Osaka Mathematical Journal
Vol. 7, No. 1, June, 1955.

On the Correlation of Efficient Estimates of Unknown Parameters

By Junjiro OGAWA and Kusunori KOGO

1. Introduction. Let the density function of the population distribution be $f(x;\alpha)$, which is assumed to be continuous with respect to x. Here α is the unknown parameter to be estimated from the sample. Let $\alpha_1{}^*$ be an efficient estimate in the sense of H. Cramér[1], while $\alpha_2{}^*$ be any regular unbiased estimate of efficiency e. Then, as is well known, the correlation coefficient of $\alpha_1{}^*$ and $\alpha_2{}^*$, $\rho(\alpha_1{}^*, \alpha_2{}^*)$, is equal to $\sqrt{e}$. For the proof of this theorem, use has been made usually of the fact that the linear combination of regular estimates is also regular[2]. But we have not yet succeeded in proving this, and the proof has never been published, at least as far as authors know. Therefore, it seems necessary to prove the above theorem in a different manner. If we apply our method to the case of several unknown parameters, we can derive the similar formula concerning ρ and e to that in the case of a single unknown parameter.

2. The case of a single unknown parameter.

The probability element of the joint distribution of the random sample $x_1, x_2, \cdots, x_n$ drawn from the population with the density function $f(x;\alpha)$ is

$$f(x_1;\alpha)f(x_2;\alpha)\cdots f(x_n;\alpha)dx_1\,dx_2\cdots dx_n\,.$$

If $\alpha^*(x_1, x_2,\cdots, x_n)$ is a regular unbiased estimate of α, we can select auxiliary variables $\xi_1, \cdots, \xi_{n-1}$ such that by the transformation

$$\left.\begin{aligned} \alpha^* &= \alpha^*(x_1, x_2, \cdots, x_n) \\ \xi_1 &= \xi_1(x_1, x_2, \cdots, x_n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \xi_{n-1} &= \xi_{n-1}(x_1, x_2, \cdots, x_n) \end{aligned}\right\}, \tag{2. 1}$$

the sample space W can be mapped to the $(\alpha^*, \xi_1, \cdots, \xi_{n-1})$-space in 1–1 correspondence almost everywhere. Hence the Jacobian

$$\frac{\partial(\alpha^*, \xi_1, \cdots, \xi_{n-1})}{\partial(x_1, x_2, \cdots, x_n)}$$

does not vanish almost everywhere. Therefore the Jacobian

$$J=\frac{\partial(x_1, x_2, \cdots, x_n)}{\partial(\alpha^*, \xi_1, \cdots, \xi_{n-1})}$$

remains finite almost everywhere in the $(\alpha^*, \xi_1, \cdots, \xi_{n-1})$-space. Thus we have

$$\begin{aligned} &f(x_1 ; \alpha)f(x_2 ; \alpha) \cdots f(x_n ; \alpha)|J| d\alpha^* d\xi_1 \cdots d\xi_{n-1} \\ &\quad = g(\alpha^* ; \alpha) d\alpha^* h(\xi_1, \cdots, \xi_{n-1} | \alpha^* ; \alpha) d\xi_1 \cdots d\xi_{n-1}, \end{aligned} \tag{2.2}$$

where $g(\alpha^* ; \alpha)$ is the density function of the marginal distribution of α^* and $h(\xi_1, \cdots, \xi_{n-1} | \alpha^* ; \alpha)$ is the density function of the conditional distribution of $\xi_1, \cdots, \xi_{n-1}$ given α^*. From (2.2) we get

$$\frac{\partial \log g}{\partial \alpha}+\frac{\partial \log h}{\partial \alpha}=\sum_{i=1}^{n} \frac{\partial \log f(x_i ; \alpha)}{\partial \alpha}. \tag{2.3}$$

For an efficient estimate α_0^* of α whose density function is $g_0(\alpha_0^* ; \alpha)$, (2.3) becomes

$$\frac{\partial \log g_0}{\partial \alpha}=\sum_{i=1}^{n} \frac{\partial \log f(x_i ; \alpha)}{\partial \alpha}, \tag{2.4}$$

and by the condition of efficiency we have

$$\frac{\partial \log g_0}{\partial \alpha}=k\cdot(\alpha_0^*-\alpha), \tag{2.5}$$

where k is independent of $x_1, x_2, \cdots, x_n$. From (2.3), (2.4) and (2.5) we obtain

$$k\cdot(\alpha_0^*-\alpha)=\frac{\partial \log g}{\partial \alpha}+\frac{\partial \log h}{\partial \alpha}. \tag{2.6}$$

