Ratio Estimate in Line-grid Sampling

Masuyama, Motosaburo
Institute of Physical Therapy & Intenal Medicine, Tokyo University | Meteorogical Research Institute

https://doi.org/10.5109/12978

出版情報：統計数理研究. 7 (3/4), pp.73-76, 1957-03. Research Association of Statistical Sciences
バージョン：
権利関係：

RATIO ESTIMATE IN LINE-GRID SAMPLING

By

Motosaburô MASUYAMA

(*Received* *Feb.* 11, 1957)

Let us suppose that there are two sets of areas, say rice fields and vegetable gardens, in a domain of known size T, which satisfies the condition of unbiasedness, and that the total ϕ_1 of areas in the first set is known and the total ϕ_2 of areas in the second set is to be estimated by line-grid sampling.[5]

We assume that two sets are disjoint, i.e. no part of a rice field is used as a vegetable garden and no part of a vegetable garden is used as a rice field either. Let the length of line grid be L and that of the meet of the line-grid and area or areas be l. Let us call two ends of the line-grid P and Q respectively. We select at random one point (x, y), say A, in the domain T and P is identified with A, the direction and the sense of the segment PQ being that of x-axis. Then we have[1]

$$(1)\qquad E\{l\} = L\phi/T,$$

where ϕ is the total of areas in which l is to be measured.

As a special case of the inequality (2.9) in the previous paper,[3] we get

$$(2)\qquad \iint l^2 dx dy \leqq L^2\phi .^{*}$$

Thus the variance of l is equal to

$$(3)\qquad \sigma_l^2 \leqq L^2\phi/T - (L\phi/T)^2 = L^2 p(1-p),$$

where $p = \phi/T$ is the ratio, the total of areas to the whole domain, so that the variance of l is not greater than that of the binomial case with parameter p, when we use the grid of unit length.

In a large sample, let the variance of the estimate obtained by the line-grid sampling be σ_ϕ^2, then we have

$$(4)\qquad \sigma_\phi^2 \doteqdot T^2\sigma_l^2/L^2 \leqq T^2 p(1-p) = \phi(T-\phi)$$

and the coefficient of variation of ϕ

$$(5)\qquad \text{C.V.}_\phi \leqq \sqrt{T/\phi - 1},$$

* G. Kallianpur has shown the author the formula

$$E\{l^n\} = \left[L\int_a^b Dw^{n-1}dy - (n-1)\int_a^b w^{n+1}dy/(n+1)\right]/T$$

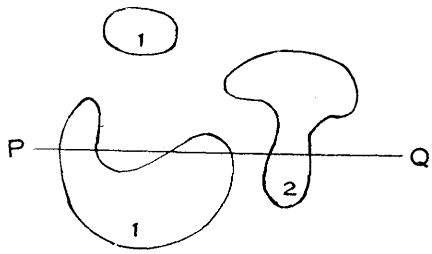
under the condition of unbiasedness, assuming a single convex area in T. The area is cut by $y=y$ at $x=x_1$ and x_2. w is equal to the minimum of L and $D = x_2 - x_1$. a and b $(>a)$ are the extrema of ordinate y of the boundary curve of the area.

so that it is desirable to make T as small as possible. A compromise would be necessary in cases where we need the extended area to ensure the condition of unbiasedness.

Now let us use suffices 1 and 2 pertinent to the first and the second set respectively. Our problem is to estimate the covariance of l_1 and l_2.

Having assumed the non-overlapping of two kinds of areas, an infinitesimal increase of l_i is always accompanied with an infinitesimal non-increase of l_j $(i \neq j)$. However, l_1 and l_2 are not always negatively correlated. For example, suppose a part of xy-plane, say $0 \leqq x \leqq A$, is cut into strips of width Δx each and l_1 and l_2 be functions of x, then the locus of a point (l_1, l_2) on the $l_1 l_2$-plane is a pantograph-like figure, which is reduced to a segment through the origin 0, if Δx tends to 0. This means a case where l_1 and l_2 are positively correlated.

To be realistic, we assume in the sequal that L is fairly small so that the line-grid meets at most two areas of different kinds simultaneously.



Let us start from the simplest case of two areas of different kinds and assume the convexity of each area.

