ANNALES
UNIVERSITATIS MARIAE CURIE-SKLODOWSKA
LUBLIN — POLONIA
VOL. III, 3 SECTIO A 1949

Z Zakładu Statystyki Matematycznej Wydziału Matematyczno-Przyrodniczego U. M. C. S.
Kierownik: z. prof. dr M. Olekiewicz

M. Olekiewicz

On the Efficiency of Biased Estimates

O wydajności ocen obciążonych

In this paper the term „closeness" is used to denote concentration of sampling values of a statistic, α^*, expressed as a function of observations in a sample of n ($\alpha^* = \alpha^* (x_1, x_2, \dots x_n)$), around the estimated parameter α. This concentration can be determined by means of the complementary property dispersion, as measured either by mean square error of estimate

$$\mathcal{E}^2 (\alpha^*) = E (\alpha^* - \alpha)^2, \qquad [A]$$

or, in an alternative and more natural way, by mean absolute error of estimate

$$\epsilon (\alpha^*) = E |\alpha^* - \alpha|. \qquad [B]$$

Inasmuch as the latter measure has a limited tractability, the former will be generally taken for criterion of „goodness" of a given estimate when compared with other estimates of the same parameter. It will be called A-criterion. In more tractable cases, however, B-criterion, i. e., mean absolute error of estimate, will be preferred. In cases of divergent determination of „better estimates" by the two criteria, the B-determination will be taken for „ultimately better". We could always speak, of course, of one estimate as „better in A-sense", and of the other as „better in B-sense", but since we shall be using mostly A-criterion, the above qualification will not be made explicit, even if tacitly understood. Thus by „better estimates" such statistics will be generally meant for which mean square error of estimate is smaller, and by „best estimates" — such statistics for which mean square error of estimate is least.

The above principle of least mean square error differs from the one advocated by R. A. Fisher, who restricts „good estimates" to unbiased estimates exclusively. The latter principle, now in common use, postulates:

1. that estimate should be unbiased;
2. that of all unbiased estimates the one with minimum sampling variance should be chosen.

Since mean square error of estimate can be resolved into two components

$$\mathcal{E}^2(a^*) = D^2(a^*) + b^2(a^*), \tag{1}$$

where $D^2(a^*)$ is sampling variance of a^* and $b(a^*)$ its bias defined by

$$b(a^*) = E(a^*) - a, \tag{2}$$

it can be seen that Fisher's principle is based on a weaker criterion than that of least mean square error of estimate, for instead of postulating the smallest possible value for the whole expression [1] it postulates the smallest possible values for its two components separately ($b^2(a^*) = 0$ and $D^2(a^*) = \text{minimum}$).

In this respect his principle is akin to the less comprehensive principle of least squares, as well as to the principle of Maximum Likelihood, both of which are based on weaker criteria than that of least error of estimate.

In fact, the principle of least squares postulates that the „best" estimate should be taken as that for which the sum of squared deviations of the observed values is least. As long as the meaning of „an observed value" implies that it is an unbiased estimate of the „true value", the principle of least squares is limited to unbiased estimates by definition. If we extend this principle to sampling values of a statistic a^*, we shall have to minimize $E(a^* - a)^2$ *with respect to* a instead of with respect to a^* as it should be done by the principle of least mean square error of estimate. The solution, $a = E(a^*)$, illustrates the real meaning of the principle of least squares, which is to assign for estimation the optimum parameter function, $E(a^*)$, to a given statistic, a^*, rather, than to restrict statistics as estimates of a given parameter. To take an example, $E(s^2) = \frac{n-1}{n} \cdot \sigma^2$, according to this interpretation, should be considered best parameter function to be estimated by $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, and similarly, $E(\frac{n}{n-1} \cdot s^2) = \sigma^2$ should be considered best parameter to be estimated by $\frac{n}{n-1} \cdot s^2$. But this is not equivalent to saying that $\frac{n}{n-1} \cdot s^2$ is a better estimate of σ^2 than other statistics could possibly be.

The principle of Maximum Likelihood postulates that the „best" estimate should be taken as that for which the probability density at the sample point $(x_1, x_2, \dots x_n)$, or the probability of obtaining the given sample (depending on whether X is a continuous or a discrete variable), is a maximum *with respect to* α. The principle is an extension of the idea of choosing the most probable value as the best guess at the „true value". This may have a strong intuitive appeal, but cannot be defended on purely logical grounds, since the position of the „true value" must not necessarily coincide with the most frequent one, or be closer to it than to any other calculable value. The great merit of the method of Maximum Likelihood lies in the fact that its solutions are often given in the form of not quite obvious functions of $x_1, x_2 \dots x_n$ which, once discovered, can be used as basic functions in estimation. The principle itself is rarely used in its pure form, since its solutions are corrected for bias. Thus in practical application the principle of Maximum Likelihood is combined with the principle of unbiased estimates. From the point of view of our criterion, as it will be seen later, Maximum Likelihood solutions should be corrected in the opposite direction, which, making them still more biased, will diminish mean square error of estimate.

When speaking of „goodness of estimate" in Fisher's sense, we shall use, following Cramér, the qualification „unbiased". To be explicit, the term „efficient estimate" by which Cramér means an unbiased estimate $\tilde{\alpha}$ with $\mathcal{E}^2$ equal to the minimum $\mathcal{E}^2$ for unbiased estimates we shall substitute by the term „efficient unbiased estimate".

