dq\Big\}\,dp + ns^2\Big]$$

$$= \frac{s^2}{n} + \frac{s^2}{n^2}\Big[\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{|D_i|\,|D_j|}\iint_{D_i}\Big\{\iint_{D_j} f(\overrightarrow{pq})\,dq\Big\}\,dp - \sum_{j=1}^{n}\frac{1}{|D_j|^2}\iint_{D_j}\Big\{\iint_{D_j} f(\overrightarrow{pq})\,dq\Big\}\,dp\Big]$$

$$= \frac{s^2}{n} + \frac{s^2}{|D|^2}\iint_{D}\Big\{\iint_{D} f(\overrightarrow{pq})\,dq\Big\}\,dp - \sum_{j=1}^{n}\frac{s^2}{n^2|D_j|^2}\iint_{D_j}\Big\{\iint_{D_j} f(\overrightarrow{pq})\,dq\Big\}\,dp$$

$$= \frac{s^2}{n} + \mathrm{D}^2\eta(D) - \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}^2\eta(D_i)$$

$$= \frac{s^2}{n} + \mathrm{D}^2\eta(D) - \frac{1}{n}\mathrm{D}^2\eta(D_1),$$

the last equality following from the fact that because of the assumed congruence of the domains $D_1, \ldots, D_n$ we have $\mathrm{D}^2\eta(D_1) = \ldots = \mathrm{D}^2\eta(D_n)$.

From (4.13), (4.14) and (4.15) we get

$$(4.16)\qquad s_{st}^2 = \mathrm{D}^2\eta(D) - 2\mathrm{D}^2\eta(D) + \frac{s^2}{n} + \mathrm{D}^2\eta(D) - \frac{1}{n}\mathrm{D}^2\eta(D_1)$$

$$= \frac{1}{n}\big(s^2 - \mathrm{D}^2\eta(D_1)\big),$$

which proves (b).

In the third case, (c), we have

$$\bar{\eta}_{sy} = \frac{1}{n}\big(y(P_1)+\ldots+y(P_n)\big),$$

where $P_1, \ldots, P_n$ are now random variables independent of the process $y(p)$ and such that P_i has uniform probability distribution over D_i. But the variables $P_1, \ldots, P_n$ themselves are now dependent: their values always correspond to one another in the sense of translations establishing the congruence of particular parts of D. If we denote by v_i, $i = 1, 2, \ldots, n$, a vector by which D_1 should be translated to obtain the coincidence with D_i, we can write with probability 1 the equalities

$$(4.17)\qquad P_i = P_1 + v_i$$

(obviously $v_1 = 0$).

We have again

$$(4.18)\qquad s_{sy}^2 = \mathrm{E}(\eta(D) - \bar{\eta}_{sy})^2$$

$$= \mathrm{D}^2\eta(D) - \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\eta(D)y(P_i) + \mathrm{E}\bar{\eta}_{sy}^2.$$

We have further

$$(4.19)\qquad \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\eta(D)y(P_i) = 2\mathrm{D}^2\eta(D),$$

for the same reasons as in the case of (4.14).

As to the last term of (4.18), we obviously have

$$\mathrm{E}\bar{\eta}_{sy}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}y(P_i)y(P_j).$$

But if $P_1 = p$, then, according to (4.17), $P_i = p + v_i$. Therefore, in view of the mutual independence of the system of random variables $P_1, \ldots, P_n$ and the process $y(p)$, the expected value of $\bar{\eta}_{sy}^2$, under condition $P_1 = p$, is equal to

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}s^2 f(\overrightarrow{p+v_i,\, p+v_j}),$$

and thus is independent of p. This proves that

$$(4.20)\qquad \mathrm{E}\bar{\eta}_{sy}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}s^2 f(\overrightarrow{p_ip_j}),$$

where $p_1, \ldots, p_n$ are any points such that $\overrightarrow{p_1p_i} = v_i$. From (4.18), (4.19) and (4.20) it follows that (4.4) is true.

Theorem 1 is thus fully proved.

§ 5. Comparison of errors. We now proceed to prove the following theorem concerning the comparison of estimation errors:

THEOREM 2. (e) *If a domain D is the sum of disjoint domains $D_1, D_2, \ldots, D_n$ congruent by translation, or if the process $y(p)$ is isotropic and the domains $D_1, \ldots, D_n$ are congruent, then*

$$s_{st}^2 \leqslant s_r^2. \tag{5.1}$$

(f) *In order that, for a domain D which is the sum of disjoint domains $D_1, \ldots, D_n$ congruent by translation, we have the relation*

$$s_{sy}^2 \leqslant s_{st}^2, \tag{5.2}$$

it is sufficient that for each i and j we have

$$f(\overrightarrow{p_ip_j}) \leqslant \frac{1}{|D_1|^2} \iint_{D_i} \Big\{ \iint_{D_j} f(\overrightarrow{pq})\, dq \Big\}\, dp, \tag{5.3}$$

p_i being the centre of gravity of D_i; if we want relation (5.2) to hold also for each domain D' of the form $D_i \cup D_j$, $i \neq j$, then condition (5.3) is also necessary.