Multiplying both sides of (2.6) by $(\alpha^*-\alpha)$ and taking expectation, we have

$$k\cdot D(\alpha_0^*)\cdot D(\alpha^*)\cdot \rho(\alpha_0^*, \alpha^*)=1. \tag{2.7}$$

On the other hand, multiply (2.5) by $(\alpha_0^*-\alpha)$ and take its expectations, then it turns out to be

$$k\cdot D^2(\alpha_0^*)=1. \tag{2.8}$$

From (2.7) and (2.8), it follows that

$$\rho(\alpha_0^*, \alpha^*)=D(\alpha_0^*)/D(\alpha^*)=\sqrt{e(\alpha^*)}, \tag{2.9}$$

as was to be proved.

3. The case of several unknown parameters.

Let the density function of the population distribution be $f(x;\alpha)$, which is assumed to be continuous with respect to x, and

$$\boldsymbol{\alpha}=\begin{pmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_k\end{pmatrix},\qquad k\geqq 2,$$

is a vector of unknown parameters to be estimated from the sample. Let $x_1, x_2, \cdots, x_n$ be a random sample of size n drawn from this parent population. Then the probability element of the joint distribution of the sample is

$$f(x_1;\boldsymbol{\alpha})f(x_2;\boldsymbol{\alpha})\cdots f(x_n;\boldsymbol{\alpha})dx_1dx_2\cdots dx_n.$$

We shall assume now on that there exists an estimate $\boldsymbol{\alpha}^*$ for $\boldsymbol{\alpha}$ such that together with properly selected auxiliary variables $\xi_1, \cdots, \xi_{n-k}$ the transformation

$$\left.\begin{aligned}\boldsymbol{\alpha}^* &= \boldsymbol{\alpha}^*(x_1,\cdots,x_n)\\ \xi_1 &= \xi_1(x_1,\cdots,x_n)\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ \xi_{n-k} &= \xi_{n-k}(x_1,\cdots,x_n)\end{aligned}\right\}\tag{3.1}$$

maps the sample space on the $(\alpha_1^*, \cdots, \alpha_k^*, \xi_1, \cdots, \xi_{n-k})$–space in 1–1 correspondence almost everywhere and the regularity conditions on the marginal frequency function $g(\boldsymbol{\alpha}^*;\boldsymbol{\alpha})$ and the conditional frequency function $h(\xi_1, \cdots, \xi_{n-k}|\boldsymbol{\alpha}^*;\boldsymbol{\alpha})$ remain valid.

Then we have

$$f(x_1;\boldsymbol{\alpha})\cdots f(x_n;\boldsymbol{\alpha})|J| = g(\boldsymbol{\alpha}^*;\boldsymbol{\alpha})\cdot h(\xi_1,\cdots,\xi_{n-k}|\boldsymbol{\alpha}^*;\boldsymbol{\alpha})\tag{3.2}$$

almost everywhere in $(\boldsymbol{\alpha}^*;\xi_1,\cdots,\xi_{n-k})$–space and

$$J=\frac{\partial(x_1, x_2, \cdots, x_k, x_{k+1}, \cdots, x_n)}{\partial(\alpha_1^*, \alpha_2^*, \cdots, \alpha_k^*, \xi_1, \cdots, \xi_{n-k})}$$

does not vanish almost everywhere.