For the sake of simplicity, we assume further that the meet l_1 lies to the left of the meet l_2. Let us consider two areas of different kinds F_1 and F_2 and let their widths at $y=y$ be D_1 and D_2 respectively. If the shortest distance of these areas at $y=y$, say δ, is greater than L, then we have $l_1 l_2 = 0$, so that we assume $L-\delta=a>0$ in the following 5 cases i)–v).

i) when $D_1 \geqq a$ and $D_2 \geqq a$, we have

$$\int l_1 l_2 dx = a^3/6. \tag{6}$$

ii) when $D_1 \geqq a$ and $D_2 \leqq a$, we have

$$\int l_1 l_2 dx = [(D_1 - a)^3 + a^3]/6. \tag{7}$$

iii) when $D_1 \leqq a$ and $D_2 \geqq a$, we have

$$\int l_1 l_2 dx = [(D_2 - a)^3 + a^3]/6. \tag{8}$$

iv) when $D_1 \leqq a$, $D_2 \leqq a$ and $D_1 + D_2 \geqq a$, we have

$$\int l_1 l_2 dx = [(D_1 - a)^3 + (D_2 - a)^3 + a^3]/6. \tag{9}$$

v) when $D_1 + D_2 \leqq a$, we have

$$\int l_1 l_2 dx = D_1 D_2 (2a - D_1 - D_2)/2. \tag{10}$$

In all these cases we have

$$\int l_1 l_2 dx \leqq a^3/6, \tag{11}$$

which is apparent from geometrical consideration. If we do not assume the convexity, the right side of the above integral is complex but still the inequality (11) holds, because l_1 and l_2 are always majorated by l_1 and l_2 of the case i) respectively.
Thus we have

$$\iint l_1 l_2 dx dy \leqq \int a^3 dy/6 \leqq L^2 \int a dy/6 . \tag{12}$$

We displace every point of F_2 in parallel to the left by L to form a new area F_2^*. Next suppose that F_1 is a cross section of an obstacle and a light source at $x = +\infty$ makes its umbra $\overline{F}_1$. The union of $\overline{F}_1$ and F_1 is denoted by F_1'. Likewise a light source at $x = -\infty$ makes the umbra $\overline{F}_2^*$ and the union of F_2^* and $\overline{F}_2^*$ is denoted by F_2'. Then the union of F_1' and F_2' is equal to

$$\psi_{12} = \int a\, dy . \tag{13}$$

Accordingly we get the inequality

$$\mathrm{cov}\,(l_1 l_2) \leqq L^2(\psi_{12}/6 - F_1 F_2/T)/T , \tag{14}$$

which means this: If the ratio ψ_{12}/F_1 is less than $6F_2/T$, then the covariance between l_1 and l_2 is negative. When l_1 lies to the right of l_2, the similar inequality holds for ψ_{21}.

In general case, combining any one area F_{1i} of the first set and any one area F_{2j} of the second set, we get

$$\psi_{ij}/6 < F_{1i} F_{2j}/T , \tag{15}$$

where the area ψ_{ij} is similarly defined as above, as a sufficient condition of negative covariance. In this case, we take

$$f_2 = \overline{l}_2 (T - \phi_1)/(L - \overline{l}_1) \tag{16}$$

as an estimate of ϕ_2, where bar denotes sample mean as usual. As is well-known, the efficiency of this estimate depends not only upon the correlation coefficient ρ between l_2 and $(-l_1)$ but also upon the ratio of two coefficients of variation $\mathrm{C.V.}_{l_1}$ and $\mathrm{C.V.}_{l_2}$ and the latter $\times 2\rho$ is desirable to be greater than the former. As we have

$$\mathrm{C.V.}_l^2 = \sigma_l^2 T^2/(L^2 \phi^2) , \tag{17}$$

the variance of l_2 is desirable to large and ϕ_2 is desirable to be small. With regards to l_1, the converse is desirable.

This sort of ratio estimate would be useful to estimate ϕ_2, when sub-areas of ϕ_2 are not so dense in T, if we insert artificially areas of known shape and size which need not be distributed at random in T. To be efficient, $\mathrm{C.V.}_{l_1}$ should be less than $\sqrt{2}$.

METEOROGICAL RESEARCH INSTITUTE, TOKYO;
INSTITUTE of PHYSICAL THERAPY & INTERNAL MEDICINE, TOKYO UNIVERSITY.

References

[1] M. MASUYAMA; *Rapid methods of estimating the sum of specified areas in a field of given size*, Rep. Stat. Appl. Res., **2** (1953), 133-119, **3** (1954), 32.

[2] M. MASUYAMA; *On geometrical methods of survey*, Operations Research, **1** (1956), 41-49. (in Japanese)

[3] M. MASUYAMA; *On a fundamental formura in bulk sampling from the viewpoint of integral geometry*, Rep. Stat. Appl. Res., **4** (1956), 85-89.

[4] M. MASUYAMA; *On an application of integral geometry*, Kagaku (Science), **26** (1956), 579. (in Japanese)

[5] J. M. SENGUPTA; *Some experiments with different types of area sampling for winter paddy*, in Giridih, Bihar, Sankhyā, **13** (1954), 235-240.