Title	Contributions to the theory of systematic statistics. I
Author(s)	Ogawa, Junjiro
Citation	Osaka Mathematical Journal. 1951, 3(2), p. 175-213
Version Type	VoR
URL	https://doi.org/10.18910/5329
rights	
Note	

Osaka Mathematical Journal,
Vol. 3, No. 2, December, 1951

Contributions to the Theory of Systematic Statistics, I.

By Junjiro OGAWA

CONTENTS

Introduction

Our main concerns in the theory of mathematical statistics have been "efficient estimates" and "most powerful tests".[1] But from the point of view of economy, it seems reasonable to inquire whether the output of information is comparable in value to the input measured in money, man-hours, or others. Alternatively we may inquire whether comparable results could have been obtained by smaller expenditures.

Recently Dr. Frederick Mosteller[2] has proposed the use of systematic statistics for such purposes, basing on the fact that, however large the sample size is, all individuals of the sample are easily (with low costs and quickly) ordered by punched-card equipment. F. Mosteller considered the estimations of the mean and standard deviation of an univariate normal population and the estimation of the coefficient of correlation of a bivariate normal population. After that Prof. Ziro Yamanouchi[3] contributed greatly in the former case. Yamanouchi's results are, from the point of view of our present stage, essentially that best linear unbiased estimates are considerably more efficient than those used by F. Mosteller.

The purposes of this paper are to extend the results already obtained by F. Mosteller and Z. Yamanouchi on the basis of the general theory of statistical estimation (Chapter I) on one hand, and on the other hand, to develop the theory of testing statistical hypotheses concerning unknown parameters of an univariate normal population in the case of systematic statistics (Chapter II).

Soon later, in a separate paper, we shall deal with problems concerning Dosage-Mortality Curves and Time-Mortality Curves[4] as applications of the theory here developed.

The author expresses here his hearty thanks to Mr. Y. Miyamoto, and Mr. F. Maruyama who born part in the computations of Tables 6.1-6.6, and to Prof. Z. Yamanouchi and others for their valuable advices. And the author was also much indebted to Prof. K. Kunugi for the encouragement during the time when he was preparing the manuscript.

Chapter I. Theory of Estimation

In this chapter we shall deal with estimation of parameters specifying a normal population by means of systematic statistics. In § 1, § 2, we summarize the necessary results from the general theory of statistical estimation[5].

§ 1. Regular unbiased estimates. For the sake of simplicity of explanations, we assume that the parent population under consideration is of the continuous type, i.e., the distribution of the population has the density function which is continuous almost everywhere. The reasonings given below will, of course, be valid for populations of the discrete type, provided the necessary modifications being made.

We shall consider here two cases when the number of parameters to be estimated is one and two. The arguments for cases when the parameters to be estimated are more than two are essentially the same as those for cases when unknown parameters two, but we shall not need such cases for the time being.

Case I. The case when the number of parameters to be estimated is one. Let the frequency function of the population under consideration be $f(x;\alpha)$, where the functional form of f is assumed to be known, and α the unknown parameter to be estimated, the region of consideration of α being a certain non-degenerate interval A.

Let $x_1, x_2, \dots, x_n$ be a random sample of size n drawn from the population. The statistic

$$\alpha^* = \alpha^*(x_1, x_2, \dots, x_n) ,$$

which is Borel-measurable and independent of α, is called a *regular estimate* of α, when it satisfies the following conditions 1 and 2.

Condition 1. We can choose a new system of variables $\xi_1, \xi_2, \dots, \xi_{n-1}$ such that the transformation of variables

$$\begin{aligned} \alpha^* &= \alpha^* (x_1, \dots, x_n) \\ \xi_1 &= \xi_1 \ (x_1, \dots, x_n) \\ &\cdots\cdots\cdots\cdots \\ \xi_{n-1} &= \xi_{n-1} (x_1, \dots, x_n) \end{aligned} \tag{1.1}$$

has the following properties:

(1a) $\alpha^*(x_1, \dots, x_n)$, $\xi_1(x_1, \dots, x_n), \dots, \xi_{n-1}(x_1, \dots, x_n)$ are one-valued and continuous functions of $x_1, x_2, \dots, x_n$ everywhere in the x-space, and have continuous partial derivatives

$$\frac{\partial \alpha^*}{\partial x_i}, \ \frac{\partial \xi_j}{\partial x_i}, \quad i = 1, 2, \dots, n \, ; \ j = 1, 2, \dots, n-1$$

in all points, except possibly in certain points belonging to a finite number of hypersurfaces.

(1b) The transformation (1.1) defines one-to-one correspondence between the points $(x_1, \dots, x_n)$ and $(\alpha^*, \xi_1, \dots, \xi_{n-1})$.

Remarks: Consider a point $(x_1, \dots, x_n)$ which does not belong to any of the exceptional hypersurfaces, and is such that the Jacobian $\frac{\partial(\alpha^*, \xi_1, \dots, \xi_{n-1})}{\partial(x_1, x_2, \dots, x_n)}$ is non-vanishing. The Jacobin of the inverse transformation $J = \frac{\partial(x_1, x_2, \dots, x_n)}{(\alpha^*, \xi_1, \dots, \xi_{n-1})}$ is then finite in the point $(\alpha^*, \xi_1, \dots, \xi_{n-1})$ corresponding to the point $(x_1, \dots, x_n)$, since we have

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

Therefore, it follows that

$$f(x_1 \, ; \alpha) \dots f(x_n \, ; \alpha) \, dx_1 \dots dx_n = f(x_1 \, ; \alpha) \dots f(x_n \, ; \alpha) \cdot |J| \, d\alpha^* d\xi_1 \dots d\xi_{n-1} \tag{1.2}$$

where in the right-hand side of (1.2), $x_1, \dots, x_n$ should be represented in terms of $\alpha^*, \xi_1, \dots, \xi_{n-1}$.

Now, let the frequency functions of the marginal distribution of α^*

and the conditional distribution of $\xi_1, \dots, \xi_{n-1}$ given α^* be $g(\alpha^*; \alpha)$ and $h(\xi_1, \dots, \xi_{n-1} | \alpha^*; \alpha)$ respectively, then we have

$$f(x_1; \alpha) \dots f(x_n, \alpha) |J| d\alpha^* d\xi_1 \dots d\xi_{n-1} = g(\alpha^*; \alpha)\, d\alpha^* h(\xi_1, \dots, \xi_{n-1} | \alpha^*; \alpha)\, d\xi_1 \dots d\xi_{n-1} \quad (1.3)$$

Condition 2. For almost all values of $(x_1, \dots, x_n); \alpha^*, \xi_1, \dots, \xi_{n-1}$, there exist partial derivatives

$$\frac{\partial f}{\partial \alpha}, \quad \frac{\partial g}{\partial \alpha}, \quad \frac{\partial h}{\partial \alpha}$$

in each point α in A, and the following relations

$$\left|\frac{\partial f}{\partial \alpha}\right| < F_0(x), \quad \left|\frac{\partial g}{\partial \alpha}\right| < G_0(\alpha^*), \quad \left|\frac{\partial h}{\partial \alpha}\right| < H_0(\xi_1, \dots, \xi_{n-1} | \alpha^*) \quad (1.4)$$

hold, where $F_0(x)$, $G_0(\alpha^*)$, $H_0(\xi_1, \dots, \xi_{n-1} | \alpha^*)$ are all independent of α and

$$F_0(x), \ G_0(\alpha^*), \ \alpha^* G_0(\alpha^*), \ H_0(\xi_1, \dots, \xi_{n-1} | \alpha^*) \quad (1.5)$$

are integrable in the whole space of of their variables.

Remark: By the Condition 2, we can interchange integration and differentiation, when it is necessary, for example,

$$\frac{\partial}{\partial \alpha} \int_{-\infty}^{\infty} f(x; \alpha) dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial \alpha} f(x; \alpha) dx .$$

In the sequel, we consider only such estimates as regular, so we shall omit the adjective "regular".

If an estimate α^* of α has the property, that

$$\int_{-\infty}^{\infty} \alpha^* g(\alpha^*; \alpha) d\alpha^* = \alpha \quad (1.6)$$

holds for all values of α in A, then we call α^* the *unbiased estimate* of α.

Differentiating both sides of (1.6) with respect to α, we have

$$\int_{-\infty}^{\infty} \alpha^* \frac{\partial g(\alpha^*; \alpha)}{\partial \alpha} d\alpha^* = 1 . \quad (1.7)$$

Since, of course,

$$\int_{-\infty}^{\infty} g(\alpha^*; \alpha) d\alpha^* = 1 ,$$

it follows that

$$\int_{-\infty}^{\infty} \frac{\partial g(\alpha^*; \alpha)}{\partial \alpha} d\alpha^* = 0 , \quad (1.8)$$

From (1.7) and (1.8), we have

$$\int_{-\infty}^{\infty}(\alpha^*-\alpha)\frac{\partial g(\alpha^*;\alpha)}{\partial\alpha}d\alpha^*=1\ ,$$

which will be rearranged as follows:

$$\int_{-\infty}^{\infty}(\alpha^*-\alpha)\sqrt{g(\alpha^*;\alpha)}\times\frac{1}{\sqrt{g(\alpha^*;\alpha)}}\frac{\partial g(\alpha^*;\alpha)}{\partial\alpha}d\alpha^*=1\ , \tag{1.9}$$

assuming that $g(\alpha^*;\alpha)>0$ for all values of α^*, when otherwise the interval of integration should be chosen such that $g(\alpha^*;\alpha)>0$. By means of Schwarz's inequality, we have

$$D^2(\alpha^*)\cdot E\left(\frac{\partial\log g(\alpha^*;\alpha)}{\partial\alpha}\right)^2\geqq 1\ ,$$

whence it follows that

$$D^2(\alpha^*)\geqq\frac{1}{E\left(\frac{\partial\log g(\alpha^*;\alpha)}{\partial\alpha}\right)^2}\ . \tag{1.10}$$

The equality sign holds when and only when

$$\frac{\partial g(\alpha^*;\alpha)}{\partial\alpha}=k(\alpha^*-\alpha)\ , \tag{1.11}$$

where k is a constant independent of $x_1, \ldots, x_n$.

From (1.2) and (1.3), we have after some easy calculations

$$nE\left(\frac{\partial\log f(x;\alpha)}{\partial\alpha}\right)^2=E\left(\frac{\partial\log g}{\partial\alpha}\right)^2+E\left(\frac{\partial\log h}{\partial\alpha}\right)^2. \tag{1.12}$$

Hence, we have

$$D^2(\alpha^*)\geqq\frac{1}{nE\left(\frac{\partial\log f}{\partial\alpha}\right)^2}\ , \tag{1.13}$$

where the equality sign holds when and only when (1.11) and further

$$\frac{\partial h}{\partial\alpha}=0 \tag{1.14}$$

hold, i.e., $h(\xi, \ldots, \xi_{n-1}\,|\,\alpha^*;\alpha)$ is independent of α.

When the equality in (1.13) holds, we call α^*, after H. Cramér, the *efficient estimate* of α. For any unbiased estimate α^* of α, the quantity

$$e(\alpha^*)=\frac{1}{D^2(\alpha^*)\,nE\left(\frac{\partial\log f}{\partial\alpha}\right)^2} \tag{1.15}$$

lies between 0 and 1. We shall define the *efficiency* of the unbiased estimate α^* be $e(\alpha^*)$ of (1.15). Of course, for efficient estimate α^* of α,

$$e(\alpha^*)=1\ .$$

Now, let us suppose $\alpha^*=\alpha^*(x_1,\dots,x_n)$ is defined for all sufficiently large values of n, and α^* converges in probability to α as n tends to infinity, i. e., α^* is a *consistent estimate* of α. In many important cases the standard deviation of the estimate α^* is of order $n^{-1/2}$ for large n, so that we have $D(\alpha^*)\propto cn^{-1/2}$, where c is a constant independent of n. In such cases we define *asymptotic efficiency* of α^* by

$$e_0(\alpha^*)=\frac{1}{c^2E\left(\dfrac{\partial\log f}{\partial\alpha}\right)^2}\ , \tag{1.16}$$

and the estimate for which $e_0(\alpha^*)=1$ is called an *asymptotically efficient estimate*.

The quantity

$$E\left(\frac{\partial\log f}{\partial\alpha}\right)^2$$

is named by R. A. Fisher an *intrinsic accuracy* of the pupulation and $nE\left(\dfrac{\partial\log f}{\partial\alpha}\right)^2$ *amount of information* of the sample given[6], and these give the upper bound of information, in a certain sense, which the sample will possibly offer.

If we can differentiate $\displaystyle\int_{-\infty}^{\infty}\frac{\partial f}{\partial\alpha}dx=0$ under the integral sign with respect to α, we have

$$E\left(\frac{\partial\log f}{\partial\alpha}\right)^2=-E\left(\frac{\partial^2\log f}{\partial\alpha^2}\right)\ . \tag{1.17}$$

We shall mention two examples which will be necessary in the following.

