ARKIV FÖR MATEMATIK Band 4 nr 29

Communicated 22 February 1961 by H. Cramér and O. Frostman

On a maximizing problem with several applications in statistical theory

By Carl-Erik Särndal

SUMMARY

For a certain distribution, the three problems of determining

(i) optimum spacings of observations in linear estimation of location and scale parameters
(ii) optimum strata limits in proportionate sampling
(iii) optimum interval boundaries in grouping a sample

are closely related to each other.

In section two of this paper, two basic lemmas are stated. Applied to problem (i) above, these lemmas enable us to make conclusions about the rate at which the efficiency of linear estimates increases with the number of observations used to form the estimate. Similarly, in problem (ii), statements can be made concerning the gain in accuracy due to increasing the number of strata.

Furthermore, the laborious procedure of calculating the optimum solutions of the above mentioned problems is replaced by a simplified approximation technique which will cause negligible loss of accuracy.

2. Two lemmas

Let $H_r(\lambda)$, $(r=1, 2)$, be real-valued functions, defined for $0 \leqslant \lambda \leqslant 1$, and form the sums

$$K_{rs}(\lambda_1, \ldots, \lambda_k) = \sum_{i=1}^{k+1} \frac{[H_r(\lambda_i) - H_r(\lambda_{i-1})]\,[H_s(\lambda_i) - H_s(\lambda_{i-1})]}{\lambda_i - \lambda_{i-1}}, \quad (r, s=1, 2), \tag{1}$$

where $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_k < \lambda_{k+1} = 1$.

Two groups of conditions are imposed upon $H_r(\lambda)$, $(r=1, 2)$, and its derivatives.

A1. The functions $H_r(\lambda)$, $(r=1, 2)$, are continuous for $0 \leqslant \lambda \leqslant 1$ and tend to zero as $\lambda \to 0$ and $\lambda \to 1$.

A2. The derivatives $H_r^{v}(\lambda)$, $(r=1, 2)$, of order $1, \ldots, 5$ are continuous and limited for $0 < \lambda < 1$.

A3. The integrals

$$e_{rs} = \int_0^1 H_r'(t)\, H_s'(t)\, dt, \quad (r, s=1, 2), \tag{2}$$

exist and $e_{11} e_{22} - e_{12}^2 > 0$.

A4. $$e_{11}[H_2''(\lambda)]^2 - 2e_{12}H_1''(\lambda)H_2''(\lambda) + e_{22}[H_1''(\lambda)]^2 \neq 0$$

for all λ; $0 < \lambda < 1$.

A5. The integral

$$c = \int_0^1 \{e_{11}[H_2''(t)]^2 - 2e_{12}H_1''(t)H_2''(t) + e_{22}[H_1''(t)]^2\}^{\frac{1}{5}}\,dt \tag{3}$$

exists.

In the second group of conditions, the subscript r can take on either of the values 1 and 2.

B1. The function $H_r(\lambda)$ is continuous for $0 \leqslant \lambda \leqslant 1$ and tends to zero as $\lambda \to 0$ and $\lambda \to 1$.

B2. The derivatives $H_r^v(\lambda)$ or order 1, ..., 5 are continuous and limited for $0 < \lambda < 1$.

B3. The integral

$$e_{rr} = \int_0^1 [H_r'(t)]^2\,dt$$

exists and $e_{rr} > 0$.

B4. $H_r''(\lambda) \neq 0$ for all λ; $0 < \lambda < 1$.

B5. The integral

$$c_r = \int_0^1 [H_r''(t)]^{\frac{2}{5}}\,dt \tag{4}$$

exists.

Also, define

$$\pi_i = \frac{i}{k+1}, \quad (i = 0, 1, \dots k, k+1). \tag{5}$$

We can now state the two lemmas (although the proofs are omitted here).

Lemma 1. *If conditions A1–A5 are fulfilled, then the determinant*

$$K(\lambda_1, \dots, \lambda_k) = \begin{vmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{vmatrix}$$

attains a maximum for

$$\lambda_i = Q(\pi_i) + O(k+1)^{-2}, \quad (i = 1, \dots, k), \tag{6}$$

where π_i *is defined by* (5) *and*

$$Q(y) = M^{-1}(y)$$

is the inverse function of

$$M(y) = \frac{1}{c}\int_0^y \{e_{11}[H_2''(t)]^2 - 2e_{12}H_1''(t)H_2''(t) + e_{22}[H_1''(t)]^2\}^{\frac{1}{5}}\,dt,$$

where c *is given by* (3).