Subject to certain general conditions of regularity[1]) it is possible to determine the minimum value of $\mathcal{E}^2$ for statistics possessing bias of a fixed magnitude. Two inequalities determine this bottom value, one for the case of continuous type, and the other for the case of discrete type[2]). The first is

$$\mathcal{E}^2(\alpha^*) \geqslant \frac{\left(1 + \frac{d\,b(\alpha^*)}{d\alpha}\right)^2}{n \int\limits_{-\infty}^{\infty} \left(\frac{\partial \log f(x)}{\partial \alpha}\right)^2 f(x)\,dx}, \qquad [3]$$

[1]) Cf. Cramér H. Mathematical Methods of Statistics, sections 32.2 and 32.3.
[2]) Ibid. pp. 478—483 and 486—487.

where $f(x)$ is probability density of value x of random variable X; and the second is

$$\mathcal{E}^2(\alpha^*) \geqslant \frac{\left(1+\dfrac{d\,b(\alpha^*)}{d\alpha}\right)^2}{n\sum\limits_i \left(\dfrac{d\log p_i}{d\alpha}\right)^2 p_i} \qquad [4]$$

where p_i is probability of value i of random variable X.

For unbiased estimates, $\hat{\alpha}$'s, the bottom value of $\mathcal{E}^2(\hat{\alpha})$ is, for the case of continuous type

$$\mathcal{E}_0^2(\hat{\alpha}) = D_0^2(\hat{\alpha}) = \frac{1}{n\int\limits_{-\infty}^{\infty}\left(\dfrac{\partial \log f(x)}{\partial\alpha}\right)^2 f(x)\,dx}, \qquad [5]$$

and for the case of discrete type

$$\mathcal{E}_0^2(\hat{\alpha}) = D_0^2(\hat{\alpha}) = \frac{1}{n\sum\limits_i \left(\dfrac{d\log p_i}{d\alpha}\right)^2 p_i}, \qquad [6]$$

Under rather restricted conditions there exist unbiased estimates $\tilde{\alpha}$'s with sampling variance equal to $\mathcal{E}_0^2(\alpha) = D_0^2(\hat{\alpha})$. When they do, they can be always found by the method of Maximum Likelihood.

Generally, for estimates with bias equal to b, say, we shall have for the bottom value of $\mathcal{E}^2(\alpha_b^\bullet)$

$$\mathcal{E}_0^2(\alpha_b^\bullet) = D_0^2(\hat{\alpha})\cdot\left(1+\frac{db}{d\alpha}\right)^2. \qquad [7]$$

If bias can be expressed as a linear function of α, as it is generally the case:

$$b = a\alpha,\ \frac{db}{d\alpha} = a,\ \mathcal{E}(\alpha^*) = (1+a)\,\alpha,\ \alpha^* = (1+a)\,\hat{\alpha},\ \text{say}, \qquad [8]$$

then, according to [7], the bottom value of $\mathcal{E}^2(\alpha_{a\alpha}^\bullet)$ should be

$$\mathcal{E}_0^2(\alpha_{a\alpha}^\bullet) = D_0^2(\hat{\alpha})\cdot(1+a)^2. \qquad [9]$$

It can be easily shown howewer that, when [8] *is satisfied, there exist no biased estimates with* $\mathcal{E}^2$ *equal to the bottom value given in* [9]. *The inequalities* [3] *and* [4] *can be improved then, and the optimum bias* $b_0 = a_0\,\alpha$ *found, such that the estimate* $(1+a_0)\,\tilde{\alpha}$, *when it exists (which depends on the existance of* $\tilde{\alpha}$), *will be the „best“,* *that is, „closest" out of all possible linearly biased and unbiased estimates of α.* We shall call it „linearly efficient estimate" of α.

To show that $\mathcal{E}^2(\overset{*}{\alpha}_{a\alpha}) > \mathcal{E}_0{}^2(\overset{*}{\alpha}_{a\alpha})$, when $a \neq 0$, we shall write

$$\mathcal{E}^2(\overset{*}{\alpha}_{a\alpha}) = \mathcal{E}^2[(1+a)\hat{\alpha}] = D^2[(1+a)\hat{\alpha}] + a^2\alpha^2 = (1+a)^2\ D^2(\hat{\alpha}) + a^2\alpha^2. \quad [10]$$

Then $$\mathcal{E}^2(\alpha^*_{a\alpha}) - \mathcal{E}_0{}^2(\alpha^*_{a\alpha}) = (1+a)^2\,(D^2(\hat{\alpha}) - D_0{}^2(\hat{\alpha})) + a^2\alpha^2 > 0, \quad [11]$$
what was to be shown.

To find the optimum bias, we first minimize $\mathcal{E}^2(\alpha_{a\alpha}{}^*)$ with respect to a, and obtain solution

$$a_0 = -\frac{D^2(\hat{\alpha})}{\alpha^2 + D^2(\hat{\alpha})} = -\frac{D^2(\alpha^*)}{E^2(\alpha^*) + D^2(\alpha^*)},\quad 1 + a_0 = \frac{\alpha^2}{\alpha^2 + D^2(\hat{\alpha})} =$$

$$= \frac{E^2(\alpha^*)}{E^2(\alpha^*) + D^2(\alpha^*)}. \quad [12]$$

Thus, for any unbiased estimate, $\hat{\alpha}$, solution [12] determines the optimum linear function, $(1+a_0)\hat{\alpha}$, as well as for any linearly biased estimate, α^*, the optimum linear function, $\frac{1+a_0}{1+a}\cdot\alpha^*$, for estimating α

$$\alpha_0{}^* = (1+a_0)\,\hat{\alpha} = \frac{\alpha^2}{\alpha^2 + D^2(\hat{\alpha})}\cdot\hat{\alpha} = \frac{1+a_0}{1+a}\cdot\alpha^* = \frac{\alpha E(\alpha^*)}{E^2(\alpha^*) + D^2(\alpha^*)}\cdot\alpha^*\ . \quad [13]$$