Example 1. For the normal population with unknown mean m and unit variance, as is well known, the frequency function is

$$f(x\,;m)=(2\pi)^{-1/2}\exp\left\{-(x-m)^2/2\right\},$$

When we take the sample mean $\bar{x}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$ as an estimate of m, it is easily seen that regularity conditions are satisfied. Since

$$\frac{\partial\log f(x\,;m)}{\partial m}=x-m\ ,$$

the amount of information is

$$nE\left(\frac{\partial\log f(x\,;m)}{\partial m}\right)^2=n\ . \tag{1.18}$$

Hence the lower bound of variances of any unbiased estimate of m is $1/n$. Consequently, the sample mean $\bar{x}$ is an efficient estimate of m.

Example 2. For the normal population with 0 mean and unknown variance σ, the frequency function is

$$f(x\,;\sigma) = (2\pi\sigma^2)^{-1/2} \exp\left\{-x^2/2\sigma^2\right\},$$

Then, we have

$$\frac{\partial \log f(x\,;\sigma)}{\partial\sigma} = -\frac{1}{\sigma} + \frac{x^2}{\sigma^3},$$

hence the amount of information is

$$nE\left(\frac{\partial \log f(x\,;\sigma)}{\partial\sigma}\right)^2 = \frac{2n}{\sigma^2}. \tag{1.19}$$

The lower bound of variances of any unbiased estimate of σ is $2n/\sigma^2$.

If we take the statistic

$$\sigma^* = \sqrt{\frac{n}{2}}\,\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}\,s_0\,, \quad \text{where} \quad s_0^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2\,,$$

then the variance of σ^* is for large values of n

$$D^2(\sigma^*) = \left(\frac{n}{2}\,\frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n+1}{2}\right)} - 1\right)\sigma^2 = \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right),$$

hence, σ^* is an asymptotically efficient estimate of σ.

Case II. The case when the number of parameters to be estimated is two. Let the frequency function of the population be $f(x\,;\alpha,\beta)$, where the functional form of f is assumed to be known and α and β are the unknown parameters to be estimated. The regularity conditions for a pair of estimates $\alpha^*(x_1, \dots, x_n)$ and $\beta^*(x_1, \dots, x_n)$ of α and β are as follows:

Condition 1. We can choose a new system of variables $\xi_1, \dots, \xi_{n-2}$ such that the transformation of variables

$$\begin{aligned} \alpha^* &= \alpha^*(x_1, \dots, x_n) \\ \beta^* &= \beta^*(x_1, \dots, x_n) \\ \xi_1 &= \xi_1(x_1, \dots, x_n) \\ &\cdots\cdots\cdots\cdots \\ \xi_{n-2} &= \xi_{n-2}(x_1, \dots, x_n) \end{aligned} \tag{1.20}$$

have the following properties:

(1a) $\alpha^*(x_1, \dots, x_n)$, $\beta^*(x_1, \dots, x_n)$, $\xi_1(x_1, \dots, x_n)$, $\dots$, $\xi_{n-2}(x_1, \dots, x_n)$ are one-valued and continuous functions of $x_1, \dots, x_n$ everywhere in the x-space, and have continuous partial derivatives

$$\frac{\partial\alpha^*}{\partial x_i}, \ \frac{\partial\beta^*}{\partial x_i}, \ \frac{\partial\xi_j}{\partial x_i}; \quad i=1,2,\dots,n; \ j=1,2,\dots,n-2,$$

for all values of $x_1, \dots, x_n$, except possibly in certain points belonging to a finite number of hypersurfaces.

(1b) The transformation (1.20) defines one-to-one correspondence between the points $(x_1, \dots, x_n)$ and $(\alpha^*, \beta^*, \xi_1, \dots, \xi_{n-2})$.

Remark: The Jacobian

$$J=\frac{\partial(x_1, x_2, x_3, \dots, x_n)}{\partial(\alpha^*, \beta^*, \xi_1, \dots, \xi_{n-2})}$$

is non-vanishing almost everywhere, hence we have

$$\prod_{i=1}^{n} f(x_i;\alpha,\beta)\cdot dx_1 \dots dx_n=\prod_{i=1}^{n} f(x_i;\alpha,\beta)\cdot|J|\cdot d\alpha^* d\beta^* d\xi_1 \dots d\xi_{n-2}, \quad (1.21)$$

and

$$\prod_{i=1}^{n} f(x_i;\alpha,\beta)\cdot|J|\cdot d\alpha^* d\beta^* d\xi_1 \dots d\xi_{n-2}=g(\alpha^*, \beta^*;\alpha,\beta)d\alpha^* d\beta^* \times$$
$$h(\xi_1, \dots, \xi_{n-2}\mid\alpha^*, \beta^*;\alpha,\beta)\, d\xi_1 \dots d\xi_{n-2}, \quad (1.22)$$

where $g(\alpha^*, \beta^*;\alpha,\beta)$ is the frequency function of the marginal distribution of α^* and β^* and $h(\xi_1, \dots, \xi_{n-2}\mid\alpha^*, \beta^*;\alpha,\beta)$ is the conditional frequency function of $\xi_1, \dots, \xi_{n-2}$ with given α^*, β^*.

Condition 2. For almost all values of $(x_1, \dots, x_n)$, $\alpha^*, \beta^*, \xi_1, \dots, \xi_{n-2}$ in each point (α, β) of the region of parameters, exist partial derivatives

$$\frac{\partial f}{\partial\alpha}, \frac{\partial f}{\partial\beta}; \ \frac{\partial g}{\partial\alpha}, \frac{\partial g}{\partial\beta}; \ \frac{\partial h}{\partial\alpha}, \frac{\partial h}{\partial\beta},$$

which satisfy the following restrictions;

$$\left|\frac{\partial f}{\partial\alpha}\right|<F_1(x), \qquad \left|\frac{\partial f}{\partial\beta}\right|<F_2(x);$$

$$\left|\frac{\partial g}{\partial\alpha}\right|<G_1(\alpha^*, \beta^*), \quad \left|\frac{\partial g}{\partial\beta}\right|<G_2(\alpha^*, \beta^*);$$

$$\left|\frac{\partial h}{\partial\alpha}\right|<H_1(\xi_1, \dots, \xi_{n-2}\mid\alpha^*, \beta^*), \quad \left|\frac{\partial h}{\partial\beta}\right|<H_2(\xi_1, \dots, \xi_{n-2}\mid\alpha^*, \beta^*),$$

where $F_1(x)$, $F_2(x)$, $G_1(\alpha^*, \beta^*)$, $G_2(\alpha^*, \beta^*)$, α^*G_1, α^*G_2, β^*G_1, β^*G_2, H_1, H_2 are all independent α, β and are integrable in the whole space of variables.

In this case, the important results will be sketched as follows:

(1) The *ellipse of concentration* of the unbiased estimates α^*, β^* of α, β

$$\frac{1}{1-\rho^2}\left\{\frac{(u-\alpha)^2}{\sigma_1^2}-2\rho\frac{(u-\alpha)(v-\beta)}{\sigma_1\sigma_2}+\frac{(v-\beta)^2}{\sigma_2^2}\right\}=4 , \tag{1.24}$$

where

$$\sigma_1^2=D^2(\alpha^*), \; \sigma_2^2=D^2(\beta^*), \; \sigma_1\sigma_2\rho=C(\alpha^*,\beta^*) ,$$

contains a fixed ellipse

$$nE\left(\frac{\partial \log f}{\partial\alpha}\right)^2\cdot(u-\alpha)^2+2nE\left(\frac{\partial \log f}{\partial\alpha}\frac{\partial \log f}{\partial\beta}\right)\cdot(u-\alpha)(v-\beta)$$
$$+nE\left(\frac{\partial \log f}{\partial\beta}\right)^2\cdot(v-\beta)^2=4 \tag{1.24}$$

in its interior[7].

(2) Comparing areas of two ellipses, we have

$$n^2\Delta \geqq \frac{1}{\sigma_1^2\sigma_2^2(1-\rho)^2} ,$$

where $\Delta=E\left(\frac{\partial \log f}{\partial\alpha}\right)^2 E\left(\frac{\partial \log f}{\partial\beta}\right)^2-E^2\left(\frac{\partial \log f}{\partial\alpha}\cdot\frac{\partial \log f}{\partial\beta}\right)$, hence the quantity

$$e(\alpha^*,\beta^*)=\frac{1}{n^2\Delta\sigma_1^2\sigma_2^2(1-\rho^2)} \tag{1.25}$$

lies between 0 and 1. If two ellipses (1.23) and (1.24) coincide, it is obvious that

$$e(\alpha^*,\beta^*)=1 .$$

We define by the quantity $e(\alpha^*,\beta^*)$ the *efficiency of the joint estimates* α^*, β^* of α, β, and, when

$$e(\alpha^*,\beta^*)=1 ,$$

we say that α^*, β^* are *jointly efficient estimates* of α, β.

(3) Now let α_0^*, β_0^* be a system of jointly efficient estimates of α, β, then by the definition, we get

$$D^2(\alpha_0^*)=\frac{1}{n\Delta}E\left(\frac{\partial \log f}{\partial\beta}\right)^2, \quad D^2(\alpha_0^*)=\frac{1}{n\Delta}E\left(\frac{\partial \log f}{\partial\alpha}\right)^2,$$

$$\rho(\alpha_0^*,\beta_0^*)=-E\left(\frac{\partial \log f}{\partial\alpha}\cdot\frac{\partial \log f}{\partial\beta}\right)\Big/\sqrt{E\left(\frac{\partial \log f}{\partial\alpha}\right)^2\cdot E\left(\frac{\partial \log f}{\partial\beta}\right)^2} . \tag{1.26}$$

Hence, it follows at once

$$D^2(\alpha_0^*)=\frac{1}{1-\rho^2(\alpha_0^*,\beta_0^*)}\,\frac{1}{nE\left(\frac{\partial \log f}{\partial\alpha}\right)^2} . \tag{1.27}$$

Consequently, the variance of any one of the jointly efficient estimates is greater than that of the efficient estimate of the corresponding parameter assuming that the other being known, provided the coefficient of correlation $\rho(\alpha_0^*, \beta_0^*)$ is not zero.

(4) In a case when there are two unknown parameters α, β, it often arrives that we are only interested in estimating one of them, say α, and we then ask if it should be possible to find some pair of unbiased estimates α^*, β^*, yielding $D^2(\alpha^*) < D^2(\alpha_0^*)$, no matter how large the corresponding $D^2(\beta^*)$ becomes. However, from the result of (1), we get for any pair of unbiased estimates

$$\frac{1}{1-\rho^2}\left\{\frac{(u-\alpha)^2}{\sigma_1^2}-2\rho\frac{(u-\alpha)(v-\beta)}{\sigma_1\sigma_2}+\frac{(v-\beta)^2}{\sigma_2^2}\right\} \leqq nE\left(\frac{\partial \log f}{\partial \alpha}\right)\cdot(u-\alpha)^2$$
$$+2nE\left(\frac{\partial \log f}{\partial \alpha}\cdot\frac{\partial \log f}{\partial \beta}\right)\cdot(u-\alpha)(v-\beta)+nE\left(\frac{\partial \log f}{\partial \beta}\right)^2\cdot(v-\beta)^2, \quad (1.28)$$

it follows that

$$D^2(\alpha^*)=\sigma_1^2 \geqq \frac{1}{n\Delta}E\left(\frac{\partial \log f}{\partial \beta}\right)^2 = D^2(\alpha_0^*) .$$

(5) Thus far, we have assumed that $x_1, \dots, x_n$ are mutually independent, but if they are dependent, and let the joint frequency function be

$$f(x_1, \dots, x_n; \alpha) \quad \text{or} \quad f(x_1, \dots, x_n; \alpha, \beta),$$

then the whole arguments go parallel with slight modifications, for example, the inequality (1.13) becomes in this case

$$D^2(\alpha^*) \geqq \frac{1}{E\left(\frac{\partial \log f}{\partial \alpha}\right)^2} . \quad (1.29)$$

Example 3. Let the population be normal with unknown mean m and unknown standard deviation σ, then the frequency function is

$$f(x; m, \sigma)=(2\pi\sigma^2)^{-1/2}\exp\left\{-(x-m)^2/2\sigma^2\right\} .$$

In this case, it is easily seen that the amount of information is

$$\frac{2n^2}{\sigma^4} . \quad (1.30)$$

If we take as a pair of joint estimates of m and σ

$$m^* = \bar{x}, \quad \text{and} \quad \sigma^* = \sqrt{\frac{n-1}{2}}\,\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}s ,$$

where

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 ,$$

then, as is easily calculated, their variances and covariance are

$$D^2(m^*) = \frac{\sigma^2}{n}, \; D^2(\sigma^*) = \left(\frac{n-1}{2} \frac{\Gamma^2\left(\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)} - 1 \right) \sigma^2 = \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right), C(m^*, \sigma^*) = 0.$$

Hence, when n tends to infinity, we have

$$e(m^*, \sigma^*) \to 1 ,$$

therefore, m^*, σ^* are a pair of asymptotically efficient estimates of m, σ.