Furthermore, the maximum value of K is

$$K_{\max} = e_{11}\, e_{22} - e_{12}^2 - \frac{c^3}{12\,(k+1)^2} + o\,(k+1)^{-2}. \tag{7}$$

The value taken by K for

$$\lambda_i = Q\,(\pi_i), \quad (i = 1, \ldots, k), \tag{8}$$

can also be written in the form (7).

Lemma 2. *If conditions B1–B5 with $r=1$ or 2 are fulfilled, then the sum K_{rr}, ($r=1$ or 2), attains a maximum for*

$$\lambda_i = Q_r\,(\pi_i) + 0\,(k+1)^{-2}, \quad (i = 1, \ldots, k;\ r = 1 \text{ or } 2), \tag{9}$$

where π_i is defined by (5) *and*

$$Q_r\,(y) = M_r^{-1}\,(y)$$

is the inverse function of

$$M_r\,(y) = \frac{1}{c_r} \int_0^y [H_r''\,(t)]^{\frac{1}{3}}\, dt,$$

where c_r is given by (4).

The maximum value of K_{rr} is

$$K_{rr,\,\max} = e_{rr} - \frac{c_r^3}{12\,(k+1)^2} + o\,(k+1)^{-2}. \tag{10}$$

The value of K_{rr} for

$$\lambda_i = Q_r\,(\pi_i), \quad (i = 1, \ldots, k;\ r = 1 \text{ or } 2), \tag{11}$$

can also be expressed in the form (10).

3. Optimum spacing in linear estimation

A large sample of n observations from a population with a continuous cdf of type $F\,[(z-\alpha_1)/\alpha_2]$ is available. The unknown parameters α_1 and α_2 are to be estimated by linear estimates using only k out of the n observations namely

$$z_{(n_1)} < z_{(n_2)} < \cdots < z_{(n_k)}. \tag{12}$$

These are to be optimally selected. If $x = (z-\alpha_1)/\alpha_2$, we have

$$E\,x_{(n_i)} = x_i + 0\,(n^{-\frac{1}{2}}),$$

where $x_i = G\,(\lambda_i)$, $\lambda_i = n_i/(n+1)$, and $G\,(x) = F^{-1}\,(x)$ is the inverse of F.

Also define

$$H_1(\lambda) = f[G(\lambda)],$$
$$H_2(\lambda) = G(\lambda)\, f[G(\lambda)], \tag{13}$$

where $f(x) = F'(x)$ is the density function.

The general form of linear unbiased estimates α_1^* and α_2^*, based on the k sample quantiles (12), was given by Ogawa [1]. We are now going to investigate the asymptotic efficiencies of these estimates. Let K_{rs} and e_{rs}, $(r, s = 1, 2)$, be defined by (1) and (2), respectively, with $H_r(\lambda)$, $(r = 1, 2)$, given by (13). Three cases are distinguished.

Case 1. Both α_1 and α_2 unknown.

The joint asymptotic efficiency of α_1^* and α_2^* is

$$E_{12} = \frac{K_{11} K_{22} - K_{12}^2}{e_{11}\, e_{22} - e_{12}^2}.$$

The quantities λ_i which maximize E_{12} will be called the joint optimum spacings for α_1 and α_2, and the corresponding estimates will be termed the joint optimum linear estimates (JOLE) of α_1 and α_2. By application of lemma 1, we have

Theorem 1. *If $H_r(\lambda)$, $(r = 1, 2)$, fulfills conditions A1–A5, then the joint optimum spacings for α_1 and α_2 are given by* (6), *and the joint asymptotic efficiency of the JOLE of α_1 and α_2 is*

$$E_{12} = 1 - \frac{1}{12(k+1)^2} \frac{c^3}{e_{11}\, e_{22} - e_{12}^2} + o(k+1)^{-2}.$$

Cases 2a and 2b. Only one of the two parameters α_r, $(r = 1 \text{ or } 2)$, is unknown; the remaining one is known.