Mean square error of estimate for this function is

$$\mathcal{E}^2(\alpha_0{}^*) = \mathcal{E}^2(\alpha^*_{a_0\alpha}) = \frac{\alpha^2 D^2(\hat{\alpha})}{\alpha^2 + D^2(\hat{\alpha})} = \frac{\alpha^2 D^2(\alpha^*)}{E^2(\alpha^*) + D^2(\alpha^*)}\ . \quad [14]$$

It can be seen from [14] that minimum $\mathcal{E}^2(\alpha_0{}^*)$ will be obtained when $D^2(\hat{\alpha}) = D_0{}^2(\hat{\alpha})$, that is, when $\hat{\alpha}$ is $\tilde{\alpha}$. Thus the linearly efficient estimate of α will be given by

$$\overset{*}{\alpha} = \frac{\alpha^2}{\alpha^2 + D_0{}^2(\hat{\alpha})}\cdot\tilde{\alpha}\ . \quad [15]$$

Its mean square error of estimate is

$$\mathcal{E}^2(\overset{*}{\alpha}) = \min.\ \mathcal{E}^2(\alpha^*{}_{\text{linear } b}) = D_0{}^2(\hat{\alpha})\cdot\frac{\alpha^2}{\alpha^2 + D_0{}^2(\hat{\alpha})}\ . \quad [16]$$

Thus the unattainable limit of $\mathcal{E}^2(\underset{a_1}{a^*})$ given in [9] can be substituted now by the attainable bottom value given in [16], and inequalities [3] and [4] by the improved inequalities

$$\mathcal{E}^2(a^*_{\text{linear } b}) > \frac{a^2}{1+a^2 n \int\limits_{-\infty}^{\infty} \left(\frac{\partial \log f(x)}{\partial a}\right)^2 f(x)dx}, \qquad [17]$$

and

$$\mathcal{E}^2(a^*_{\text{linear } b}) > \frac{a^2}{1+a^2 n \sum\limits_i \left(\frac{d \log p_i}{da}\right)^2 p_i}, \qquad [18]$$

for the cases of continuous and discrete types respectively.

When linearly efficient estimate does not exist, the solution [12] can be used for determining a relatively best linear estimate derived from such a statistic, for which $\frac{D^2(a^*)}{E^2(a^*)+D^2(a^*)}$ is smaller than for any other known statistic (cf. [14])[1]). Denoting such a statistic by $a^\cdot$, and relatively best linear estimate by $a_0^\cdot$, we shall have

$$a_0^\cdot = \frac{aE(a^\cdot)}{E^2(a^\cdot)+D^2(a^\cdot)}\, a^\cdot . \qquad [19]$$

The expression for its mean square error of estimate can be written from [14].

When there exists a sufficient joint unbiased estimate, $\check{a}$, the best linear estimate derived from it will be called „sufficient linear estimate of a with other parameters unknown"

$$\dot{a} = \frac{a^2}{a^2+D^2(\check{a})} \cdot \check{a} . \qquad [20]$$

It can be seen that linearly efficient estimate of σ^2 in normal population is, according to [15]

$$\overset{*}{\sigma}{}^2 = \frac{n}{n+2} \cdot s_0^2 \qquad [21]$$

(since $\tilde{\sigma}^2$ exists and is equal to $s_0^2 = \frac{1}{n}\sum\limits_{i=1}^{n} (x_i - \mu)^2$,

and $D^2(s_0^2) = D_0^2(\hat{\sigma}^2) = \frac{2\sigma^4}{n}$).

---

[1]) The functional form of such a statistic can be often determined by the method of Maximum Likelihood.

Mean square error of $\overset{*}{\sigma}{}^2$ is, according to [16]

$$\mathcal{E}^2(\overset{*}{\sigma}{}^2) = \frac{2\sigma^4}{n+2} \cdot \qquad [22]$$

Similarly, the optimum linear function of s^2 for the estimation of σ^2 in normal population is, according to [13]

$$\sigma_0^{2*} = \frac{1+a_0}{1+a} \cdot s^2 = \frac{n}{n+1} \cdot s^2 \qquad [23]$$

with

$$\mathcal{E}^2(\sigma_0^{2*}) = \frac{2\sigma^4}{n+1} \qquad [24]$$

(since for normal population $D^2(s^2) = \frac{2(n-1)\sigma^4}{n^2}$) .

On the basis of Maximum Likelihood solution for joint estimation of μ and σ^2 in normal population, $\bar{x}$ and $\frac{n}{n-1} \cdot s^2$ are found to be sufficient unbiased joint estimates of these parameters. It follows then from [20] that we can take $\frac{n}{n+1} \cdot s^2$ for sufficient linear estimate of σ^2, with μ unknown, writing instead of [23]

$$\dot{\sigma}^2 = \frac{n}{n+1} \cdot s^2 \qquad [25]$$

with mean square error given in [24].

From the foregoing it can be seen that as an alternative to Fisher's conception of efficiency based on minimum variance for unbiased estimates a different conception can be adduced based on minimum square error of linearly biased estimates. The efficiency in this sense may be called *linear efficiency*. Its measure will be given by the following considerations.