§ 2. **The best linear unbiased estimates and extensions of A. Markoff's theorem on least squares.** When an estimate $\alpha^*(x_1, \dots, x_n)$ of unknown parameter α satisfies the following three conditions, J. Neyman[8] named it *the best linear unbiased estimate* of α :

(1) $\alpha^*(x_1, \dots, x_n)$ is a linear form of a random sample $x_1, \dots, x_n$, and

(2) it is unbiased, i. e.,

$$E(\alpha^*) = \alpha$$

holds identically for all values of α.

(3) $\alpha^*(x_1, \dots, x_n)$ has the smallest variance among those which satisfy the conditions (1) and (2).

Because of the fact that in many cases of practical importance distributions of the linear estimates are asymptotically normal for large values of n, the sample size, we frequently make use of the best linear unbiased estimates.

As a powerful tool of obtaining the best linear unbiased estimates in certain important situations, we have the famous theorem due to A. Markoff.

Theorem 2. 1. (Extension of Markoff's Theorem by J. Neyman and F. N. David[9])

(i) $x_1, \dots, x_n$ are mutually independent ramdom variables.

(ii) The mean values of $x_1, \dots, x_n$ are linear forms of $s(< n)$ unknown parameters $\theta_1, \dots, \theta_s$, i. e.,

$$E(x_i) \equiv m_i = a_{i1}\theta_1 + \cdots + a_{is}\theta_s , \quad i = 1, 2, \dots, n , \tag{2 1}$$

where the coefficients a_{ij} are known constants.

(iii) The rank of the coefficient matrix

$$A=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1s} \\ a_{21} & a_{22} & \dots & a_{2s} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{ns} \end{pmatrix} \tag{2.2}$$

is s.

(iv) The variances of $x_i, (i=1,2,\dots,n)$ are of the following forms, i. e.,

$$D^2(x_i)=\sigma_i^2=\frac{\sigma^2}{P_i}, \quad i=1,2,\dots,n, \tag{2.3}$$

where $P_i, (i=1,2,\dots,n)$ are all known constants and σ^2 unknown.

If the above four conditions are satisfied, then (*i*) the best linear unbiased estimate F of a linear form of $\theta_1,\dots,\theta_s$

$$b_1\theta_1+\cdots+b_s\theta_s$$

with known constant coefficients $b_1,\dots,b_s$ is

$$F=b_1\theta_1^0+\cdots+b_s\theta_s^0, \tag{2.4}$$

where $\theta_1^0,\dots,\theta_s^0$ are the values of $\theta_1,\dots,\theta_s$ for which the quadratic form

$$S=\sum_{i=1}^{n}P_i(x_i-a_{i1}\theta_1-\cdots-a_{is}\theta_s)^2 \tag{2.5}$$

takes its minimum value.

And (*ii*) denoting the minimum value of S by S_0, i. e.,

$$S_0=\sum_{i=1}^{n}P_i(x_i-a_{i1}\theta_1^0-\cdots-a_{is}\theta_s^0)^2, \tag{2.6}$$

the statistic

$$S_0/(n-s)$$

is an unbiased estimate of σ^2.

For the convenience of later uses, we extend the above theorem slightly, and state as follows:

Theorem 2. 2.[10)]

(i) $x_1,\dots,x_n$ are distributed according to the n-dimensional non-singular distribution with means $m_1,\dots,m_n$ and the variance-covariance matrix

$$(\sigma^2 d_{ij}), \quad i,j=1,2,\dots,n; \tag{2.7}$$

where d_{ij} are known constants and σ^2 unknown.

(ii) The means $m_1,\dots,m_n$ are linear forms of $s(<n)$ unknown parameters $\theta_1,\dots,\theta_s$ with known constant coefficients, i. e.,

$$E(x_i)\equiv m_i=a_{i1}\theta_1+\cdots+a_{is}\theta_s, \quad i=1,2,\dots,n. \tag{2.8}$$

(iii) The rank of the coefficient matrix $A=(a_{ij})$ is s.

(iv) Let

$$(d_{ij})^{-1}=(D_{ij})$$

If the above four conditions are satisfied, the following two propositions (***i***) and (***ii***) are valid.

(***i***) The best linear unbiased estimate F of a linear form of $\theta_1, \dots, \theta_s$ with known constant coefficients

$$b_1\theta_1+\cdots+b_s\theta_s$$

is

$$F=b_1\theta_1^0+\cdots+b_s\theta_s^0, \tag{2.9}$$

where $\theta_1^0, \dots, \theta_s^0$ are the values of $\theta_1, \dots, \theta_s$ for which the quadratic from

$$S=\sum_{i=1}^{n}\sum_{j=1}^{n}D_{ij}(x_i-\sum_{\alpha=1}^{s}a_{i\alpha}\theta_\alpha)(x_j-\sum_{\beta=1}^{s}a_{j\beta}\theta_\beta)\,, \tag{2.10}$$

takes its minimum value.

(***ii***) Denoting the minimum value of S by S_0, i.e.,

$$S_0=\sum_{i=1}^{n}\sum_{j=1}^{n}D_{ij}(x_i-\sum_{\alpha=1}^{s}a_{i\alpha}\theta_\alpha^0)(x_j-\sum_{\beta=1}^{s}a_{j\beta}\theta_\beta^0)\,, \tag{2.11}$$

the statistic

$$S_0/(n-s)$$

is an unbiased estimate of σ^2.

§ 3. **Order statistics and their limiting distributions, Systematic statistics.** Rearranging a random sample $x_1, \dots, x_n$, of size n, in ascending order of their magnitudes, we write

$$x(1)<x(2)<\dots<x(n)\,,$$

and call them *order statistics*. If we consider the parent population of the continuous type, then

$$P(x(i)=x(j))=0\,, \quad \text{for all} \quad i\neq j\,,$$

hence we may disregard the cases when equalities occur.

Now, let the frequency function of the parent population be $g(x)$, and for any given number λ, which is

$$0<\lambda<1\,,$$

we define the λ-*quantile* or *100 λ percent point* of the population as the value $x=x_\lambda$ for which

$$\int_{-\infty}^{x_\lambda} g(t)dt = \lambda \tag{3.1}$$

holds. For example, when $\lambda = 0.5$, the 50 percent point of the population $x_{0\cdot 5}$ is the so-called "*median*" of the population.

For the sake of simplicity, assuming that $n\lambda$ is not an integer, we call the order statistic

$$z_\lambda = x([n\lambda]+1) \tag{3.2}$$

the *λ-quantile of the sample*, where the symbol [ ] is Gauss'.

Theorem 3. 1.[10] *If $g(x)$ is differentiable in the neighbourhood of $x = x_\lambda$ and*

$$g(x_\lambda) \neq 0 , \tag{3.3}$$

then the distribution of the statistic

$$\sqrt{\frac{n}{\lambda(1-\lambda)}} \cdot g(x_\lambda) \cdot (z_\lambda - x_\lambda) \tag{3.4}$$

tends asymptotically normal $N(0, 1)$ according as n tends to infinity. Hence the frequency function of the distribution of z_λ is asymptotically

$$\sqrt{\frac{n}{2\pi\lambda(1-\lambda)}} \cdot g(x_\lambda) \cdot \exp\left\{-\frac{ng^2(x_\lambda)}{2\lambda(1-\lambda)}(z_\lambda - x_\lambda)^2\right\} \tag{3.5}$$

for sufficiently large values of n.

As an extension of the above theorem, we have the following

Theorem 3. 2. (F. Mosteller)[12] *For k given real numbers*

$$0 < \lambda_1 < \lambda_2 < \cdots < \lambda_k < 1 ,$$

let the λ_i-quantile of the population be x_i, $i = 1, 2, \dots, k$, i. e.

$$\int_{-\infty}^{x_i} g(t)dt = \lambda_i , \quad i = 1, 2, \dots, k . \tag{3.6}$$

And assume that the frequency function $g(x)$ of the population is differentiable in the neighbourhoods of $x = x_i$, $i = 1, 2, \dots, k$, and

$$g_i = g(x_i) \neq 0 , \quad i = 1, 2, \dots, k , \tag{3.7}$$

then the joint distribution of k order statistics $x(n_1), x(n_2), \dots, x(n_k)$, where

$$n_i = [n\lambda_i]+1 , \quad i = 1, 2, \dots, k ,$$

tends to k-dimensional normal distribution with means $x_1, \dots, x_k$ and variance-covariance matrix

$$\begin{pmatrix} \frac{\lambda_1(1-\lambda_1)}{ng_1^2} & \frac{\lambda_1(1-\lambda_2)}{ng_1g_2} & \cdots\cdots\cdots & \frac{\lambda_1(1-\lambda_k)}{ng_1g_k} \\ \frac{\lambda_1(1-\lambda_2)}{ng_1g_2} & \frac{\lambda_2(1-\lambda_2)}{ng_2^2} & \cdots\cdots\cdots & \frac{\lambda_2(1-\lambda_k)}{ng_2g_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\lambda_1(1-\lambda_k)}{ng_1g_k} & \frac{\lambda_2(1-\lambda_k)}{ng_2g_k} & \cdots\cdots\cdots & \frac{\lambda_k(1-\lambda_k)}{ng_k^2} \end{pmatrix}, \tag{3.8}$$

according as n tends to infinity. Hence the frequency function of the limiting distribution becomes

$$(2\pi)^{-k/2}g_1\ldots g_k[\lambda_1(\lambda_2-\lambda_1)\ldots(\lambda_k-\lambda_{k-1})(1-\lambda_k)]^{-1/2}\cdot n^{k/2}\times$$
$$\exp\left[-\frac{n}{2}\left\{\sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})}g_i^2\cdot\left(x(n_i)-x_i\right)^2\right.\right.$$
$$\left.\left.-2\sum_{i=2}^{k}\frac{g_ig_{i-1}}{\lambda_i-\lambda_{i-1}}(x(n_i)-x_i)(x(n_{i-1})-x_{i-1})\right\}\right]. \tag{3.9}$$

In particular, if we consider a normal population $N(m,\sigma)$, then the frequency function of the population is

$$g(x)=(2\pi\sigma^2)^{-1/2}\exp\left\{-(x-m)^2/2\sigma^2\right\}. \tag{3.10}$$

Let the frequency function of the standard normal population $N(0,1)$ be

$$f(x)=(2\pi)^{-1/2}\exp(-x^2/2), \tag{3.11}$$

and let the λ_i-quantile and the ordinate at that point of the standard normal population be u_i and f_i respectively, i. e.,

$$\left.\begin{aligned}\int_{-\infty}^{u_i}f(t)dt&=\lambda_i,\\ f(u_i)&=f_i,\end{aligned}\right. \qquad i=1,2,\ldots,k \tag{3.12}$$

then (3.6) can be written as follows

$$\int_{-\infty}^{\frac{x_i-m}{\sigma}}f(t)dt=\lambda_i,\qquad i=1,2,\ldots,k.$$

Comparing these with (3.12), we obtain relations

$$x_i=m+u_i\sigma,\qquad i=1,2,\ldots,k. \tag{3.13}$$

Consequently, the frequency function of the limiting distribution of $x(n_1),\ldots,x(n_k)$ becomes

$$h(x(n_1), \ldots, x(n_k)) = (2\pi\sigma^2)^{-k/2} f_1 \ldots f_k \Big[\lambda_1(\lambda_2-\lambda_1)\ldots(1-\lambda_k)\Big]^{-1/2} n^{k/2}$$

$$\times \exp\Big[-\frac{n}{2\sigma^2}\Big\{\sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2\cdot(x(n_i)-m-u_i\sigma)^2$$

$$-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}}(x(n_i)-m-u_i\sigma)(x(n_{i-1})-m-u_{i-1}\sigma)\Big\}\Big]. \tag{3.14}$$

F. Mosteller[13] named such statistics as are functions of order satistics *systematic ones*. In the following, we shall develop the large sample theory of such systematic statistics as are linear functions of $x(n_1), \ldots, x(n_k)$.

§ 4. The efficiencies of systematic statistics for estimating parameters of a normal population. We shall be concerned with the theory of estimation of the mean and standard deviation of a normal population $N(m, \sigma)$, making use of the limiting distribution of $x(n_1), \ldots, x(n_k)$, when the sample size is sufficiently large. In this section, we shall consider the efficiencies of systematic statistics for estimating the mean m and standard deviation σ separately and for estimating m and σ jointly.