Using the vector notation

$$\frac{\partial \log f(x\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \begin{pmatrix} \dfrac{\partial \log f(x\,;\boldsymbol{\alpha})}{\partial \alpha_1} \\ \vdots \\ \dfrac{\partial \log f(x\,;\boldsymbol{\alpha})}{\partial \alpha_k} \end{pmatrix}, \quad \text{etc.,}$$

we obtain the following equation from (3.2):

$$\sum_{i=i}^{n} \frac{\partial \log f(x_i\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{\partial \log g(\boldsymbol{\alpha}^*\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} + \frac{\partial \log h(\boldsymbol{\xi}\,|\,\boldsymbol{\alpha}^*\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}. \tag{3.3}$$

Hence it follows that

$$\begin{aligned} &\left(\sum_{i=1}^{n} \frac{\partial \log f(x_i\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right)\cdot\left(\sum_{i=1}^{n} \frac{\partial \log f(x_i\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right)' \\ &= \left(\frac{\partial \log g}{\partial \boldsymbol{\alpha}}\right)\left(\frac{\partial \log g}{\partial \boldsymbol{\alpha}}\right)' + \left(\frac{\partial \log g}{\partial \boldsymbol{\alpha}}\right)\left(\frac{\partial \log h}{\partial \boldsymbol{\alpha}}\right)' + \left(\frac{\partial \log h}{\partial \boldsymbol{\alpha}}\right)\left(\frac{\partial \log g}{\partial \boldsymbol{\alpha}}\right)' \\ &\quad + \left(\frac{\partial \log h}{\partial \boldsymbol{\alpha}}\right)\left(\frac{\partial \log h}{\partial \boldsymbol{\alpha}}\right)'. \end{aligned} \tag{3.4}$$

Now taking expectations of both sides of (3.4), we have

$$n\boldsymbol{F} = \boldsymbol{G} + \boldsymbol{H}, \tag{3.5}$$

where

$$\boldsymbol{F} = E\left[\left(\frac{\partial \log f}{\partial \boldsymbol{\alpha}}\right)\cdot\left(\frac{\partial \log f}{\partial \boldsymbol{\alpha}}\right)'\right],$$

$$\boldsymbol{G} = E\left[\left(\frac{\partial \log g}{\partial \boldsymbol{\alpha}}\right)\cdot\left(\frac{\partial \log g}{\partial \boldsymbol{\alpha}}\right)'\right]$$

and

$$\boldsymbol{H} = E\left[\left(\frac{\partial \log h}{\partial \boldsymbol{\alpha}}\right)\cdot\left(\frac{\partial \log h}{\partial \boldsymbol{\alpha}}\right)'\right].$$

As is easily seen, the matrix $\boldsymbol{H}$ is non-negative, so

$$n\boldsymbol{F} - \boldsymbol{G} \tag{3.6}$$

is non-negative.

Lemma 1. *For two random k-vector variables $\boldsymbol{x}$ and $\boldsymbol{y}$ with means $\boldsymbol{O}$, the following statements hold:*

$$\boldsymbol{\Lambda}_{yy} - \boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy} \quad \textit{is non-negative,} \tag{3.7}$$

where $\boldsymbol{\Lambda}_{xx}$, $\boldsymbol{\Lambda}_{yy}$, $\boldsymbol{\Lambda}_{xy}$, $\boldsymbol{\Lambda}_{yx}$ are the variances and covariances matrices, and the equality

$$\boldsymbol{\Lambda}_{yy} = \boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy} \tag{3.8}$$

holds when and only when

$$\boldsymbol{y}=\boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\cdot\boldsymbol{x}\,. \tag{3.9}$$

Proof: For any constant k–vector $\mathfrak{x}$, we shall form an quadratic form

$$Q=\mathfrak{x}'\cdot\boldsymbol{M}\cdot\mathfrak{x}\,,$$

where

$$\boldsymbol{M}=E[(\boldsymbol{y}-\boldsymbol{\beta}\boldsymbol{x}-\boldsymbol{\gamma})(\boldsymbol{y}-\boldsymbol{\beta}\boldsymbol{x}-\boldsymbol{\gamma}')]\,,$$

and minimize it with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, where

$$\boldsymbol{\beta}=\begin{pmatrix}\beta_{11}\,\beta_{12}\cdots\beta_{1k}\\ \vdots\quad\vdots\qquad\vdots\\ \beta_{k1}\beta_{k2}\cdots\beta_{kk},\end{pmatrix},\quad \boldsymbol{\gamma}=\begin{pmatrix}\gamma_1\\ \vdots\\ \gamma_k\end{pmatrix}.$$

$$-\frac{1}{2}\frac{\partial Q}{\partial\beta_{ij}}=-\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\beta}_{ij}}E\{\mathfrak{x}'(\boldsymbol{y}-\boldsymbol{\beta}\boldsymbol{x}-\boldsymbol{\gamma})\}^2{}_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}},\ \boldsymbol{\gamma}=\hat{\boldsymbol{\gamma}}}$$

$$=\mathfrak{x}'\cdot E[(\boldsymbol{y}-\hat{\boldsymbol{\beta}}\boldsymbol{x}-\hat{\gamma})x_j]_{\mathfrak{x}_i}=0,\quad i,j=1,2,\cdots,k\,.$$