If by using *n* observations we obtain an estimate of such precision („closeness“) which can be attained by an efficient estimate based on n' observations ($n' < n$), then it cannot be said that we make full use of the available information, but it can be said instead that the use we make of *n* observations by employing such an estimate is equivalent to the full use of n'. The proportion, $\frac{n'}{n}$, will be thus considered a measure of efficiency of the employed estimate. This measure will depend on our determination of precision („closeness“). If [*A*] is chosen for this determination, then we shall have linear efficiency in A-sense, if [B], then in B-sense. Saying simply „linear efficiency", we shall understand efficiency in A-sense. Thus linear officiency will be defined by

$$e_l(a_n^{\bullet}) = \frac{n'}{n}, \tag{26}$$

where $e_l(a_n^{\bullet})$ stands for linear efficiency of a linearly biased (or unbiased) estimate a^* based on n observations, and n'— for number of observations sufficient and necessary for obtaining a linearly efficient estimate of the same precision.

To determine n' (for the efficiency in A-sense) we shall write the equation

$$\mathcal{E}^2(a_n^{\bullet}) = \min\ \mathcal{E}^2(a^{\bullet}_{n'\,\text{linear}\,b}) = \frac{a^2 D_0^2(\hat{a}_{n'})}{a^2 + D_0^2(\hat{a}_{n'})}. \tag{27}$$

For instance, in estimating σ^2 in normal population we shall write

$$\mathcal{E}^2(\sigma_n^{2\,*}) = \frac{2\sigma^4}{n'+2}. \tag{28}$$

Taking $\frac{n}{n-1}\cdot s^2$ for σ_n^{2*}, we shall have

$$\frac{2\sigma^4}{n-1} = \frac{2\sigma^4}{n'+2}. \tag{29}$$

Solving for n', we obtain

$$n' = n-3, \tag{30}$$

and

$$e_l(\hat{\sigma}_n^2) = e_l\left(\frac{n}{n-1}\cdot s^2\right) = \frac{n-3}{n}. \tag{31}$$

Similarly we find

$$e_l(\tilde{\sigma}_n^2) = e_l(s_0^2) = \frac{n-2}{n}. \tag{32}$$

$$e_l(\dot{\sigma}_n^2) = e_l\left(\frac{n}{n+1}\cdot s^2\right) = \frac{n-1}{n}. \tag{33}$$

$$e_l(s^2) = \frac{2(n-1)^2}{n(2n-1)} > e_l(s_0^2)\,! \tag{34}$$

The efficiency in Fisher's sense can be obtained from [26] by putting $D_0^2(\hat{a}_{n'})$ in [27] in place of min $\mathcal{E}^2(a^{\bullet}_{n'}$ linear $b)$. The results will be identical with those calculated in accordance with Fisher's[1]), or, for that matter, with **Kendall's**[2]), definition of efficiency

$$e(\hat{a}_n) = \frac{D_0^2(\hat{a}_n)}{D^2(a_n)}, \qquad [35]$$

as long as denominator in the expressions for $D^2_0(\hat{a}_n)$ remains equal to n without any constant number added. For this reason (see [22]) the analogous identity of $\frac{n'}{n}$ with $\frac{min\ \mathcal{E}^2(a_n^{\bullet}\ linear\ b)}{\mathcal{E}^2(a_n^{\bullet})}$ generally does not exist, and therefore linear efficiency cannot be expressed as ratio of minimum $\mathcal{E}^2$ to mean square error of a given estimate.

It can be seen that some estimates which in Fisher's sense are more efficient are not so in our sense, and vice versa. The prevailing practice of correcting s^2 for bias is from our point of view unwarranted, inasmuch as s^2 is a „closer" estimate of σ^2 than $\frac{n}{n-1}\cdot s^2$ is. As it has been already said, the indicated correction would be in different direction: s^2 to be multiplied by $\frac{n}{n+1}$ rather than by $\frac{n}{n-1}$. From the point of view of estimation we do not see any special reason, except expediency, for preferring unbiased estimates to biased, once mean square error of estimate accounts for all errors, constant including. The correcting of an individual value for constant error cannot be logically distinguished from correcting it for variable error. When correcting for constant error introduces a still larger variable error, we cannot help but feel that such correction makes the estimate more unreliable. It is when the combining of individual values is concerned that constant error assumes importance. In such cases, however, as for instance in estimating variance, the knowledge of sample sizes is all that is needed to calculate a combined biased estimate that will be „closer" to the parameter in question than any unbiased combined estimate.

In fact, applying weights $w_1, w_2, \ldots w_k$ to individual linear estimates of a, $a_1^{\bullet}, a_2^{\bullet}, \ldots a_k^{\bullet}$, taken from independent samples of

[1]) **Fisher** R. A. „Statistical Methods for Research Workers", p. 312. Oliver and Boyd, London, 1948. (On page 13, though, Fisher gives an alternative definition of efficiency which is essentially the same as ours, when restricted to unbiased estimates).

[2]) **Kendall**, M. G. „The Advanced Theory of Statistics" vol. 2, p 6. Griffin a Co, London, 1948.

sizes n_1, $n_2, \ldots n_k$, such that combined estimate $\sum_{i=1}^{k} w_i a_i^*$ be the „closest‘ possible, we shall obtain from

$$\mathcal{E}^2\left(\sum_{i=1}^{k} w_i a_i^*\right) = \sum_{i=1}^{k} w_i^2 D^2(a_i^*) + \left(\sum_{i=1}^{k} w_i E(a_i^*) - a\right)^2 = \text{minimum} \qquad [36]$$

a system of solutions

$$w_i = \frac{E(a_i^*)}{D^2(a_i^*)} \cdot \frac{a}{1 + \sum_{i=1}^{k} \frac{E^2(a_i^*)}{D^2(a_i^*)}}, \quad i = 1, 2, \ldots k. \qquad [37]$$