Case 1. The case when σ is known. Let

$$S = \sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2\cdot(x(n_i)-m-u_i\sigma)^2$$

$$-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}}(x(n_i)-m-u_i\sigma)(x(n_{i-1})-m-u_{i-1}\sigma), \tag{4.1}$$

then we have from (3.14)

$$\log h = -nS/2\sigma^2 + \text{terms independent of } m.$$

Hence, it follows that

$$-\frac{\partial^2 \log h}{\partial m^2} = \frac{n}{2\sigma^2}\frac{\partial^2 S}{\partial m^2} = \frac{n}{\sigma^2}K_1, \tag{4.2}$$

where

$$K_1 = \sum_{i=1}^{k+1}\frac{(f_i-f_{i-1})^2}{\lambda_i-\lambda_{i-1}}, \tag{4.3}$$

assuming that

$$\lambda_0 = 0, \quad \lambda_{k+1} = 1, \quad f_0 = f_{k+1} = 0.$$

Hence the efficiency of estimation by means of systematic statistics is

$$\eta_m = -E\left(\frac{\partial^2 \log h}{\partial m^2}\right)\Big/\frac{n}{\sigma^2} = K_1. \tag{4.4}$$

Case 2. The case when m is known. In this case

$$\log h = -k \log \sigma - nS/2\sigma^2 + \text{terms independent of } \sigma,$$

hence it follows that

$$\frac{\partial^2 \log h}{\partial \sigma^2} = \frac{k}{\sigma^2} - \frac{3n}{\sigma^4} S + \frac{2n}{\sigma^3}\frac{\partial S}{\partial \sigma} - \frac{n}{2\sigma^2} \cdot \frac{\partial^2 S}{\partial \sigma^2} .$$

If we put

$$K_2 = \sum_{i=1}^{k+1} \frac{(f_i u_i - f_{i-1} u_{i-1})^2}{\lambda_i - \lambda_{i-1}} , \tag{4.5}$$

where, as before,

$$\lambda_0 = 0 , \quad \lambda_{k+1} = 1 , \quad f_0 = f_{k+1} = 0 ,$$

it is easily seen that

$$\frac{\partial^2 S}{\partial \sigma^2} = K_2 , \quad E\left(\frac{\partial S}{\partial \sigma}\right) = 0 , \quad E(S) = \frac{\sigma^2}{n} k ,$$

whence we have

$$-E\left(\frac{\partial^2 \log h}{\partial \sigma^2}\right) = \frac{2k}{\sigma^2} + \frac{n}{\sigma^2} K_2 = \frac{n}{\sigma^2} K_2 \left(1 + 2\frac{k}{n} K_2^{-1}\right) . \tag{4.6}$$

Here, of course, k is very small compared with n, so we may put

$$\frac{k}{n} \doteqdot 0 .$$

Consequently, we have

$$-E\left(\frac{\partial^2 \log h}{\partial \sigma^2}\right) \doteqdot \frac{n}{\sigma^2} K_2 .$$

The efficiency η_σ of systematic statistics for estimating σ is

$$\eta_\sigma = \frac{n}{\sigma^2} K_2 \Big/ \frac{2n}{\sigma^2} = \frac{1}{2} K_2 . \tag{4.7}$$

Case 3. The case when both m and σ are unknown. In this case, we have

$$\log h = -k \log \sigma - nS/2\sigma^2 + \text{terms independent of } m \text{ and } \sigma .$$

If we put

$$K_3 = \sum_{i=1}^{k+1} \frac{(f_i - f_{i-1})(f_i u_i - f_{i-1} u_{i-1})}{\lambda_i - \lambda_{i-1}}, \quad \text{where } \lambda_0 = 0 , \ \lambda_{k+1} = 1, \ f_0 = f_{k+1} = 0 , \tag{4.8}$$

and

$$\Delta = K_1 K_2 - K_3^2 , \tag{4.9}$$

then the intrinsic accuracy of the joint estimates of m and σ by means of systematic statistics is

$$E\left(\frac{\partial^2 \log h}{\partial m^2}\right)\cdot E\left(\frac{\partial^2 \log h}{\partial \sigma^2}\right)-E^2\left(\frac{\partial^2 \log h}{\partial m\partial \sigma}\right)=\frac{n^2}{\sigma^4}\Delta+\frac{2nk}{\sigma^4}\cdot K_1 . \tag{4.10}$$

Hence, the efficiency of the joint estimates of m and σ by means of systematic statistics is

$$\eta=\left(\frac{n^2}{\sigma^4}\Delta+\frac{2nk}{\sigma^4}K_1\right)\Big/\frac{2n^2}{\sigma^4}=\frac{1}{2}\Delta+\frac{k}{n}K_1\doteqdot\frac{1}{2}\Delta . \tag{4.11}$$

§ 5. **The best linear unbiased estimates of the mean and standard deviation of a normal population by means of systematic statistics.** The basic frequency function of $x(n_1), \dots, x(n_k)$ is given by $h(x(n_1), \dots, x(n_k))$ of (3.14), and to obtain the best linear unbiased estimates of m and σ, we can apply the extended Markoff's theorem (Theorem 2.2). In this case, the quadratic form S corresponding to (2.10) is given by (4.1).

We shall consider step by step three cases of the preceeding section.

Case 1. The case when σ is known. Let the best linear unbiased estimate of m be $\hat{m}_0$, then it should be obtained by solving the equation

$$\left.\frac{\partial S}{\partial m}\right|_{m=\hat{m}_0}=0 .$$

Whence we have

$$K_1\cdot\hat{m}_0=\sum_{i=1}^{k}\left(\frac{f_i-f_{i-1}}{\lambda_i-\lambda_{i-1}}-\frac{f_{i+1}-f_i}{\lambda_{i+1}-\lambda_i}\right)\cdot f_i\cdot(x(n_i)-u_i\sigma) . \tag{5.1}$$

Put

$$X=\sum_{i=1}^{k+1}\frac{(f_i-f_{i-1})(f_i\cdot x(n_i)-f_{i-1}\cdot x(n_{i-1}))}{\lambda_i-\lambda_{i-1}} , \tag{5.2}$$

then, (5.1) is written in the form

$$K_1\cdot\hat{m}_0=X-\sigma K_3 . \tag{5.1'}$$

Hence, it follows that

$$\hat{m}_0=\frac{X}{K_1}-\sigma\cdot\frac{K_3}{K_1} . \tag{5.3}$$

Since the variance of X is

$$D^2(X)=\frac{\sigma^2}{n}\cdot K_1 , \tag{5.4}$$

it follows at once that

$$D^2(\hat{m}_0) = \frac{\sigma^2}{n}\frac{1}{K_1}, \tag{5.5}$$

whence we see readily that $\hat{m}_0$ is an efficient estimate of m so far as we use systematic statistics.

Case 2. The case when m is known. Let the best linear unbiased estimate of σ be $\hat{\sigma}_0$, then it should be obtained by solving the equation

$$\left.\frac{\partial S}{\partial \sigma}\right|_{\sigma=\hat{\sigma}_0} = 0 .$$

Whence we have

$$K_2\hat{\sigma}_0 = \sum_{i=1}^{k}\left(\frac{f_i u_i - f_{i-1}u_{i-1}}{\lambda_i - \lambda_{i-1}} - \frac{f_{i+1}u_{i+1} - f_i u_i}{\lambda_{i+1} - \lambda_i}\right)\cdot f_i \cdot (x(n_i) - m) . \tag{5.6}$$

Put

$$Y = \sum_{i=1}^{k+1}\frac{(f_i u_i - f_{i-1}u_{i-1})(f_i \cdot x(n_i) - f_{i-1}\cdot x(n_{i-1}))}{\lambda_i - \lambda_{i-1}}, \tag{5.7}$$

then (5.6) is written in the form

$$K_2 \cdot \hat{\sigma}_0 = Y - mK_3 . \tag{5.6'}$$

Hence, it follows that

$$\hat{\sigma}_0 = \frac{Y}{K_2} - m \cdot \frac{K_3}{K_2} . \tag{5.8}$$

Since the variance of Y is

$$D^2(Y) = \frac{\sigma^2}{n}K_2 , \tag{5.9}$$

it follows at once that

$$D^2(\hat{\sigma}_0) = \frac{\sigma^2}{n}\frac{1}{K_2} . \tag{5.10}$$

Whence we may consider the estimate $\hat{\sigma}_0$ as an efficient one so far as we use systematic statistics.

Case 3. The case when both m and σ are unknown. Let the best linear unbiased estimates m and σ be $\hat{m}$ and $\hat{\sigma}$ respectively. Then they should be obtained by solving a system of equations

$$\left.\frac{\partial S}{\partial m}\right|_{\substack{m=\hat{m}\\ \sigma=\hat{\sigma}}} = 0 , \qquad \left.\frac{\partial S}{\partial \sigma}\right|_{\substack{m=\hat{m}\\ \sigma=\hat{\sigma}}} = 0 .$$

If we use the notations introduced by (4.3), (4.5), (5.2), and (5.7), the above equations may be written as follows;

$$K_1\hat{m}+K_3\hat{\sigma}=X\ .$$
$$K_3\hat{m}+K_2\hat{\sigma}=Y\ . \tag{5.11}$$

Whence we have

$$\hat{m}=\frac{1}{\Delta}(K_2X-K_3Y)\ ,$$
$$\hat{\sigma}=\frac{1}{\Delta}(-K_3X+K_1Y)\ . \tag{5.12}$$

From (5.4), (5.9) and the fact that

$$C(X,Y)=\frac{\sigma^2}{n}K_3\ , \tag{5.13}$$

we have

$$D^2(\hat{m})=\frac{1}{\Delta^2}(K_2^2D^2(X)-2K_2K_3\cdot C(X,Y)+K_3^2D^2(Y))=\frac{\sigma^2}{n}\frac{K_2}{\Delta}\ ,$$
$$D^2(\hat{\sigma})=\frac{1}{\Delta^2}(K_3^2D^2(X)-2K_3K_1\cdot C(X,Y)+K_1^2D^2(Y))=\frac{\sigma^2}{n}\frac{K_1}{\Delta}\ , \tag{5.14}$$
$$C(\hat{m},\hat{\sigma})=\frac{1}{\Delta^2}(-K_2K_3D^2(X)+(K_1K_2+K_3^2)\cdot C(X,Y)-K_1K_3D^2(Y))=-\frac{\sigma^2}{n}\frac{K_3}{\Delta}\ .$$

Therefore it follows that

$$D^2(\hat{m})D^2(\hat{\sigma})-C^2(\hat{m},\hat{\sigma})=\frac{\sigma^4}{n^2}\cdot\frac{1}{\Delta}\ . \tag{5.15}$$

Hence, in this case also, we may consider $\hat{m}$, $\hat{\sigma}$ are jointly efficient estimates of m, σ, so far as we use systematic statistics.

In particular, if the relations

$$n_i+n_{k-i+1}=n\ ,\quad i=1,2,\dots,k\ ,$$

or in terms of λ,

$$\lambda_i+\lambda_{k-i+1}=1\ ,\quad i=1,2,\dots,k\ , \tag{5.16}$$

hold, we say that the order statistics $x(n_1),\dots,x(n_k)$ are *symmetrically spaced* or *in symmetric spacing*. In such a case, it follows that

$$u_i+u_{k-i+1}=0\ ,\quad i=1,2,\dots,k\ ,$$

hence

$$K_3=0\ .$$

Therefore

$$D^2(\hat{m})=\frac{\sigma^2}{n}\frac{1}{K_1}\ ,\qquad D^2(\hat{\sigma})=\frac{\sigma^2}{n}\frac{1}{K_2}\ ,$$

hence, they coincide with the variances of $\hat{m}_0$ and $\hat{\sigma}_0$ respectively. F. Mosteller and Z. Yamanouchi dealt solely with those cases.

§ 6. Determination of the optimum spacing. We have seen in § 5 that $\hat{m}_0$, $\hat{\sigma}_0$ and $\hat{m}$, $\hat{\sigma}$ are efficient estimates in each case, given $\lambda_1, \dots,$ λ_k, however, we may raise their efficiencies by choosing the spacing suitably. The values of $\lambda_1, \dots, \lambda_k$ for which the efficiencies of estimates attain their maximum are called the *optimum spacings*. In this section we shall consider the problems of determination of the optimum spacings.