Proof. For the sake of simplicity we shall assume throughout the proof that the expected value m is equal to 0. Let us begin with (e).

It immediately follows from definition (4.1) that

$$\eta(D) = \frac{1}{n}\big(\eta(D_1)+\ldots+\eta(D_n)\big). \tag{5.4}$$

By virtue of our assumptions, $\eta(D)$ is thus an average of n random variables with equal variances. Further, we have

$$D^2\eta(D) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E\eta(D_i)\eta(D_j). \tag{5.5}$$

Since, by the Schwarz inequality, we have for each i and j

$$E\eta(D_i)\eta(D_j) \leqslant D^2\eta(D_1), \tag{5.6}$$

the equality being possible for $i \neq j$ only if the variables $\eta(D_i)$ and $\eta(D_j)$ are equal with probability 1, it follows from (5.4)-(5.6) that $D^2\eta(D) \leqslant D^2\eta(D_1)$. This, together with (4.2) and (4.3), proves (5.1).

As to (f), let us consider the difference $s_{sy}^2 - s_{st}^2$. Owing to (4.3), (4.4), (4.5) and (5.5) we can write the following chain of equalities:

$$\begin{aligned} s_{sy}^2 - s_{st}^2 &= \frac{s^2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} f(\overrightarrow{p_ip_j}) - D^2\eta(D) - \frac{s^2}{n} + \frac{1}{n} D^2\eta(D_1) \\ &= \frac{1}{n^2}\Big[\sum_{i=1}^{n} \sum_{j=1}^{n} \big(s^2 f(\overrightarrow{p_ip_j}) - E\eta(D_i)\eta(D_j)\big) - ns^2 + n\cdot D^2\eta(D_1)\Big] \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \big[s^2 f(\overrightarrow{p_ip_j}) - s^2\delta_{ij} + \delta_{ij} D^2\eta(D_1) - E\eta(D_i)\eta(D_j)\big]; \end{aligned} \tag{5.7}$$

we have used here δ_{ij} to denote 1 for $i = j$ and 0 for $i \neq j$. Since the variables $\eta(D_1), \ldots, \eta(D_n)$ have equal variances and since, according to (c) of § 3, $f(\overrightarrow{pp}) = 1$, we get from (5.7)

$$\begin{aligned} s_{sy}^2 - s_{st}^2 &= \frac{1}{n^2} \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \big[s^2 f(\overrightarrow{p_ip_j}) - E\eta(D_i)\eta(D_j)\big] \\ &= \frac{s^2}{n^2} \sum_{\substack{i,j=1 \\ i\neq 1}}^{n} \Big[f(\overrightarrow{p_ip_j}) - \frac{1}{|D_1|^2} \iint_{D_i} \Big\{\iint_{D_j} f(\overrightarrow{pq})\, dq\Big\}\, dp\Big]. \end{aligned}$$

This identity proves the sufficiency of (5.3). Now in the case of a domain D' of the form $D_i \cup D_j$, $i \neq j$, the equality (5.8) reduces to

$$s_{sy}^2 - s_{st}^2 = \frac{s^2}{4} f(\overrightarrow{p_ip_j}) - \frac{1}{|D_1|^2} \iint_{D_i} \Big\{\iint_{D_j} f(\overrightarrow{pq})\, dq\Big\}\, dp.$$

This proves the necessity of (5.3).

The proof of theorem 2 is thus completed.

§ 6. Exponential correlation function. In this section we shall be concerned with isotropic processes, that is with the processes whose correlation function $f(\overrightarrow{pq})$ depends only on the length $|\overrightarrow{pq}|$ of the vector $\overrightarrow{pq}$ or, which amounts to the same, on the distance $d = d(p, q)$ of the points p and q, $f(\overrightarrow{pq}) = g(d)$.

Among those processes we are especially interested in those having the exponential correlation function, i. e. in those for which $g(d) = e^{-ad}$, where a is a positive constant. It is remarkable that it is exactly the exponential function that fits observations of different kinds considerably well (see e. g. [5], [10]).