Mean square error of so combined estimate will be given by

$$\mathcal{E}^2\left(\sum_{i=1}^{k} w_i a_i^*\right) = \frac{a^2}{1 + \sum_{i=1}^{k} \frac{E^2(a_i^*)}{D^2(a_i^*)}} \cdot \qquad [38]$$

When $a = \sigma^2$, and

$$a_i^* = s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 ,$$

then

$$w_i = \frac{n_i - 1}{n_i D^2(s_i^2)} \cdot \frac{\sigma^4}{1 + \sigma^4 \sum_{i=1}^{k} \frac{(n_i - 1)^2}{n_i^2 D^2(s_i^2)}}, \qquad [39]$$

and

$$\mathcal{E}^2\left(\sum_{i=1}^{k} w_i s_i^2\right) = \frac{\sigma^4}{1 + \sigma^4 \sum_{i=1}^{k} \frac{(n_i - 1)^2}{n_i^2 D^2(s_i^2)}} \cdot \qquad [40]$$

-If population is normal, then

$$w_i = \frac{n_i}{n + 2 - k}, \qquad [41]$$

where $n = \sum_{i=1}^{k} n_i$,

and $\mathcal{E}^2$ of so weighted combined biased estimate, $\frac{\sum_{i=1}^{k} n_i s_i^2}{n + 2 - k}$, is

$$\mathcal{E}^2\left(\frac{\sum_{i=1}^{k} n_i s_i^2}{n + 2 - k}\right) = \frac{2\sigma^4}{n + 2 - k} \qquad [42]$$

which is smaller than $\mathcal{E}^2$ of best weighted unbiased combined estimate

$$\frac{\sum_{i=1}^{k} n_i s_i^2}{n-k},$$

which is

$$\mathcal{E}^2\left(\frac{\sum_{i=1}^{k} n_i s_i^2}{n-k}\right)=D^2\left(\frac{\sum_{i=1}^{k} n_i s_i^2}{n-k}\right)=\frac{2\sigma^4}{n-k}. \qquad [43]$$

The above weightings, of course, have been deduced on the assumption that means of the samples are not availab'e for information, or that the samples had been drawn from different populations differing in respect to means, but not to variances, all of which are supposed to be equal. Were there no such assumption, the best estimate would be the one calculated with all observations put together, i. e.,

$$\hat{\sigma}^2=\frac{n}{n+1}\cdot s^2=\frac{1}{n+1}\sum_{i=1}^{k} n_i s_i^2+\frac{1}{n+1}\sum_{i=1}^{k} n_i(\bar{x}_i-\bar{x})^2, \qquad [44]$$

where $\bar{x}=\frac{1}{n}\sum_{i=1}^{k} n_i \bar{x}_i$.

It may be argued that our negatively biased estimates, if used for testing statistical hypotheses, would tend in the long run to overstate significance. This objection could be valid, if the estimates were used, as they had been before, in inexact tests of significance. Since the introduction, however, of exact Student's ratio tests there is no need to estimate the unknown σ, and the objection becomes pointless. It just so happens that in these tests the expressions for unbiased estimates of σ^2 do enter, but even then the square roots of these expressions, which appear in the tests, are not in themselves unbiased estimates of σ. The important thing to note is that *the functional form of any exact critical ratio can be deduced quite independently from any statements pertaining to estimation.* It seems desirable, even if some authors do otherwise, not to mix problems of statistical tests with the problem of estimating a parameter by a single number.

* * *

We shall pass now to some special case where B-criterion is applicable. An attempt to introduce a biased estimate of σ^2 in preference to the unbiased $\frac{n}{n-1} \cdot s^2$ was made by Pitman[1]), who based his criterion of „closeness" on absolute magnitudes instead of on squares. According to him, estimate α_1^* will be considered „closer" than estimate α_2^* for all values of α, if the probability that $|\alpha_1^* - \alpha| < |\alpha_2^* - \alpha|$ is greater than $\frac{1}{2}$. His principle of finding the „closest" estimate is, however, analogous to the principle of least squares, and as such is subject to the same limitations. The logic of Pitman's procedure can be described briefly as follows.

Seeing that minimizing $\epsilon(\alpha^*) = E|\alpha^* - \alpha|$ with respect to α gives solution $\alpha = Me(\alpha^*)$, where $Me(\alpha^*)$ is median value of α^* in sampling distribution, Pitman chooses for best estimate of α such α^* the median value of which equals α, e. g., putting $\alpha = \sigma^2$, he choses

$$\sigma^{2*} = \frac{\sigma^2}{Me(s^2)} \cdot s^2 . \tag{45}$$

Thus he is proceeding similarly to those who, in accordance with least square solution, $\alpha = E(\alpha^*)$, take for best estimate of α such α^* the mathematical expectation of which equals α, e. g., putting $\alpha = \sigma^2$, take $\sigma^{2*} = \frac{n}{n-1} \cdot s^2$. In reality Pitman's expression does estimate σ^2 better than it does any other parameter from the point of view of B-criterion, just as $\frac{n}{n-1} \cdot s^2$ estimates σ^2 better than it does any other parameter from the point of view of A-criterion. As estimate of σ^2, however, Pitman's expression is far from being best, for, being of the form As^2, it can be obtained by minimizing but one of the two components (the second) of the following expression of $E|As^2 - \sigma^2|$ with respect to A .[2])

$$E|As^2 - \sigma^2| = \begin{cases} E|As^2 - Me(As^2)| + 2\int\limits_{Me(As^2)}^{\sigma^2} (\sigma^2 - As^2)\,dF(As^2) & \text{for } Me(As^2) \leqslant \sigma^2 \\ E|As^2 - Me(As^2)| + 2\int\limits_{\sigma^2}^{Me(As^2)} (As^2 - \sigma^2)\,dF(As^2) & \text{for } Me(As^2) \geqslant \sigma^2, \end{cases} \tag{46}$$

where $F(As^2)$ is distribution function of random variable As^2.