Case 1. The case when σ is known. The required optimum spacing is such that it makes the efficiency $\eta_m = K_1$ maximum. Since

$$\frac{df_i}{du_i} = -u_i f_i \ , \quad \frac{d\lambda_i}{du_i} = f_i \ , \qquad i = 1, 2, \dots, k \ ,$$

we have by differentiating K_1 with respect to u_i, $i = 1, 2, \dots, k$,

$$\frac{\partial K_1}{\partial u_i} = f_i \cdot \left(\frac{f_{i+1} - f_i}{\lambda_{i+1} - \lambda_i} - \frac{f_i - f_{i-1}}{\lambda_i - \lambda_{i-1}} \right) \left(2u_i + \frac{f_i - f_{i-1}}{\lambda_i - \lambda_{i-1}} + \frac{f_{i+1} - f_i}{\lambda_{i+1} - \lambda_i} \right), \quad i = 1, 2, \dots, k \ .$$

Whence we have the following conditions for optimum spacings

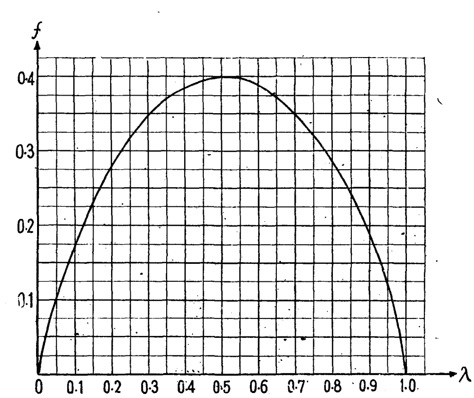
$$\frac{f_{i+1} - f_i}{\lambda_{i+1} - \lambda_i} = \frac{f_i - f_{i-1}}{\lambda_i - \lambda_{i-1}} \ , \qquad i = 1, 2, \dots, k \ , \tag{6.1}$$

and

$$2u_i + \frac{f_i - f_{i-1}}{\lambda_i - \lambda_{i-1}} + \frac{f_{i+1} - f_i}{\lambda_{i+1} - \lambda_i} = 0 \ , \qquad i = 1, 2, \dots, k \ . \tag{6.2}$$

If we consider the ordinate f of the standard normal function as a function of the cumulative frequency λ, then its graph is convex upwards, as will be seen in the next Fig. 6.1, therefore, (6.1) is inconsistent unless any two of f_{i-1}, f_i, f_{i+1} coincide with each other. Consequently, we may consider only the equations (6.2) for the determination of the optimum spacing.

Fig. 6.1. The graph of f as a function of λ.



We conjecture that the values of $u_1, \dots, u_k$ which satisfy (6.2) are in symmetric spacing, but at present, we can not prove it[14].

Assuming the symmetric spacing of $u_1, \dots, u_k$, we solved (6.2) numerically for $k = 1, 2, \dots, 10$, and computed the maximum efficiencies corresponding to those spacings. The results are shown in Table 6.1. The two lowest rows of Table 6.1 are the values of

$\frac{1}{2}K_2$ and $\frac{1}{2}\Delta$ corresponding to those spacings.

In Table 6.2, there are comparisons of efficiencies corresponding to the optimum spacings with those corresponding to equal probability spacings $(\lambda_{i+1}-\lambda_i=\lambda_i-\lambda_{i-1})$ and with those corresponding to such spacing as $\lambda_i=\left(i-\frac{1}{2}\right)\Big/k,\ i=1, 2, \dots, k$.

Table 6.1.

Table of the optimum spacings for estimating the mean m and the maximum efficiencies (assuming symmetric spacings).

λ, u \ k	1	2	3	4	5	6	7	8	9	10
λ_1	0.500	0.270	0.163	0.107	0.074	0.055	0.040	0.031	0.024	0.020
u_1	0.000	−0.613	−0.982	−1.243	−1.447	−1.598	−1.751	−1.866	−1.977	−2.054
λ_2		0.730	0.500	0.351	0.255	0.195	0.147	0.115	0.092	0.076
u_2		0.613	0.000	−0.383	−0.659	−0.860	−1.049	−1.200	−1.329	−1.433
λ_3			0.837	0.649	0.500	0.395	0.308	0.247	0.202	0.167
u_3			0.982	0.383	0.000	−0.266	−0.502	−0.634	−0.834	−0.966
λ_4				0.893	0.745	0.605	0.500	0.412	0.343	0.288
u_4				1.243	0.659	0.266	0.000	−0.222	−0.404	−0.559
λ_5					0.926	0.305	0.692	0.588	0.500	0.427
u_5					1.447	0.860	0.502	0.222	0.000	−0.134
λ_6						0.945	0.852	0.753	0.657	0.573
u_6						1.598	1.049	0.684	0.404	0.184
λ_7							0.960	0.885	0.798	0.712
u_7							1.751	1.200	0.834	0.559
λ_8								0.969	0.908	0.833
u_8								1.866	1.329	0.966
λ_9									0.976	0.924
u_9									1.977	1.433
λ_{10}										0.980
u_{10}										2.504
K_1	0.6366	0.8097	0.8800	0.9342	0.9420	0.9559	0.9654	0.9722	0.9771	0.9808
$\frac{1}{2}K_2$	0.0000	0.3303	0.5326	0.6566	0.7392	0.7902	0.8516	0.8620	0.8858	0.9016
$\frac{1}{2}\Delta$	0.000	0.267	0.469	0.614	0.696	0.752	0.822	0.838	0.866	0.884

Table 6. 2.

Comparisons of efficiencies corresponding to various spacings.

Spacings / k	Optimum Spacing	Equal Probability Spacing	$\lambda_i = \frac{i-1/2}{k}$
1	0.6366	0.6366	0.637
2	0.8097	0.7926	0.808
3	0.8300	0.8606	0.878
4	0.9342	0.8969	0.913
5	0.9420	0.9172	0.934
6	0.9559	0.9352	0.948
7	0.9654	0.9450	0.957
8	0.9722	0.9521	0.963
9	0.9771	0.9591	0.969
10	0.9808	0.9634	0.973

Table 6. 3.

The expressions of the best linear unbiased estimates of the mean m corresponding to the optimum spacings.

2	$\frac{1}{2}[(x(0.270n)+x(0.730n))]$
3	$0.297[x(0.163n)+x(0.837n)]+0.407\cdot x(0.500n)$
4	$0.197[x(0.107n)+x(0.893n)]+0.303[x(0.351n)+x(0.649n)]$
5	$0.133[x(0.074n)+x(0.926n)]+0.233[x(0.255n)+x(0.745n)]+0.269\cdot x(0.500n)$
6	$0.099[x(0.055n)+x(0.945n)]+0.181[x(0.195n)+x(0.805n)]$ $+0.220[x(0.395n)+x(0.605n)]$
7	$0.071[x(0.040n)+x(0.960n)]+0.140[x(0.147n)+x(0.853n)]$ $+0.186[x(0.308n)+x(0.692n)]+0.203\cdot x(0.500n)$
8	$0.049[x(0.031n)+x(0.969n)]+0.111[x(0.115n)+x(0.885n)]$ $+0.155[x(0.247n)+x(0.753n)]+0.178[x(0.412n)+x(0.588n)]$
9	$0.044[x(0.024n)+x(0.976n)]+0.091[x(0.092n)+x(0.908n)]$ $+0.130[x(0.202n)+x(0.798n)]+0.155[x(0.343n)+x(0.657n)]$ $+0.163\cdot x(0.500n)$
10	$0.036[x(0.020n)+x(0.980n)]+0.075[x(0.076n)+x(0.924n)]$ $+0.109[x(0.167n)+x(0.833n)]+0.133[x(0.288n)+x(0.712n)]$ $+0.147[x(0.427n)+0.573n)]$

n is the sample size.

Case 2. The case when *m* is known. In this case the required spacing is such that it makes the efficiency $\eta_\sigma = \frac{1}{2}K_2$ maximum. Differentiating K_2 with respect to u_i, as in the former case, we have

$$\frac{\partial K_2}{\partial u_i} = f_i \cdot \left(\frac{f_{i+1}u_{i+1} - f_i u_i}{\lambda_{i+1} - \lambda_i} - \frac{f_i u_i - f_{i-1}u_{i-1}}{\lambda_i - \lambda_{i-1}} \right) \times$$
$$\left(2u_i^2 - 2 + \frac{f_i u_i - f_{i-1}u_{i-1}}{\lambda_i - \lambda_{i-1}} + \frac{f_{i+1}u_{i+1} - f_i u_i}{\lambda_{i+1} - \lambda_i} \right), \quad i = 1, 2, \dots, k,$$

hence, it follows the conditions for the optimum spacing

$$\frac{f_{i+1}u_{i+1} - f_i u_i}{\lambda_{i+1} - \lambda_i} = \frac{f_i u_i - f_{i-1}u_{i-1}}{\lambda_i - \lambda_{i-1}}, \quad i = 1, 2, \dots, k. \tag{6.4}$$

and

$$2u_i^2 - 2 + \frac{f_i u_i - f_{i-1}u_{i+1}}{\lambda_i - \lambda_{i+1}} + \frac{f_{i+1}u_{i+1} - f_i u_i}{\lambda_{i+1} - \lambda_i} = 0, \quad i = 1, 2, \dots, k. \tag{6.5}$$

The graph of fu as a function of λ is seen in the following Fig. 6.2, this is symmetric with respect to $\lambda = 0.5$, and in the interval $0 \leqq \lambda \leqq 0.5$ it is convex upwards and in $0.5 \leqq \lambda \leqq 1.0$ convex downwards. We can easily see that when k is odd, the solution of the system of equations (6.5) is not in symmetric spacing[15]. But the general characters of the solutions of (6.4) and (6.5) are yet unknown to us[16]. We solved (6.5) numerically for $k = 1, \dots, 6$. The results similar to the former case are shown in Tables 6.4-6.6.

Fig. 6.2. The graph of fu as a function of λ.

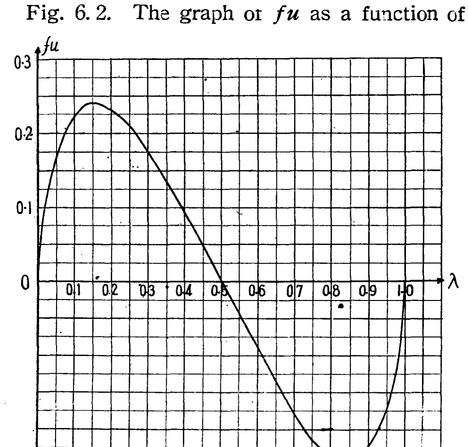


Table 6. 4.

Table of the optimum spacings for estimating the standard deviation σ and the corresponding maximum efficiencies.

k	1		2			3				4			5		6
λ_1	0.058	0.942	0.069	(*) 0.905	(**) 0.018	0.021	0.072	(***) 0.007	0.882	0.023	(****) 0.009	0.072	0.010	0.025	0.010
u_1	−1.572	1.572	−1.483	1.311	−2.097	−2.034	−1.463	−2.457	1.185	−1.995	−2.366	−1.461	−2.326	−1.960	−2.318
		λ_2	0.931	0.982	0.095	0.118	0.882	0.039	0.961	0.127	0.048	0.852	0.053	0.133	0.056
		u_2	1.483	2.097	−1.311	−1.185	1.185	−1.762	1.762	−1.141	−1.665	1.045	−1.618	−1.112	−1.589
					λ_3	0.928	0.979	0.118	0.993	0.873	0.148	0.952	0.163	0.837	0.171
					u_3	1.468	2.034	−1.185	2.457	1.141	−1.045	1.665	−0.982	0.982	−0.950
									λ_4	0.977	0.928	0.991	0.867	0.947	0.829
									u_4	1.995	1.461	2.366	1.112	1.618	0.950
												λ_5	0.975	0.990	0.944
												u_5	1.960	2.326	1.589
														λ_6	0.990
														u_6	2.318
$\frac{1}{2}K_2$	0.304		0.653	0.374		0.729		0.399		0.824	0.745		0.858		0.893
K_1	0.246		0.512	0.352		0.605		0.370		0.715	0.658		0.764		0.814
K_3	0.387		0	∓0.482		0.393		±0.500		0	0.823		±0.035		0
$\frac{1}{2}\Delta$	0.003		0.334	0.061		0.367		0.002		0.589	0.151		0.656		0.727

(*), (**), and (****) are non-symmetric solutions of (6.5) for even k, and (***) one-sided solutions of (6.5) for cdd k, As is seen above, they do not give maximum efficiencies.

Table 6.5

Comparisons of efficiencies corresponding to various spacings.

Spacings / k	Optimum Spaicng	Equal Probability Spacing	$\lambda_i = \frac{i-1/2}{k}$
1	0.304	0.000	0.000
2	0.653	0.221	0.413
3	0.729	0.368	0.526
4	0.824	0.468	0.619
5	0.858	0.541	0.681
6	0.893	0.595	0.725

Table 6.6

The expressions of the best linear unbiased estimates of the standard deviation σ corresponding to the optimum spacings. (n is the sample size)

k	Expression
2	$0.674[x(0.931n)-x(0.069n)]$
3	$0.070m+0.305\cdot x(0.928n)-0.253\cdot x(0.118n)-0.123\cdot x(0.021n)$ $-0.070m-0.305\cdot x(0.072n)+0.253\cdot x(0.882n)+0.123\cdot x(0.979n)$
4	$0.115[x(0.977n)-x(0.023n)]+0.237[x(0.873n)-x(0.127n)]$
5	$0.020\cdot m+0.117\cdot x(0.975n)+0.230\cdot x(0.867n)-0.186\cdot x(0.169n)-0.126\cdot x(0.053)-0.056x(0.010n)$ $-0.020\cdot m-0.117\cdot x(0.025n)-0.230\cdot x(0.133n)+0.186\cdot x(0.831n)+0.126\cdot x(0.947n)+0.056x(0.990n)$
6	$0.056[x(0.990n)-x(0.011n)]+0.126[x(0.944n)-x(0.056n)]+0.181[x(0.829n)-x(0.171n)]$

Case 3. The case when both m and σ are unknown. In this case, we must obtain the spacing which makes

$$\eta = \frac{1}{2}\Delta = \frac{1}{2}(K_1K_2 - K_3^2)$$

maximum. Differentiating Δ with respect to u_i, $i=1, \dots, k$, as before, we have

$$\frac{\partial K_1}{\partial u_i}K_2 + K_1\frac{\partial K_2}{\partial u_i} = 2K_3\frac{\partial K_3}{\partial u_i}, \quad i=1, 2, \dots, k.$$

It is difficult to solve the above equations numerically, even we assume the symmetric spacings.