Hence we have

$$\mathfrak{x}'\cdot E[(\boldsymbol{y}-\hat{\boldsymbol{\beta}}\cdot\boldsymbol{x}-\hat{\gamma})\boldsymbol{x}']=\boldsymbol{O}$$

or

$$\mathfrak{x}'(\boldsymbol{\Lambda}_{yx}-\hat{\boldsymbol{\beta}}\cdot\boldsymbol{\Lambda}_{xx})=\boldsymbol{O}\,. \tag{3.10}$$

Since this must have been valid for any constant k–vector $\mathfrak{x}$, we have

$$\hat{\boldsymbol{\beta}}=\boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\,. \tag{3.11}$$

By similar arguments we have

$$\hat{\gamma}=\boldsymbol{O}. \tag{3.12}$$

Therefore the minimum value of Q for any $\mathfrak{x}$ is

$$\mathfrak{x}'(\boldsymbol{\Lambda}_{yy}-\boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy})\mathfrak{x}\geqq 0\,,$$

whence (3.7) follows.

The minimum value Q_0 of Q is

$$Q_0=\mathfrak{x}'\cdot\hat{\boldsymbol{M}}\cdot\mathfrak{x}=E\{\mathfrak{x}'(\boldsymbol{y}-\hat{\boldsymbol{\beta}}\cdot\boldsymbol{x})\}^2$$

$$=\mathfrak{x}'\cdot(\boldsymbol{\Lambda}_{yy}-\boldsymbol{\Lambda}_{yx}\boldsymbol{\Lambda}_{xx}^{-1}\boldsymbol{\Lambda}_{xy})\cdot\mathfrak{x}\,, \tag{3.13}$$

therefore the validity of the equality (3.8) is equivalent to the following statement; i.e. for all constant vector $\mathfrak{x}$

$$\mathfrak{x}'\cdot(\boldsymbol{y}-\hat{\boldsymbol{\beta}}\cdot\boldsymbol{x})=\boldsymbol{O}\,,$$

that is

$$y = \hat{\beta} \cdot x ,$$

which is the same as (3. 9) by the relation (3. 11) .

Theorem 3. 1. *If* α^* *is a regular unbiased estimator of* α *and* $g(\alpha^* ; \alpha)$ *is the marginal joint frequency function of* α^*, *then the matrix* $G - \Lambda^{-1}$ *is non-negative, and the equality*

$$G = \Lambda^{-1} \tag{3. 14}$$

holds, when and only when

$$\frac{\partial \log g}{\partial \alpha} = \Lambda^{-1} \cdot (\alpha^* - \alpha) . \tag{3. 15}$$

Here the matrix Λ *denotes the moment matrix of* α^* .

Proof : Put $y = \frac{\partial \log g}{\partial \alpha}$ and $x = \alpha^* - \alpha$. in Lemma 1, then

$$E(x) = E(y) = O$$

$$\Lambda_{yy} = G, \ \Lambda_{xx} = \Lambda, \ \Lambda_{yx} = \Lambda_{xy} = I .$$

therefore we have

$$\Lambda_{yx} \Lambda_{xx}^{-1} \Lambda_{xy} = \Lambda^{-1}.$$

Hence this theorem is an immediate consequence of Lemma 1.

From Theorem 3. 1 it follows that the following three statements are equivalent ; i. e.