---

[1]) Pitman, E. J. G. The „closest" estimates of statistical parameters. Proceedings of the Cambridge Philosophical Society, 33 (1937), p. 212.

[2]) Cf. Cramér, H. Mathematical Methods of Statistics, p. 179.

In fact, since either integral becomes minimum (zero) when lower and upper limits of integration are equal, i. e., when $A=\frac{\sigma^2}{Me(s^2)}$, this weak criterion gives $\frac{\sigma^2}{Me(s^2)}\cdot s^2$, i. e., precisely the Pitman's expression given in [45], for estimating σ^2. When X is normal, $\frac{\sigma^2}{Me(s^2)}$ can be expressed as a function of n alone, since then $s^2=\frac{\sigma^2\chi^2_{\nu=n-1}}{n}$, and therefore $\frac{\sigma^2}{Me(s^2)}=\frac{n}{Me(\chi^2)_{\nu=n-1}}$, distribution of χ^2 depending but on one parameter, ν. It remains to evaluate $Me(\chi^2)_{\nu=n-1}$ from

$$\int_0^{Me(\chi^2)_{\nu=n-1}} f_{n-1}(\chi^2)\,d\chi^2=\frac{1}{2}, \qquad [47]$$

where $f_{n-1}(\chi^2)$ is frequency function of χ^2 with $\nu=n-1$ as given by

$$f_\nu(\chi^2)=\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\cdot(\chi^2)^{\frac{\nu-2}{2}}\cdot e^{\frac{-\chi^2}{2}}\cdot \qquad [48]$$

In the result of this evaluation Pitman arrives at an approximate expression

$$Me(\chi^2)\approx\nu-\frac{2}{3}+\frac{0.09}{\nu} \qquad [49]$$

which gives an occasional error of 1 in the third decimal place for $\nu>4$. The exact values of $Me(\chi^2)$ for $\nu=1, 2,\ldots 30$ can be, of course, read off directly from Chi-Square Table at $P=0.50$.

It can be seen that Pitman's estimate of σ^2, which we shall denote by

$$\overset{(\cdot)}{\sigma^2}=\frac{n}{Me(\chi^2)_{\nu=n-1}}\cdot s^2\approx\frac{n}{n-1-\frac{2}{3}+\frac{0.09}{n-1}}\cdot s^2 \qquad [50]$$

is positively biased and, from the point of view of A-criterion, even farther removed from σ^2 than the unbiased sufficient joint estimate, $\breve{\sigma}^2=\frac{n}{n-1}\cdot s^2$, is.

The best linear estimate of σ^2 of the form As^2 will be obtained, from the point of view of B-criterion, through minimizing $\epsilon(As^2)=E\,|\,As^2-\sigma^2\,|$ with respect to A. We shall give first a general expression for derivative of $E\,|\,A\alpha^*-\alpha\,|$ with respect to A, equated to zero, where α^* is any statistic, and α any parameter.

For this purpose we shall write

$$\varphi(A) = E\,|\,Aa^* - a\,| = \int_{-\infty}^{\infty} |\,Aa^* - a\,|\, dF(Aa^*) = \int_{-\infty}^{a} (a - Aa^*)\, dF(Aa^*) + $$

$$+ \int_{a}^{\infty} (Aa^* - a)\, dF(Aa^*)$$

$$= a\int_{-\infty}^{a} dF(Aa^*) - \int_{-\infty}^{a} Aa^* dF(Aa^*) + \int_{a}^{\infty} Aa^* dF(Aa^*) -$$

$$- a\int_{a}^{\infty} dF(Aa^*)$$

$$= \int_{-\infty}^{\infty} Aa^* dF(Aa^*) - 2\int_{-\infty}^{a} Aa^* dF(Aa^*) - a\int_{-\infty}^{\infty} dF(Aa^*) +$$

$$+ 2a\int_{-\infty}^{a} dF(Aa^*)$$

$$= E(Aa^*) - a - 2\int_{-\infty}^{a} Aa^*)\, dF(Aa^*) + 2a\int_{-\infty}^{a} dF(Aa^*), \qquad [51]$$

Changing the variable of integration from Aa^* to a^*,

$$\varphi(A) = AE(a^*) - a - 2A\int_{-\infty}^{\frac{a}{A}} a^* dF(a^*) + 2a\int_{-\infty}^{\frac{a}{A}} dF(a^*)$$

$$= AE(a^*) - a - 2A\int_{-\infty}^{0} a^* dF(a^*) - 2A\int_{0}^{\frac{a}{A}} a^* dF(a^*) + 2a\int_{-\infty}^{0} dF(a^*) +$$

$$+ 2a\int_{0}^{\frac{a}{A}} dF(a^*). \qquad [52]$$

Minimizing this expression with respect to A, we obtain

$$\varphi'(A) = E(a^*) - 2\int_{-\infty}^{0} a^* dF(a^*) - 2\int_{0}^{\frac{a}{A}} a^* dF(a^*) - 2A\left(\frac{d}{d\frac{a}{A}}\int_{0}^{\frac{a}{A}} a^* dF(a^*)\right)\frac{d\frac{a}{A}}{dA} +$$

$$+ 2a\left(\frac{d}{d\frac{a}{A}}\int_{0}^{\frac{a}{A}} dF(a^*)\right)\frac{d\frac{a}{A}}{dA} = 0, \qquad [53]$$