For $k=2$, assuming symmetricity, we have

$$K_1 = 2\frac{f_1^2}{\lambda_1}\,, \quad K_2 = 2\frac{f_1^2 u_1^2}{\lambda_1(1-2\lambda_1)}\,,$$

hence

$$\eta = 2\frac{f_1^4 u_1^2}{\lambda_1^2(1-2\lambda_1)}\,.$$

In this particular case (6. 6) becomes

$$2u_1 - \frac{1}{u_1} + \frac{f_1}{\lambda_1} - \frac{f_1}{1-2\lambda_1} = 0\,. \tag{6.7}$$

Whence, solving (6. 7) numerically, we get

$$\lambda_1 = 0.134\,, \quad u_1 = -1{,}10768\,, \quad f_1 = 0{,}21601\,,$$
$$\lambda_2 = 0{,}866\,, \quad u_2 = 1{,}10768\,, \quad f_2 = 0{,}21601\,.$$

Consequently the maximum efficiency is

$$\eta_{max} = 0{,}4066\,.$$

Chapter II. Theory of Testing Statistical Hypotheses.

§ 7. **On tests of general linear hypotheses.**[17] Let there be given n normal populations $N(m_i, \sigma)$, $i = 1, 2, \dots, n$, with common variance σ^2 and assume that population means $m_i (i = 1, 2, \dots, n)$ are linear forms of $s(< n)$ unknown parameters $\theta_1, \dots, \theta_s$ with known constant coefficients a_{ij}, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, s$, i. e.,

$$m_i = a_{i1}\theta_1 + \cdots + a_{is}\theta_s\,, \quad i = 1, 2, \dots, n\,. \tag{7.1}$$

The statistical hypothesis H that

$$\begin{array}{l} b_{11}\theta_1 + \cdots + b_{1s}\theta_s = B_1^0 \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ b_{r1}\theta_1 + \cdots + b_{rs}\theta_s = B_r^0 \end{array} \quad (0 < r \leq s)\,, \tag{7.2}$$

where b_{ij} and B_i^0 are known constants, is called *a general linear hypothesis*. General linear hypothesis contains as its special cases almost all situations of practical importance[18].

Without any loss of generality, we can assume that the rank of the matrix (b_{ij}) is r. Hence, solving (7. 2) with respect to $\theta_1, \dots, \theta_r$, say, we get the following relations under the hypothesis H;

$$\theta_i = c_{i1}B_1^0 + \cdots + c_{ir}B_r^0 + c_{ir+1}\theta_{r+1} + \cdots + c_{is}\theta_s\,, \quad i = 1, 2, \dots, r\,, \tag{7.3}$$

where c_{ij} are known constants. Substituting (7. 3) into (7. 1), we have

$$m_i = d_{i1}B_1^0 + \cdots + d_{ir}B_r^0 + d_{ir+1}\theta_{r+1} + \cdots + d_{is}\theta_s\,, \quad i = 1, 2, \dots, n\,, \tag{7.4}$$

where the coefficients d_{ij} are also known constants.

Consequently, without any loss of generality, we can substitute the general linear hypothesis H by the hypothesis

$$H_0 : \theta_1 = \theta_1^0, \dots, \theta_r = \theta_r^0, \quad 0 < r \leqq s \tag{7.5}$$

Let x_i, $i = 1, 2, \dots, n$, be a random sample of size 1 drawn from the population $N(m_i, \sigma)$, $i = 1, 2, \dots, n$ respectively, then the frequency function of the joint distribution of $x_1, x_2, \dots, x_n$ is

$$(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a_{i1}\theta_1 - \cdots - a_{is}\theta_s)^2\right\},$$

Now, let it be

$$S = \sum_{i=1}^{n}(x_i - a_{i1}\theta_1 - \cdots - a_{is}\theta_s)^2, \tag{7.6}$$

then, as is well known, the maximum likelihood estimates $\hat{\theta}_1, \dots, \hat{\theta}_s$ of $\theta_1, \dots, \theta_s$ and the minimum value

$$S_0 = \sum_{i=1}^{n}(x_i - a_{i1}\hat{\theta}_1 - \cdots - a_{i1}\hat{\theta}_s)^2 \tag{7.7}$$

of S are stochastically independent. The joint distribution of $\hat{\theta}_1, \dots, \hat{\theta}_s$ is a k-dimensional normal distribution, and the variable S_0/σ^2 is distributed according to χ^2-distribution of degrees of freedom $(n-s)$, whether the null hypothesis H_0 be true or not.

Under the null hypothesis H_0, the frequency function of $x_1, \dots, x_n$ becomes

$$(2\pi\sigma^2)^{-n/2} \exp.\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a_{i1}\theta_1^0 - \cdots a_{ir}\theta_r^0 - a_{is+1}\theta_{r+1} - \cdots - a_{is}\theta_s)^2\right\}.$$

Let the maximum likelihood estimates of $\theta_{r+1}, \dots, \theta_s$ be $\hat{\theta}^*_{r+1}, \dots, \hat{\theta}^*_s$ respectively, and let the minimnm value of

$$S' = \sum_{i=1}^{n}(x_i - a_{i1}\theta_1^0 - \cdots - a_{ir}\theta_r^0 - a_{ir+1}\theta_{r+1} - \cdots - a_{is}\theta_s)^2 \tag{7.8}$$

be

$$S_0' = \sum_{i=1}^{n}(x_i - a_{i1}\theta_1^0 - \cdots - a_{ir}\theta_r^0 - a_{ir+1}\hat{\theta}^*_{r+1} - \cdots - a_{is}\hat{\theta}^*_s)^2, \tag{7.9}$$

then $\hat{\theta}^*_{r+1}, \dots, \hat{\theta}^*_s$ are jointly independent of S_0' in the sense of statistics. S_0'/σ^2 is distributed according to χ^2-distribution of degrees of freedom $(n-s+r)$. Hence under the null hypothesis H_0, $S_0' - S_0$ and S_0 are mutually independent, and consequently $(S_0' - S_0)/\sigma^2$ is distributed according to χ^2-distribution of degrees of freedom r.

Whence, it follows that, under the null hypothesis H_0, the statistic

$$F^r_{n-s} = \frac{n-s}{r} \frac{S'_0 - S_0}{S_0} \tag{7.10}$$

is distributed according to Snedecor's F-distribution with degrees of freedom $(r, n-r)$.

It is known that the test by means of statistic (7.10) is the most powerful one in a certain sense.[18)]

Thus far, we have assumed that $x_1, \dots, x_n$ are mutually independent. However, we can easily extend the above results to the case when $x_1, \dots, x_n$ are distributed according to non-singular n-dimensional normal distribution.

§ 8. **Tests of statistical hypotheses concerning unknown parameters of a normal population using systematic statistics.** We shall consider the tests of statistical hypotheses concerning unknown parameters of the normal population $N(m, \sigma)$, using the limiting distribution of k order statistics $x(n_1), \dots, x(n_k)$.

The frequency function of this limiting distribution is given by (3.14), i.e.,

$$h(x(n_1), \dots, x(n_k)) = (2\pi\sigma^2)^{-k/2} f_1 \dots f_k \Big[\lambda_1(\lambda_2-\lambda_1)\dots(1-\lambda_k)\Big]^{-1/2} n^{k/2} \times$$

$$\exp\Big[-\frac{n}{2\sigma^2}\Big\{\sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2\cdot(x(n_i)-m-u_i\sigma)^2$$

$$-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}}(x(n_i)-m-u_i\sigma)(x(n_{i-1})-m-u_{i-1}\sigma)\Big\}\Big]. \tag{8.1}$$

Case 1. We shall consider the test of *Student's hypothesis*

$$H_1 : m = m_0 \tag{8.2}$$

where m_0 is a certain specified value.

Let it be

$$S = \sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2\cdot(x(n_i)-m-u_i\sigma)^2$$

$$-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}}(x(n_i)-m-u_i\sigma)(x(n_{i-1})-m-u_{i-1}\sigma), \tag{8.3}$$

and determine $\hat{m}, \hat{\sigma}$, such that

$$\left.\frac{\partial S}{\partial m}\right|_{\substack{m=\hat{m}\\ \sigma=\hat{\sigma}}}=0\ ,\qquad \left.\frac{\partial S}{\partial \sigma}\right|_{\substack{m=\hat{m}\\ \sigma=\hat{\sigma}}}=0\ ,$$

i. e.,

$$\begin{aligned} K_1\hat{m}+K_3\hat{\sigma} &= X\ ,\\ K_3\hat{m}+K_2\hat{\sigma} &= Y\ , \end{aligned} \tag{8.4}$$

then m and $\hat{\sigma}$ are given by (5. 12). The minimum value S_0 of S is

$$S_0=\sum_{i=1}^{k+1}\frac{(f_i\cdot x(n_i)-f_{i-1}\cdot x(n_{i-1}))^2}{\lambda_i-\lambda_{i-1}}-(K_1\hat{m}^2+2K_3\hat{m}\hat{\sigma}+K_2\hat{\sigma}^2)\ , \tag{8.5}$$

and it is stochastically independent of $\hat{m}$, and $\hat{\sigma}$. S_0/σ^2 is distributed according to χ^2-distribution of degrees of freedom $(k-2)$. The above facts are valid independent of the null hypothesis H_1.

Let

$$\begin{aligned} S'=&\sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})}f_i^2\cdot(x(n_i)-m_0-u_i\sigma)^2\\ &-2\sum_{i=2}^{k}\frac{f_if_{i-1}}{\lambda_i-\lambda_{i-1}}(x(n_i)-m_0-u_i\sigma)(x(n_{i-1})-m_0-u_{i-1}\sigma)\ , \end{aligned} \tag{8.6}$$

and determine $\hat{\sigma}^*$ so as to satisfy the equation

$$\left.\frac{\partial S'}{\partial \sigma}\right|_{\sigma=\hat{\sigma}^*}=0\ ,$$

i. e.,

$$K_2\hat{\sigma}^*+K_3m_0=Y\ , \tag{8.7}$$

then the minimum value of S' becomes

$$\begin{aligned} S_0'=&\sum_{i=1}^{k+1}\frac{(f_ix\cdot(n_i)-f_{i-1}\cdot x(n_{i-1}))^2}{\lambda_i-\lambda_{i-1}}-2m_0(K_1\hat{m}+K_3\hat{\sigma})\\ &-\frac{\{K_3(\hat{m}-m_0)+K_2\hat{\sigma}\}^2}{K_2}+m_0^2K_1\ . \end{aligned} \tag{8.8}$$

Under the null hypothesis H_1, S_0' and $\hat{\sigma}^*$ are stochastically independent, and S_0'/σ^2 is distributed according to χ^2-distribution of degrees of freedom $(k-1)$. By the results of the general theory of § 7, we see that $S_0'-S_0$ and S_0 are stochastically independent, and $(S_0'-S_0)/\sigma^2$ is distributed according to χ^2-distribution of degree of freedom 1. Hence, the statistic

$$t=\sqrt{k-2}\sqrt{\frac{S_0'-S_0}{S_0}}=\sqrt{\frac{(k-2)\Delta}{K_2}}\frac{\hat{m}-m_0}{\sqrt{S_0}} \tag{8.9}$$

is distributed according to Student's t-distribution of degrees of freedom $(k-2)$.

In particular, if the spacing of $x(n_1), \dots, x(n_k)$ is symmetric, then $K_3=0$, hence (8.9) becomes

$$t=\sqrt{(k-2)K_1}\cdot\frac{\hat{m}-m_0}{\sqrt{S_0}}\,. \tag{8.10}$$

Whence the confidence interval of confidence coefficient $100(1-\alpha)\%$ for the true mean m is given by

$$\hat{m}-t_\alpha\sqrt{\frac{S_0}{(k-2)K_1}}<m<\hat{m}+t_\alpha\sqrt{\frac{S_0}{(k-2)K_1}}\,, \tag{8.11}$$

where t_α is the $100\alpha\%$ point of the t-distribution.