Now, it is easy to see that for suitably narrow and long rectangles D_i stratified sampling may be more efficient than systematic sampling. But it is somewhat surprising that this can happen for domains D_i as round as circles. Namely we have

THEOREM 3. *If the process $y(p)$ has the exponential correlation function, i. e. if*

$$f(\overrightarrow{pq}) = e^{-ad}, \tag{6.1}$$

d standing here for the distance of the points p and q, and if the domain D has a diameter less than $1/a$ and is composed of disjoint circles $D_1, \ldots, D_n$ of equal size, then stratified sampling is more efficient than systematic sampling, i. e. we have

$$s_{st}^2 < s_{sy}^2. \tag{6.2}$$

This theorem shows that the roundness of circles can in some cases be a satisfactory guarantee against the doubling of observations, which occurs when the sampling points are very near and which is the main reason why systematic sampling is often more efficient than stratified sampling.

Proof. In order to prove (6.2) it suffices to show that for each i and j, $i \neq j$, we have

$$f(\overrightarrow{p_i p_j}) > \frac{1}{|D_1|^2} \iint_{D_i} \left\{ \iint_{D_j} f(\overrightarrow{pq})\, dq \right\} dp; \tag{6.3}$$

here p_i is the centre of D_i. These inequalities are opposite to (5.3).

To prove (6.3) it will be convenient to introduce in a plane of points p an orthogonal system of coordinates x and y. Then we can write our correlation function in the form

$$f(\overrightarrow{pq}) = f(x, y) = e^{-a\sqrt{x^2+y^2}},$$

where x and y play the role of coordinates of the vector $\overrightarrow{pq}$. Now the function $f(x, y)$ appears to be subharmonic in the domain $0 < \sqrt{x^2+y^2} < 1/a$, as is shown by the equality

$$\frac{\partial^2}{\partial x^2} f(x, y) + \frac{\partial^2}{\partial y^2} f(x, y) = a \cdot e^{-a\sqrt{x^2+y^2}} \left(a - \frac{1}{\sqrt{x^2+y^2}} \right).$$

It easily follows from the properties of subharmonic functions that for each point p in D_i we shall have, for $j \neq i$,

$$f(\overrightarrow{pp_j}) > \frac{1}{|D_1|} \iint_{D_j} f(\overrightarrow{pq})\, dq.$$

Integrating this inequality over D_i we get

$$\frac{1}{|D_1|} \iint_{D_i} f(\overrightarrow{pp_j})\, dp > \frac{1}{|D_1|^2} \iint_{D_i} \left\{ \iint_{D_j} f(\overrightarrow{pq})\, dq \right\} dp.$$

For the same reasons as in the case of the last but one inequality we have

$$f(\overrightarrow{p_i p_j}) > \frac{1}{|D_1|} \iint_{D_i} f(\overrightarrow{pp_j})\, dp.$$

The last two inequalities prove (6.3), which completes the proof of theorem 3.

Theorem 3 shows, roughly speaking, that for a given exponential correlation function there exist domains D for which stratified sampling is more efficient than systematic sampling. However, for any domain D systematic sampling becomes more efficient than stratified sampling provided that the constant a in the exponential correlation function is sufficiently large, as is shown by

THEOREM 4. *If the process $y(p)$ has the exponential correlation function of the form (6.1) and D is the sum of disjoint domains $D_1, \ldots, D_n$, congruent by translation, then for sufficiently large a a systematic sampling is more efficient than stratified sampling, i. e. we have*

$$s_{sy}^2 < s_{st}^2. \tag{6.4}$$

This theorem is an immediate consequence of the following

LEMMA. *If $p(x)$ is a probability density on a half-line $0 < x < \infty$ and x_0 is any fixed number such that*

$$\int_0^{x_0} p(x)\, dx > 0,$$

then for sufficiently large a we have

$$e^{-ax_0} < \int_0^{\infty} e^{-ax} p(x)\, dx. \tag{6.5}$$

Proof of the lemma. We have

$$\int_0^{\infty} e^{-ax} p(x)\, dx - e^{-ax_0} = e^{-ax_0} \int_0^{\infty} (e^{-a(x-x_0)} - 1) p(x)\, dx.$$

Now it is seen that the function under the integral sign is positive for $x < x_0$ and negative for $x_0 < x$. Moreover, on the half-line $x_0 < x < \infty$ it is bounded from below by -1, and in each interval $0 < x < x'$, where $x' < x_0$, it increases over all bounds when a tends to infinity. Since we can always choose x' so that $\int_0^{x'} p(x)dx > 0$, we can conclude that for sufficiently large a the considered integral becomes positive. The lemma is thus proved.

Proof of theorem 4. Let us consider two domains D_i and D_j. If $p(x)$ is the probability density of the distance of two points chosen at random and independently of D_i and D_j, respectively, and if x_0 is the distance between the centres of gravity of D_i and D_j, then we can apply the lemma to $p(x)$ and x_0. Thus for sufficiently large a we get inequality (6.5). This, however, is nothing else but another form of inequality (5.3), for a given i and j. Since this reasoning is applicable to each pair of domains D_i, D_j, we can choose a so large that inequalities (5.3) hold for all i and j, $i \neq j$. This proves our theorem.