and since $dF(a^*) = f(a^*)\, da^*$, where $f(a^*)$ is frequency function of a^*, it follows that

$$A\left(\frac{d}{d\frac{a}{A}}\int_0^{\frac{a}{A}} a^* dF(a^*)\right)\frac{d\frac{a}{A}}{dA} = a f\left(\tfrac{a}{A}\right)\frac{d\frac{a}{A}}{dA} = a\left(\frac{d}{d\frac{a}{A}}\int_0^{\frac{a}{A}} dF(a^*)\right)\frac{d\frac{a}{A}}{dA},$$

and therefore

$$\varphi'(A) = E(a^*) - 2\int_{-\infty}^{0} a^* dF(a^*) - 2\int_0^{\frac{a}{A}} a^* dF(a^*) = 0. \qquad [54]$$

Now, substituting s^2 for a^* and σ^2 for a, and noting that $\int_{-\infty}^{0} s^2 f(s^2)\, ds^2 = 0$ we are left with

$$\frac{(n-1)\,\sigma^2}{n} - 2\int_0^{\frac{\sigma^2}{A}} s^2 f(s^2)\, ds^2 = 0,$$

so that

$$\int_0^{\frac{\sigma^2}{A}} s^2 f(s^2)\, ds^2 = \frac{(n-1)\,\sigma^2}{2n}. \qquad [55]$$

If X is normal, we can change the variable of integration from s^2 to χ^2. Since $s^2 = \frac{\sigma^2 \chi^2}{n}{}_{\nu = n-1}$ and $f(s^2)\, ds^2 = f_{n-1}(\chi^2)\, d\chi^2$, the [55] becomes

$$\frac{\sigma^2}{n}\int_0^{\frac{n}{A}} \chi^2 f_{n-1}(\chi^2)\, d\chi^2 = \frac{(n-1)\sigma^2}{2n},$$

and therefore

$$\int_0^{\frac{n}{A}} \chi^2 f_{n-1}(\chi^2)\, d\chi^2 = \frac{n-1}{2}. \qquad [56]$$

From [48] we have

$$f_{n-1}(\chi^2) = \frac{1}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}\cdot(\chi^2)^{\frac{n-3}{2}}\cdot e^{-\frac{\chi^2}{2}}. \qquad [57]$$

Substituting this into [56],

$$\frac{1}{2^{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)}\cdot\int_0^{\frac{n}{A}} (\chi^2)^{\frac{n-1}{2}}\, e^{-\frac{\chi^2}{2}}\, d\chi^2 = \frac{n-1}{2}.$$

Now, since by [48]

$$\int_0^{\frac{n}{A}} (\chi^2)^{\frac{n-1}{2}} e^{-\frac{\chi^2}{2}} d\chi^2 = \int_0^{\frac{n}{A}} f_{n+1}(\chi^2) \cdot 2^{\frac{n+1}{2}} \Gamma\left(\tfrac{n+1}{2}\right) d\chi^2, \tag{58}$$

the [56] becomes

$$\frac{2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} \cdot \int_0^{\frac{n}{A}} f_{n+1}(\chi^2)\, d\chi^2 = \frac{n-1}{2}.$$

But $\Gamma\left(\frac{n+1}{2}\right) = \frac{n-1}{2} \Gamma\left(\frac{n-1}{2}\right)$, and therefore

$$\int_0^{\frac{n}{A}} f_{n+1}(\chi^2)\, d\chi^2 = \frac{1}{2}, \tag{59}$$

which shows that $\frac{n}{A} = Me(\chi^2)_{\nu = n+1}$, i. e.,

$$A = \frac{n}{Me(\chi^2)_{\nu = n+1}} \tag{60}$$

Using [49] for evaluating $Me(\chi^2)_{\nu = n+1}$, we obtain finally

$$A \approx \frac{n}{n+1-\frac{2}{3}+\frac{0.09}{n+1}}. \tag{61}$$

Thus the best estimate of σ^2 in normal population with unknown mean, of the form As^2, is

$$\overset{(\cdot)}{\sigma}{}^2 = \frac{n}{Me(\chi^2)_{\nu=n+1}} \cdot s^2 \approx \frac{n}{n+1-\frac{2}{3}+\frac{0.09}{n+1}} \cdot s^2. \tag{62}$$

Similarly, for the best estimate of σ^2 of the form As_0^2 in normal population with known mean, we obtain

$$\overset{(\bullet)}{\sigma}{}^2 = \frac{n}{Me(\chi^2)_{\nu=n+2}} \cdot s_0^2 \approx \frac{n}{n+2-\frac{2}{3}+\frac{0.09}{n+2}} \cdot s_0^2 \tag{63}$$

in place of Pitman's

$$\overset{(\sim)}{\sigma}{}^2 = \frac{n}{Me(\chi^2)_{\nu=n}} \cdot s_0^2 \approx \frac{n}{n-\frac{2}{3}+\frac{0.09}{n}} \cdot s_0^2. \tag{64}$$

It can be seen that our estimates differ from Pitman's in that in the expression $\frac{n}{Me(\chi^2)}$ the χ^2 in our formulas is taken with two more d. f. than in Pitman's, which makes our estimates negatively biased, while his are positively biased.