The asymptotic efficiency of α_r^*, $(r = 1 \text{ or } 2)$, is then

$$E_r = \frac{K_{rr}}{e_{rr}}, \quad (r = 1 \text{ or } 2).$$

The quantities λ_i which maximize E_r will be called the optimum spacings for α_r, and the corresponding estimate will be called the optimum linear estimate (OLE) of α_r. From lemma 2 we conclude

Theorem 2. *If $H_r(\lambda)$, $(r = 1 \text{ or } 2)$, fulfills conditions B1–B5, then the optimum spacings for α_r are given by* (9) *and the asymptotic efficiency of the OLE of α_r is*

$$E_r = 1 - \frac{1}{12(k+1)^2} \frac{c_r^3}{e_{rr}} + o(k+1)^{-2}, \quad (r = 1 \text{ or } 2).$$

Under conditions, slightly more rigorous than those needed in proving lemmas 1 and 2, one can also state

Theorem 3. *In order to obtain an asymptotically efficient estimate of* α_1 *(*α_2 *known) or* α_2 *(*α_1 *known), or joint asymptotically efficient estimates of* α_1 *and* α_2*, it is sufficient to utilize only* $k = k(n)$ *out of the n observations available, where, as* $n \to \infty$*,* $k(n) \to \infty$ *in the weakest possible way. (This means that, when* $n \to \infty$*, the relative number of observations* $k(n)/n \to 0$*.)*

For small values of k, optimum spacings were laboriously calculated by Ogawa [2] and Sarhan, Greenberg and Ogawa [3] in the following estimation cases:

(a) normal distribution, α_1 unknown, α_2 known,
(b) normal distribution, α_2 unknown, α_1 known,
(c) exponential distribution, α_2 unknown, α_1 known.

A simplified method will be described here. Comparison of formulas (9) and (11) suggests that an estimate which may be termed the nearly optimum linear estimate (NOLE) of α_r can be based on the spacings (11). These latter will be called the nearly optimum spacings for α_r. Likewise, the nearly joint optimum estimates (NJOLE) of α_1 and α_2 will be defined as the estimates which are based on the nearly joint optimum spacings (8). (Asymptotically, the efficiencies of the NJOLE and the NOLE behave according to the formulas of Theorems 1 and 2, respectively.)

To illustrate the advantage of the NOLE, we choose the estimation cases (a) and (c) mentioned above. In fact, the nearly optimum spacings are:

case (a):

$$\lambda_i = \phi\,[\sqrt{3}\cdot\phi^{-1}(\pi_i)], \quad (i = 1, \ldots, k),$$

where

$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

and ϕ^{-1} its inverse;

case (c):

$$\lambda_i = 1 - (1 - \pi_i)^3, \quad (i = 1, \ldots, k),$$

where π_i is defined by (5).

The easily calculated nearly optimum spacings yield estimates slightly less efficient than the OLE's, a fact which is illustrated by Table 1. (The efficiencies of the OLE are taken from [2] and [3].)

Table 1. Asymptotic efficiency of the OLE as compared to the NOLE.

Estimation case	Asymptotic efficiency			
	$k=5$		$k=10$	
	OLE	NOLE	OLE	NOLE
Normal, α_1 (α_2 known)	0.9420	0.9387	0.9808	0.9800
Exponential, α_2 (α_1 known)	0.9476	0.9448	0.9832	0.9826

Still better results can be obtained, if we change the definition of the quantities π_i into

$$\pi_i = \frac{i-a}{k-a-b+1}, \tag{14}$$

where the constants a and b can be properly chosen in each estimation case.

4. Optimum stratification in proportionate sampling

Consider a population with the frequency function $f(x)$, defined over the interval $A \leqslant x \leqslant B$. The mean of x,

$$m = \int_A^B t f(t) \, dt,$$

is assumed to exist. The population is to be stratified into $(k+1)$ strata by means of the points

$$A = x_0 < x_1 < \cdots < x_k < x_{k+1} = B.$$

If proportionate allocation is used, and we omit the finite population correction factor, then the variance of the usual estimate

$$\bar{x} = \sum_{i=1}^{k+1} p_i \bar{x}_i$$

can be written

$$V(\bar{x}) = \frac{1}{n}\left[\int_A^B (t-m)^2 f(t) \, dt - K_{11}(\lambda_1, \ldots, \lambda_k)\right],$$

where n is the sample size and K_{11} is given by (1) with

$$H_1(\lambda) = \int_A^{G(\lambda)} t f(t) \, dt - m\lambda.$$

Compared to the spacing problem dealt with in the preceding section, the problem of this section is much simpler to handle. To meet conditions B1, B3 and B4, we need only presuppose that $f(x)$ is continuous and limited for $A \leqslant x \leqslant B$, and that the variance of x exists. Under these assumptions we have

Theorem 4. *If conditions B2 and B5 are fulfilled, then the minimum variance of* $\bar{x}$ *is*

$$V(\bar{x})_{\min} = \frac{1}{n}\left[\frac{c_1^3}{12(k+1)^2} + o(k+1)^{-2}\right],$$

the optimum stratification points being

$$x_i = G(\lambda_i), \tag{15}$$

where λ_i *is given by* (9).