Finally let us note that the following question is still open:

P 254. Is always systematic sampling, for isotropic processes $y(p)$ with the exponential correlation function and domains D composed of squares or of regular hexagons, more efficient than stratified sampling?

REFERENCES

[1] W. G. Cochran, *Relative accuracy of systematic and stratified samples for a certain class of populations*, Annals of Mathematical Statistics 17 (1946), p. 164-177.

[2] D. J. Finney, *Random and systematic sampling in timber surveys*, Forestry 22 (1948), p. 64-99; 23 (1950), p. 96-111 (quoted after [9]).

[3] J. Hájek, *Příspěvky k teorii statistického odhadu*, kandidátská disertace, Praha 1955. *(Some contributions to the theory of estimation*, Thesis*)*.

[4] В. М. Крейтер, *Поиски и разведки полезных ископаемых*, Москва 1940.

[5] B. Matern, *Metoder att uppskata noggranhetten vid linje- och provytetaxering (Methods of estimating the accuracy of line and sample plot surveys)*, Meddelanden från Statens Skogsforskningsinstitut 36, 1 (1947), p. 1-138.

[6] M. H. Quenouille, *Problems in plane sampling*, Annals of Mathematical Statistics 20 (1949), p. 355-375.

[7] П. А. Рыжов, *Математическое определение „геологической ошибки" при подсчёте запасов полезных ископаемых*, Труды Казах. горн.-металл. института 1 (1938) (quoted after [8]).

[8] В. И. Смирнов, *Подсчёт запасов минерального сырья*, Москва 1950.

[9] R. M. Williams, *The variance of the mean of systematic samples*, Biometrika 43 (1956), p. 137-148.

[10] S. Zubrzycki, *O szacowaniu parametrów złóż geologicznych (On estimating gangue parameters)*, Zastosowania Matematyki 3 (1957), p. 105-153.

[11] — *Szacowanie złóż jako zagadnienie statystyczne (Estimating ore-gangues as a statistical problem)*, to be published in Przegląd Geologiczny.

MATHEMATICAL INSTITUTE OF THE POLISH ACADEMY OF SCIENCES

Reçu par la Rédaction le 7. 3. 1958

COLLOQUIUM MATHEMATICUM

VOL. VI DÉDIÉ À M. CASIMIR KURATOWSKI 1958

REMARKS ON COMPACT SEMIGROUPS

BY

J. ŁOŚ (TORUŃ) AND Š. SCHWARZ (BRATISLAVA)

The purpose of this note is to show that some theorems concerning Hausdorff compact (= bicompact) semigroups can be proved by a unified method based on a very simple topological lemma.

Theorem 1, Corollary to Theorem 2 and Theorem 2a are known; we bring them into our considerations to emphasize the method presented in the paper and for the convenience of the reader, since they are needed in the rest of the paper. The remaining results seem to be new, though probably they may be partly known to those who are occupied with compact semigroups. The results of section 1 and a part of the results of section 2 are closely connected with some results stated in a paper of Wallace [1] (see also Wallace [2]).

1. We shall be concerned with functions f of k variables running over a set X and with values in the same set X. We shall write $f(\overset{k}{\breve{X}})$ or $f^{(1)}(\overset{k}{\breve{X}})$ instead of $f(\underbrace{X\times\ldots\times X}_{k\text{ times}})$ and $f^{(n)}(\overset{k}{\breve{X}})$ instead of $f[f^{(n-1)}(\overset{k}{\breve{X}})\times\ldots\times f^{(n-1)}(\overset{k}{\breve{X}})]$.

LEMMA 1. *Let X be a compact space and $f(x_1, \ldots, x_k)$ a continuous function for which $f(\overset{k}{\breve{X}}) \subset X$. Let X_ξ, $\xi < a$, be a decreasing family of non-empty closed subsets of X. Suppose that for each $\xi_0 < a$ there exist ξ_1, ξ_2, $\xi_0 \leqslant \xi_1 \leqslant \xi_2 < a$ such that $f(\overset{k}{\breve{X}}_{\xi_1}) = X_{\xi_2}$. Then $f(\bigcap_{\xi<a} \overset{k}{\breve{X}}_\xi) = \bigcap_{\xi<a} X_\xi$.*

Proof. Let φ_ξ, ψ_ξ be two ordered sets of ordinal numbers both confinal in a and such that $f(\overset{k}{\breve{X}}_{\varphi_\xi}) = X_{\psi_\xi}$. (The existence of such sequences follows from the suppositions.) Since X_ξ is a decreasing family we have

$$\bigcap_{\xi<a} X_\xi = \bigcap_{\xi<a} X_{\varphi_\xi} = \bigcap_{\xi<a} X_{\psi_\xi} \neq \emptyset.$$