Case 2. *Generalized Student's hypothesis.—Test of the homogeneity of several means.* Let there be given s normal populations $N(m_\alpha, \sigma)$, $\alpha=1, \dots, s$ with common unknown variance σ^2 and unknown means m_α, $\alpha=1, \dots, s$, and let

$$x^{(\alpha)}(1), x^{(\alpha)}(2), \dots, x^{(\alpha)}(n)\,, \quad \alpha=1, 2, \dots, s\,,$$

be order statistics of size n, common to all populations, drawn from $N(m_\alpha, \sigma)$, $\alpha=1, \dots, s$ respectively. Further let

$$\lim_{n\to\infty}\frac{n_i}{n}=\lambda_i\,, \quad i=1, 2, \dots, k\,.$$

We shall deal with the test of homogeneity of means, i.e., the test of the null hypothesis

$$H_2:\ m_1=m_2=\cdots=m_s \tag{8.12}$$

utilizing the limiting distributions of

$$x^{(\alpha)}(n_1), \dots, x^{(\alpha)}(n_k)\,, \quad \alpha=1, 2, \dots, s\,.$$

Now let

$$S=\sum_{\alpha=1}^{k}\Big\{\sum_{i=1}^{s}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})}f_i^2\cdot(x^{(\alpha)}(n_i)-m_\alpha-u_i\sigma)$$
$$-2\sum_{i=2}^{k}\frac{f_i f_{i-1}}{\lambda_i-\lambda_{i-1}}(x^{(\alpha)}(n_i)-m_\alpha-u_i\sigma)(x^{(\alpha)}(n_{i-1})-m_\alpha-u_{i-1}\sigma)\Big\}, \tag{8.13}$$

and let $\hat{m}_1, \dots, \hat{m}_s, \hat{\sigma}$ be such that

$$\frac{\partial S}{\partial m_\alpha}\Big|_{\substack{m_1=\hat{m}_1\\ \cdots\cdots\\ m_s=\hat{m}_s\\ \sigma=\hat{\sigma}}}=0\,, \quad \alpha=1, \dots, s\,; \qquad \frac{\partial S}{\partial \sigma}\Big|_{\substack{m_1=\hat{m}_1\\ \cdots\cdots\\ m_s=\hat{m}_s\\ \sigma=\hat{\sigma}}}=0\,,$$

i. e.,

$$\left.\begin{array}{r}K_1\hat{m}_1 \qquad\qquad +K_3\hat{\sigma}=X_1\\ K_1\hat{m}_2 \qquad +K_3\hat{\sigma}=X_2\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ K_1\hat{m}_s+K_3\hat{\sigma}=X_s\\ K_3(\hat{m}_1+\cdots+\hat{m}_s)+sK_2\hat{\sigma}=Y_1+\cdots+Y_s\end{array}\right\} \tag{8.14}$$

where

$$\left.\begin{array}{l}X_\alpha=\sum_{i=1}^{k+1}\frac{(f_i-f_{i-1})(f_i\cdot x^{(\alpha)}(n_i)-f_{i-1}\cdot x^{(\alpha)}(n_{i-1}))}{\lambda_i-\lambda_{i-1}},\\ Y_\alpha=\sum_{i=1}^{k+1}\frac{(f_iu_i-f_{i-1}u_{i-1})(f_i\cdot x^{(\alpha)}(n_i)-f_{i-1}\cdot x^{(\alpha)}(n_{i-1}))}{\lambda_i-\lambda_{i-1}},\end{array}\right. \quad \alpha=1,2,\dots,s.$$

The minimum value of S becomes

$$S_0=\sum_{\alpha=1}^{s}\left\{\sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})}f_i^2\cdot(x^{(\alpha)}(n_i)-\hat{m}_\alpha-u_i\hat{\sigma})^2\right.$$
$$\left.-2\sum_{i=2}^{k}\frac{f_if_{i-1}}{\lambda_i-\lambda_{i-1}}(x^{(\alpha)}(n_i)-\hat{m}_\alpha-u_i\hat{\sigma})(x^{(\alpha)}(n_{i-1})-\hat{m}_\alpha-u_{i-1}\hat{\sigma})\right\}, \tag{8.15}$$

and this is stochastically independent of $\hat{m}_1, \dots, \hat{m}_s, \hat{\sigma}$. The variable S_0/σ^2 is distributed according to χ^2-distribution of degrees of freedom $(sk-s-1)$. This results ss independent of the null hypothesis H_2.

If the null hypothesis H_2 is true, we put

$$m_1=m_2=\dots=m_s=m ,$$

and let

$$S'=\sum_{\alpha=1}^{s}\left\{\sum_{i=1}^{k}\frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})}f_i^2\cdot(x^{(\alpha)}(n_i)-m-u_i\sigma)^2\right.$$
$$\left.-2\sum_{i=2}^{k}\frac{f_if_{i-1}}{\lambda_i-\lambda_{i-1}}(x^{(\alpha)}(n_i)-m-u_i\sigma)(x^{(\alpha)}(n_{i-1})-m-u_{i-1}\sigma)\right\}. \tag{8.16}$$

If we choose $\hat{m}^*, \hat{\sigma}^*$ such that

$$\left.\frac{\partial S'}{\partial m}\right|_{\substack{m=\hat{m}^*\\ \sigma=\hat{\sigma}^*}}=0, \quad \left.\frac{\partial S'}{\partial \sigma}\right|_{\substack{m=\hat{m}^*\\ \sigma=\hat{\sigma}^*}}=0,$$

i. e.,

$$\left.\begin{array}{l}sK_1\hat{m}^*+sK_3\hat{\sigma}^*=X_1+\cdots+X_s\\ sK_3\hat{m}^*+sK_2\hat{\sigma}^*=Y_1+\cdots+Y_s\end{array}\right\}, \tag{8.17}$$

then the minimum value S_0' of S'

$$S_0' = \sum_{\alpha=1}^{s} \left\{ \sum_{i=2}^{k} \frac{\lambda_{i+1}-\lambda_{i-1}}{(\lambda_{i+1}-\lambda_i)(\lambda_i-\lambda_{i-1})} f_i^2 \cdot (x^{(\alpha)}(n_i) - \hat{m}^* - u_i\hat{\sigma}^*)^2 \right.$$
$$\left. -2\sum_{i=2}^{k} \frac{f_i f_{i-1}}{\lambda_i - \lambda_{i-1}} (x^{(\alpha)}(n_i) - \hat{m}^* - u_i\hat{\sigma}^*)(x^{(\alpha)}(n_{i-1}) - \hat{m}^* - u_{i-1}\hat{\sigma}^*) \right\}, \quad (8.18)$$

is independent of $\hat{m}^*$, $\hat{\sigma}^*$ stochastically. S_r'/σ^2 is distributed according to χ^2-distribution of degrees of freedom $(sk-2)$ and $S_0'-S_0$ and S_0 are mutually independent. Since $(S_0'-S_0)/\sigma^2$ is distributed according to χ^2-distribution of degrees of freedom $(s-1)$, we have

$$F^{s-1}_{sk-s-1} = \frac{sk-s-1}{s-1} \frac{S_0'-S_0}{S_0} \quad (8.19)$$

is distributed according to Snedecor's F-distribution of degrees of freedom $(s-1,\ sk-s-1)$.

In the particular case when $s=2$, it is easily seen that

$$S_0' = \sum_{\alpha=1}^{2}\sum_{i=1}^{k+1} \frac{(f_i \cdot x^{(\alpha)}(n_i) - f_{i-1} \cdot x^{(\alpha)}(n_{i-1}))^2}{\lambda_i - \lambda_{i-1}} - \frac{1}{2} K_1(\hat{m}_1 + \hat{m}_2)^2$$
$$-2K_3 \cdot (\hat{m}_1 + \hat{m}_2)\hat{\sigma} - 2K_2\hat{\sigma}^2, \quad (8.20)$$

hence

$$S_0' - S_0 = \frac{K_1}{2}(\hat{m}_1 - \hat{m}_2)^2. \quad (8.21)$$

Consequently, (8.19) becomes

$$F^1_{2k-3} = (2k-3)\frac{S_0'-S_0}{S_0} = \frac{2k-3}{2} \cdot K_1 \cdot \frac{(\hat{m}_1-\hat{m}_2)^2}{S_0},$$

Taking square root, we have

$$t = \sqrt{\frac{2k-3}{2} \cdot K_1} \cdot \frac{\hat{m}_1 - \hat{m}_2}{\sqrt{S_0}}, \quad (8.22)$$

which is distributed according to Student's t-distribution of degrees of freedom $(2k-3)$. Whence we have the confidence interval of confidence coefficient $100(1-\alpha)\%$ for the true difference of means

$$\hat{m}_1 - \hat{m}_2 - t_\alpha\sqrt{\frac{2S_0}{(2k-3)K_1}} < m_1 - m_2 < \hat{m}_1 - \hat{m}_2 + t_\alpha\sqrt{\frac{2S_0}{(2k-3)K_1}}, \quad (8.23)$$

where t_α is the $100\alpha\%$ point of t-distribution of degrees of freedom $(2k-3)$.

§ 9. **Power functions of the tests mentioned in the preceding section and the optimum spacings for them.** In this section, we shall consider the power functions of such tests as defined by (8.10) and (8.22). Now let z be distributed according to $N(0, 1)$ and w be distributed according to χ^2-distribution of degrees of freedom f, and further it is assumed that they are mutually independent in the sense of statistics, then

$$t=\frac{z+\delta}{\sqrt{w/f}}, \quad \delta \neq 0 \tag{9.1}$$

is distributed according to the so-called "*non-central t-distribution*[19]" which has frequency function

$$\frac{\left(\frac{f}{2}\right)^{\frac{f}{2}} e^{-\frac{\delta^2}{2}}}{\sqrt{2\pi}\Gamma\left(\frac{f}{2}\right)} \cdot \sum_{\nu=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}(f+\nu+1)\right)}{\nu!} (\delta t)^\nu \left(\frac{2}{f+t^2}\right)^{\frac{f+\nu+1}{2}}. \tag{9.2}$$

Since the infinite series (9.2) can be integrated term by term, it follows that

$$P(|t| \leq t_\alpha)=e^{-\frac{\delta^2}{2}} \cdot \sum_{\nu=0}^{\infty} \frac{\left(\frac{\delta^2}{2}\right)^\nu}{\nu!} \cdot I\left(\frac{\nu+1}{2}, \frac{f}{2}; \frac{t_\alpha^2}{f+t_\alpha^2}\right), \tag{9.3}$$

where $1(p, q; x)$ denotes K. Pearson's incomplete Beta-function,[20] i.e.,

$$1(p, q; x)=\frac{\int_0^x t^{p-1}(1-t)^{q-1}dt}{\mathrm{B}(p, q)}.$$

If in case I, the null hypothesis H_1 is not true and some alternative hypothesis

$$H_1': \ m=m' \ (\neq m_0)$$

is true, then the statistic t in (8.10) becomes

$$t=\frac{\sqrt{\frac{\Delta}{K_2}}\frac{\hat{m}-m'}{\sigma}+\sqrt{\frac{\Delta}{K_2}}\frac{m'-m_0}{\sigma}}{\sqrt{\frac{S_0}{(k-2)\sigma^2}}}, \tag{9.4}$$

hence the distribution of (9.4) is the non-central t-distribution (9.3), where

$$\delta=\sqrt{\frac{\Delta}{K_2}} \cdot \frac{m'-m_0}{\sigma} \quad \text{and} \quad f=k-2. \tag{9.5}$$

If in case II $s=2$, the null hypothesis H_2 is not true, and some alternative hypothesis

$$H_2' : \ m_1 - m_2 = m' \ (\neq 0)$$

is true, then, the statistic t in (8.22) becomes

$$t = \frac{\sqrt{\frac{K_1}{2}} \frac{\hat{m}_1 - \hat{m}_2 - m'}{\sigma} + \sqrt{\frac{K_1}{2}} \frac{m'}{\sigma}}{\sqrt{\frac{S_0}{(2k-3)\sigma^2}}}, \tag{9.6}$$

hence the distribution of (9.6) is the non-central t-distribution (9.3) with

$$\delta = \sqrt{\frac{K_1}{2}} \cdot \frac{m'}{\sigma}, \quad \text{and} \quad f = 2k - 3 .$$

Consequently, in each case the power function is given by

$$P(|t| \geq t_\alpha) = 1 - e^{-\frac{\delta^2}{2}} \cdot \sum_{\nu=0}^{\infty} \frac{\left(\frac{\delta^2}{2}\right)^\nu}{\nu!} \cdot I\left(\frac{\nu+1}{2}, \frac{f}{2}; \frac{t_\alpha^2}{f + t_\alpha^2}\right), \tag{9.7}$$

where

$$\delta = \sqrt{\frac{\Delta}{K_2}} \cdot \frac{m' - m_0}{\sigma}, \quad f = k - 2 \text{ in case I,}$$

and

$$\delta = \sqrt{\frac{K_1}{2}} \cdot \frac{m'}{\sigma}, \qquad f = 2k - 3 \text{ in case II.}$$

Whence we can readily see that the test is more powerful according as δ becomes larger. In other words, the optimum spacings for testing hypotheses H_1 and H_2 are such that they make K_1 and Δ/K_2 maximum respectively. If we restrict ourselves to symmetric spacings, we have the optimum spacings when we take the ones which make K_1 maximum. Thus the spacings given in Table 6.1 are available.