For certain distributions and for small values of k, Dalenius [4] and Zindlèr [5] have calculated optimum stratification points. One attempt to get reasonable approximations was briefly outlined by Ekman [6].

Here we define nearly optimum stratification points according to (15) with λ_i given by (11). For the population $f(x) = e^{-x}$, table 2 gives a comparison between optimum and nearly optimum stratification points. These latter are given by

$$x_i = 3 \log \frac{1}{1-\pi_i},$$

and we have chosen to define π_i according to (14) with $a=0$, $b=-0.5$, corrections which are appropriate in this case. (The resulting stratification points may suitably be termed a, b-corrected nearly optimum.)

Table 2. Optimum and a, b-corrected nearly optimum stratification points for $f(x) = e^{-x}$ and the corresponding variances.

Number of strata		Points of stratification			
		x_1	x_2	x_3	$n\,V(\bar{x})$
2	Optimum	1.594			0.3524
	a, b-corrected nearly optimum	1.533			0.3530
3	Optimum	1.018	2.611		0.1797
	a, b-corrected nearly optimum	1.009	2.542		0.1800
4	Optimum	0.754	1.772	3.365	0.1090
	a, b-corrected nearly optimum	0.754	1.763	3.296	0.1091

Sometimes the spacing problem of section 3 and the stratification problem of this section yield the same optimum solutions λ_i (and also the same nearly optimum solutions). These cases can be summarized as follows:

Distribution	Spacing problem which is equivalent to the stratification problem
Normal: $\dfrac{1}{\alpha_2\sqrt{2\pi}}\, e^{-\frac{1}{2}[(z-\alpha_1)/\alpha_2]^2}$	Estimation of α_1; α_2 known
Gamma: $\dfrac{1}{\alpha_2\,\Gamma(p)}\,[(z-\alpha_1)/\alpha_2]^{p-1}\, e^{-[(z-\alpha_1)/\alpha_2]}$	Estimation of α_2; α_1 known.

5. Determination of optimum interval limits in grouping

Here the problem is to group a sample from a certain distribution by means of optimally determined group limits.

Suppose first that the grouped sample is used for estimation purposes. Kulldorff [7], [8], [9], investigated the maximum likelihood estimates of α_1 (α_2 known) and α_2 (α_1 known) in the normal distribution and of α_2 (α_1 known) in the exponential distribution. In these cases, the asymptotic efficiencies of the maximum likelihood estimates are identical with the quantities E_r for the corresponding spacing problem. Therefore, nearly optimum group limits can be derived by use of the technique described in section 3.

Cox [10], using a different approach, assumed that the grouping is done merely for convenience of exposition. By means of a quadratic loss function, the total loss of information due to grouping can be expressed. The problem of minimizing this loss can be directly transferred into the stratification problem of section 4, and the approximation technique used there can be applied for determining nearly optimum group limits.

6. Remarks

A detailed discussion of the results given in this paper will be presented subsequently. Special emphasis will be put on the spacing problem, and attention will also be paid to those frequently occurring estimation cases, where one or more of the conditions imposed in section 2 fail to be fulfilled.

This work was sponsored by the U.S. Army Research Office.

REFERENCES

1. OGAWA, J., Contribution to the theory of systematic statistics, I. Osaka Math. Journal, Vol. 3 (1951).
2. ——, Contribution to the theory of systematic statistics, II. Osaka Math. Journal, Vol. 4 (1952).
3. SARHAN, A. E., GREENBERG, B. G., and OGAWA, J., Simplified estimates for the exponential distribution. Submitted to Ann. Math. Stat.
4. DALENIUS, T., Sampling in Sweden. Stockholm, 1957.
5. ZINDLER, H. J., Über Faustregeln zur optimalen Schichtung bei Normalverteilung. Allgem. Stat. Archiv, *40* (1956).
6. EKMAN, G., Approximate expressions for the conditional mean and variance over small intervals of a continuous distribution. Ann. Math. Stat. *30* (1959).
7. KULLDORFF, G., Maximum likelihood estimation of the mean of a normal random variable when the sample is grouped. Skand. Aktuarietidskrift *41* (1958).
8. ——, Maximum likelihood estimation of the standard deviation of a normal random variable when the sample is grouped. Skand. Aktuarietidskrift *41* (1958).
9. ——, Maximum likelihood estimation for the exponential distribution when the sample is grouped. Mimeographed, Lund, 1958.
10. COX, D. R., Note on grouping. J. Am. Stat. Assoc. *52* (1957).

Tryckt den 17 november 1961

Uppsala 1961. Almqvist & Wiksells Boktryckeri AB