(i) α^* is an efficient estimate of α in Cramér's sense,

(ii) $nF = G = \Lambda^{-1}$,

(iii) α^* is a sufficient estimate of α and

$$\frac{\partial \log g}{\partial \alpha} = \Lambda^{-1} \cdot (\alpha^* - \alpha) .$$

Theorem 3. 2. *Let* α_0^* *be an efficient estimate and* α^* *be a regular unbiased estimate of* α, *and further let it be*

$$\Lambda_{00} = E[(\alpha_0^* - \alpha)(\alpha_0^* - \alpha)'], \ \Lambda_{01} = E[(\alpha_0^* - \alpha)(\alpha^* - \alpha)'] ,$$
$$\Lambda_{10} = E[(\alpha^* - \alpha)(\alpha_0^* - \alpha)'] .$$

Then we get

$$\Lambda_{00} = \Lambda_{01} = \Lambda_{10} . \tag{3. 16}$$

Proof : Following the previous notation we have

$$\sum_{i=1}^{n} \frac{\partial \log f(x_i\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{\partial \log g_0(\boldsymbol{\alpha}_0^*\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$$

and

$$\sum_{i=1}^{n} \frac{\partial \log f(x_i\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{\partial \log g(\boldsymbol{\alpha}^*\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} + \frac{\partial \log h(\boldsymbol{\xi}\,|\,\boldsymbol{\alpha}^*\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\,.$$

Since α_0^* is an efficient estimate, we have

$$\frac{\partial \log g_0(\boldsymbol{\alpha}_0^*\,;\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \boldsymbol{\Lambda}_{00}^{-1}\cdot(\boldsymbol{\alpha}_0^* - \boldsymbol{\alpha})\,.$$

From the above three equations we obtain

$$\frac{\partial \log g}{\partial \boldsymbol{\alpha}} + \frac{\partial \log h}{\partial \boldsymbol{\alpha}} = \boldsymbol{\Lambda}_{00}^{-1}\cdot(\alpha_0^* - \boldsymbol{\alpha})\,. \qquad (3.17)$$

Multiplying both sides of (3.17) by $(\boldsymbol{\alpha}^* - \boldsymbol{\alpha})'$ from the right and taking expectations, we get

$$\boldsymbol{\Lambda}_{00}^{-1}\cdot\boldsymbol{\Lambda}_{01} = \mathbf{I}\,,$$

whence we have

$$\boldsymbol{\Lambda}_{00} = \boldsymbol{\Lambda}_{01} = \boldsymbol{\Lambda}_{10}\,.$$

The efficiency $e(\alpha^*)$ of the regular unbiased estimate α^* is given by the equation

$$e(\boldsymbol{\alpha}^*) = \Lambda_{00}/\Lambda_{11} = 1/n^k\Lambda_{11}F\,, \qquad (3.18)$$

If we adopt the Dietzius' definition[3)] of the correlation coefficient between two k-dimensional random vectors, and put

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \Lambda_{xy}/\sqrt{\Lambda_{xx}\Lambda_{yy}}\,, \qquad (3.19)$$

where Λ_{ij} is the determinant of $\boldsymbol{\Lambda}_{ij}$.

Then we shall have the following generalization of the result stated in section 2.

Theorem 3.3.

$$\rho(\alpha_0^*, \boldsymbol{\alpha}^*) = \sqrt{e(\boldsymbol{\alpha}^*)}\,. \qquad (3.20)$$

If we denote the regression matrix of $\boldsymbol{y}$ *on* $\boldsymbol{x}$ *by* $\boldsymbol{\beta}(\boldsymbol{y}, \boldsymbol{x})$ *and by* $\beta(\boldsymbol{y}, \boldsymbol{x})$ *its determinant, we have*

$$\beta(\boldsymbol{\alpha}_0^*\,;\alpha^*) = e(\alpha^*),\ \beta(\boldsymbol{\alpha}^*\,;\boldsymbol{\alpha}^*) = 1\,. \qquad (3.21)$$

The proof of this theorem is easy and will be omitted.

(Received March 14, 1955)

Notes and References

[1] H. Cramér: Mathematical Methods of Statistics, 1946, Princeton Univ. Press, Chap. 32, 473–497.

[2] H. Cramér loc. cit., p. 482.

[3] R. Dietzius: Ausdehnung der Korrelationsmethode und der Methode der kleinsten Quadrate auf Vektoren, Wien, Berichte. Abt. II. **125** (1916), 3.