(Received May 23, 1951)

Notes and References

1) See for example,
R. A Fisher; On the mathematical foundations of theoretical statistics, Phil. Trans. of the Roy. Soc. of London, Ser. A, Vol. 222, (1922) pp. 307–367.
R. A. Fisher; Theory of statistical estimation, Proc. of the Cambridge Phil. Soc., Vol. XXII, (1925) pp. 700–725.
J. Neyman and E. S. Pearson; On the problem of the most efficient tests of statistical hypotheses, Phil. Trans. of the R. Soc. of London, Vol. 231, (1933) pp. 289–337.
H. Cramér; Mathematical Methods of Statistics, Princeton, (1946) Chapts. 32, 33.

2) Frederick Mosteller; On some useful "inefficient" statistics, Ann. of Math. Statist., Vol. 17, No. 4, December (1946) pp. 377–408.

3) Ziro Yamanouchi; Estimates of mean and standard deviation of a normal distribution from linear combinations of some chosen order statistics, Bull. of Math. Stat. (edited by Research Association of Statistical Sciences) Vol. III, No. 1–2, (1949) pp. 52–57 (in Japanese with English abstract).

4) Wataru Ohsawa and Sumio Nagasawa; The statistico-physiological analysis of the vital resistibilities to water, kerosene, and pyrethrin of the workers of Cremastogaster brunnea matsumurai Forel, Axel R. Elfströms Bokfryckeri A.–B., Stockholm (1950) pp. 1–22.

5) For details of this section, see H. Cramér; loc. cit., Chapt. 32.

6) R. A. Fisher; Design of Experiments, (1949), Chapt. XI.

7) The outline of the proof of this proposition is as follows: From (1.22), we have

$$\prod_{i=1}^{n} f(x_i\,;\alpha,\beta)\cdot|J| = g(\alpha^*,\beta^*\,;\alpha,\beta)\cdot h(\xi_1,\dots,\xi_{n-2}|\alpha^*,\beta^*\,;\alpha,\beta)\,,$$

hence it follows that

$$\sum_{i=1}^{n}\log f(x_i\,;\alpha,\beta)+\log|J| = \log g+\log h\,. \tag{i}$$

Taking the total differential of both sides of (i), and squaring and averaging, we of get

$$n\int_{-\infty}^{\infty}\left(\frac{\partial\log f}{\partial\alpha}d\alpha+\frac{\partial\log f}{\partial\beta}d\beta\right)^2 f dx \geqq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{\partial\log g}{\partial\alpha}d\alpha+\frac{\partial\log g}{\partial\beta}d\beta\right)^2 g d\alpha^* d\beta^*. \tag{ii}$$

The sign of equalty in (ii) holds when and only when

$$\frac{\partial h}{\partial\alpha}=\frac{\partial h}{\partial\beta}=0\,,$$

i. e., the conditional frequency function $h(\xi_1,\dots,\xi_{n-2}|\alpha^*,\beta^*\,;\alpha,\beta)$ is independent of α and β.

Since the inequality (ii) holds for all values of $d\alpha$ and $d\beta$, denoting u and v the indeterminate variables corresponding to $d\alpha$ and $d\beta$, (ii) is equivalent to the following inequality

$$n\left[E\left(\frac{\partial\log f}{\partial\alpha}\right)^2\cdot u^2+2E\left(\frac{\partial\log f}{\partial\alpha}\cdot\frac{\partial\log f}{\partial\beta}\right)\cdot uv+E\left(\frac{\partial\log f}{\partial\beta}\right)^2\cdot v^2\right]$$
$$\geqq E\left(\frac{\partial\log g}{\partial\alpha}\right)^2\cdot u^2+2E\left(\frac{\partial\log g}{\partial\alpha}\cdot\frac{\partial\log g}{\partial\beta}\right)\cdot uv+E\left(\frac{\partial\log g}{\partial\beta}\right)^2\cdot v^2\,. \tag{iii}$$

Now, since α^* and β^* are unbiased estimates of α and β respectively, it can be seen after some easy calculations that for two pairs of undeterminate variables u, v and ξ, η the following equation hold.

$$(\xi u+\eta v)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big[\xi(\alpha^*-\alpha)+\eta(\beta^*-\beta)\Big]\Big(\frac{\partial \log g}{\partial \alpha}u+\frac{\partial \log g}{\partial \beta}v\Big)g\,d\alpha^* d\beta^* . \quad \text{(iv)}$$

By means of Schwarz's inequality, we have

$$(\xi u+\eta v)^2 \leqq$$
$$\Big[\xi^2 E(\alpha^*-\alpha)^2+2\xi\eta E(\alpha^*-\alpha)(\beta^*-\beta)+\eta^2 E(\beta^*-\beta)^2\Big]\times$$
$$\Big[u^2 E\Big(\frac{\partial \log g}{\partial \alpha}\Big)^2+2uvE\Big(\frac{\partial \log g}{\partial \alpha}\frac{\partial \log g}{\partial \beta}\Big)+v^2 E\Big(\frac{\partial \log g}{\partial \beta}\Big)^2\Big] . \quad \text{(v)}$$

The sign of equality in (v) holds when and only when $d(\log g)$ is proportional to $\xi(\alpha^*-\alpha)+\eta(\beta^*-\beta)$.

Now, we shall prove the following algebraic lemma:

Lemma: Let $A(\xi, \eta)$ and $B(u, v)$ be positive definite quadratic forms, and the inequality

$$A(\xi, \eta)\cdot B(u, v) \geqq (\xi u+\eta v)^2 \quad \text{(vi)}$$

holds for all values of ξ, η and u, v, then

$$A(\xi, \eta) \geqq B^{-1}(\xi, \eta) , \quad \text{(vii)}$$

where $B^{-1}(\xi, \eta)$ denotes the reciprocal form of $B(\xi, \eta)$.

Proof: By suitably chosen orthogonal transformation

$$u = p_{11}u'+p_{12}v' , \quad v = p_{21}u'+p_{22}v' , \quad \text{(viii)}$$

transforms the matrix B into a diagonal form, and let its eigenvalues be

$$\lambda_1 > 0 , \quad \lambda_2 > 0 ,$$

then (vi) becomes

$$A(\xi, \eta)\cdot(\lambda_1 u'^2+\lambda_2 v'^2) \geqq \Big\{(p_{11}u'+p_{12}v')\xi+(p_{21}u'+p_{22}v')\eta\Big\}^2 . \quad \text{(ix)}$$

Rearranging (ix) as quadratic form in u', v', we have

$$\Big\{\lambda_1 A(\xi, \eta)-(p_{11}\xi+p_{21}\eta)^2\Big\}u'^2-2(p_{11}\xi+p_{21}\eta)(p_{12}\xi+p_{22}\eta)u'v'$$
$$+\Big\{\lambda_2 A(\xi, \eta)-(p_{12}\xi+p_{22}\eta)^2\Big\}v'^2 \geqq 0 , \quad \text{(x)}$$

Hence the matrix

$$\begin{bmatrix} \lambda_1\cdot A(\xi, \eta)-(p_{11}\xi+p_{21}\eta)^2 & -(p_{11}\xi+p_{21}\eta)(p_{12}\xi+p_{22}\eta) \\ -(p_{11}\xi+p_{21}\eta)(p_{12}\xi+p_{22}\eta) & \lambda_2\cdot A(\xi, \eta)-(p_{12}\xi+p_{22}\eta)^2 \end{bmatrix}$$

must be positive definite for all values of ξ, η. Whence we have

$$A(\xi, \eta) \geqq \frac{1}{\lambda_1}(p_{11}\xi+p_{21}\eta)^2+\frac{1}{\lambda_2}(p_{12}\xi+p_{22}\eta)^2 ,$$

i. e.,

$$A(\xi, \eta) \geqq B^{-1}(\xi, \eta)$$

q. e. d.

Applying the above lemma to (v), we have

$$\frac{1}{1-\rho^2}\left(\frac{u^2}{\sigma_1^2}-2\rho\frac{uv}{\sigma_1\sigma_2}+\frac{v^2}{\sigma_2^2}\right)\leqq E\left(\frac{\partial \log g}{\partial \alpha}\right)^2\cdot u^2+2E\left(\frac{\partial \log g}{\partial \alpha}\cdot\frac{\partial \log g}{\partial \beta}\right)\cdot uv+E\left(\frac{\partial \log g}{\partial \beta}\right)^2\cdot v^2 . \tag{xi}$$

From (iii) and (xi), it follows the proposition.

8) F. N. David and J. Neyman; Extension of the Markoff's theorem on least squares, Statistical Research Memoirs, Vol. 1, (1936) pp. 105–116.

9) For the proof of this theorem, see for example, F. N. David and J. Neyman, loc. cit. and J. Ogawa, Note on the Markoff's theorem on least squares, Osaka Math. Journ., Vol. 2, No. 2, (1950) pp. 145–150.

10) M. Masuyama; Note on a Markoff's theorem on least squares, Kokyuroku of the Inst. Statist. Math. Vol. 4, No. 11, (1948) pp. 430–432. (in Japanese).

11) H. Cramér; loc. cit. p. 368.

12) F. Mosteller; loc. cit. p. 383

13) F. Mosteller; loc. cit. p. 380.

14) It is easily seen that when $(u_1, \dots, u_k)$ is a system of solutions of (6.2), then $(-u_1, \dots, -u_k)$ is also a system of solution. Hence, if we can prove the uniqueness of the solution of (6.2), then our conjecture is answered in the affirmative.

15) For example, when $k=5$, assume that (6.5) has a system of symmetric solution, then

$$2u_3^2-2+\frac{f_3u_3-f_2u_2}{\lambda_3-\lambda_2}+\frac{f_4u_4-f_3u_3}{\lambda_4-\lambda_3}=0 ,$$

where $u_3=0$, $u_2+u_4=0$, $f_2=f_4$ and

$$\lambda_4-\lambda_3=1-\lambda_2-\frac{1}{2}=\frac{1}{2}-\lambda_2 ,$$

$$\lambda_3-\lambda_2=\frac{1}{2}-\lambda_2 ,$$

hence we have

$$\frac{f_4u_4-0}{\lambda_4-\lambda_3}=1 .$$

This means geometrically that the chord joining two points on the curve in Fig. 6.2 corresponding to λ_3 and λ_4 has the gredient equal to I. Hence, from the convexity if the curve, there should be a point between λ_3 and λ_4, for which $\frac{dfu}{d\lambda}=1$. However, on the other hand

$$\frac{dfu}{d\lambda}=1-u^2<1 ,$$

hence the assumption that (6.5) has a system of symmetric solutions must be false.

16) Mr. Maruyama kindly informed the author that if we specify the numbers of negative u's and positive u's respectively, then the uniqueness of the solution of (6.5) seems to be almost certain, but he could not prove it logically.

17) For details of this section, see for example,

(i) S. S. Wilks; Mathematical Statistics, Princeton (1943) Chapt. 9.

(ii) S. Kołodziejczyk; On an important class of statistical hypotheses, Biometrika, Vol. 27, (1935) pp. 161–190.

(iii) P. C. Tang; The power function of the analysis of variance tests with tables and illustrations of their use, Statist. Research Memoirs, Vol. II, (1937) pp. 126–159.

(iv) Palmer O. Johnson and J. Neyman; Tests of certain linear hypotheses and their application to some educational problems, Statist. Research Memoirs, Vol. I, (1936) pp. 57–93.

(v) J. Ogawa; Normal regression theory and its applications 1, Kokyuroku of the Inst. of Statist. Math. Vol. 3, No. 21–22, (1948) pp. 374–396. (in Japanese).

(vi) G. Elfving; A simple method of deducing certain distributions connected with multivariate Sampling, Skandinavisk Aktuarietidskrift (1947) pp. 56–74

18) S. Kołodziejczyk, loc. cit.

19) J. Neyman and B. Tokarska; Errors of the second kind in testing "Student's" hypothesis, Journ. of the Amer. Statist. Assoc., Vol. 31, (1936) pp. 318–326.

N. L. Johnson and B. L. Welch; Applications of non-central t–distribution, Biometrika, Vol. 31, (1940) pp 362–389.

Cecil C. Craig; Note on the distribution of non-central t with application, Annals of Math. Statist., Vol. 12 (1941) pp. 224–228.

20) Tables of the Incomplete Beta-Function, Edited by Karl Pearson, F. R. S. Published by the "Biometrika" Office, University College, London (1934).