To determine mean absolute errors of these estimates we shall consider separately estimates based on s^2 and estimates based on s_0^2. On the basis of [52] we can write for the former

$$E\,|\,As^2-\sigma^2\,| = \frac{A(n-1)\sigma^2}{n} - \sigma^2 - 2A\int_0^{\frac{\sigma^2}{A}} s^2 dF(s^2) + 2\sigma^2 \int_0^{\frac{\sigma^2}{A}} dF(s^2)\,, \qquad [65]$$

and for the latter

$$E\,|\,As_0^2-\sigma^2\,| = A\sigma^2 - \sigma^2 - 2A\int_0^{\frac{\sigma^2}{A}} s_0^2 dF(s_0^2) + 2\sigma^2 \int_0^{\frac{\sigma^2}{A}} dF(s_0^2)\,. \qquad [66]$$

Changing the variable of integration from s^2 to χ^2 on the assumption of normal population, the [65] becomes

$$E\,|\,As^2-\sigma^2\,| = \frac{A(n-1)\sigma^2}{n} - \sigma^2 - \frac{2A\sigma^2}{n}\int_0^{\frac{n}{A}} \chi^2 f_{n-1}(\chi^2)\,d\chi^2 + 2\sigma^2 \int_0^{\frac{n}{A}} f_{n-1}(\chi^2)\,d\chi^2\,, \qquad [67]$$

and the [66] becomes

$$E\,|\,As_0^2-\sigma^2\,| = A\sigma^2 - \sigma^2 - \frac{2A\sigma^2}{n}\int_0^{\frac{n}{A}} \chi^2 f_n(\chi^2)\,d\chi^2 + 2\sigma^2 \int_0^{\frac{n}{A}} f_n(\chi^2)\,d\chi^2\,. \qquad [68]$$

Now, from [48] and [58] it follows that

$$\int \chi^2 f_k(\chi^2)\,d\chi^2 = k\int f_{k+2}(\chi^2)\,d\chi^2\,, \qquad [69]$$

where k is any natural number.

Since in [67] $k=n-1$, we obtain after substituting [69]

$$E\,|\,As^2-\sigma^2\,| = \frac{A(n-1)\,\sigma^2}{n} - \sigma^2 - \frac{2A\,(n-1)\,\sigma^2}{n}\int_0^{\frac{n}{A}} f_{n+1}(\chi^2)\,d\chi^2 + 2\sigma^2 \int_0^{\frac{n}{A}} f_{n-1}(\chi^2)\,d\chi^2. \qquad [70]$$

Similarly, having in [68] $k=n$, we obtain

$$E\,|\,As_0^2-\sigma^2\,| = A\sigma^2 - \sigma^2 - 2A\sigma^2\int_0^{\frac{n}{A}} f_{n+2}(\chi^2)\,d\chi^2 + 2\sigma^2 \int_0^{\frac{n}{A}} f_n(\chi^2)\,d\chi^2. \qquad [71]$$

From [70] and [71] some convenient formulas can be derived for the evaluation of mean absolute error of any estimate based on s^2 or on s_0^2 directly from Chi-Square Table.

Putting $A = \frac{n}{Me(\chi^2)_{\nu=n-1}}$, we shall have from [70]

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n-1}}\cdot s^2-\sigma^2\right| = \sigma^2\cdot\frac{n-1}{Me(\chi^2)_{\nu=n-1}}\left[2\left(1-\int_0^{Me(\chi^2)_{\nu=n-1}} f_{n+1}(\chi^2)\,d\chi^2\right)-1\right] \quad [72]$$

which is an expression for mean absolute error of Pitman's estimate of σ^2 in normal population with unknown mean.

Putting $A = \frac{n}{Me(\chi^2)_{\nu=n+1}}$ in [70],

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n+1}}\cdot s^2-\sigma^2\right| = \sigma^2\left[1-2\left(1-\int_0^{Me(\chi^2)_{\nu=n+1}} f_{n-1}(\chi^2)\,d\chi^2\right)\right] \quad [73]$$

which is an expression for mean absolute error of our B-estimate of σ^2 in normal population with unknown mean. Similarly, putting $A = \frac{n}{Me(\chi^2)_{\nu=n}}$ in [71], we shall have

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n}}\cdot s_0^2-\sigma^2\right| = \sigma^2\cdot\frac{n}{Me(\chi^2)_{\nu=n}}\left[2\left(1-\int_0^{Me(\chi^2)_{\nu=n}} f_{n+2}(\chi^2)\,d\chi^2\right)-1\right] \quad [74]$$

which is an expression for mean absolute error of Pitman's estimate of σ^2 in normal population with known mean.

Again, putting $A = \frac{n}{Me(\chi^2)_{\nu=n+2}}$ in [71],

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n+2}}\cdot s_0^2-\sigma^2\right| = \sigma^2\left[1-2\left(1-\int_0^{Me(\chi^2)_{\nu=n+2}} f_n(\chi^2)\,d\chi^2\right)\right] \quad [75]$$

which is an expression for mean absolute error of our B-estimate of σ^2 in normal population with known mean.

Incidentally, the analogous expressions for mean absolute errors of other estimates can be obtained in the same way, vis., of $\frac{n}{n-1}\cdot s^2$, $\frac{n}{n+1}\cdot s^2, s^2$, s_0^2, and $\frac{n}{n+2}\cdot s_0^2$.

$$E\left|\frac{n}{n-1}\cdot s^2-\sigma^2\right| = 2\sigma^2\left[\left(1-\int_0^{n-1} f_{n+1}(\chi^2)\,d\chi^2\right)-\left(1-\int_0^{n-1} f_{n-1}(\chi^2)\,d\chi^2\right)\right] \quad [76]$$

which is an expression for mean absolute error of unbiased sufficient joint estimate of σ^2 in normal population.

Next

$$E\left|\frac{n}{n+1}\cdot s^2-\sigma^2\right|=\sigma^2\left\{[1-2\,(1-\int_0^{n+1} f_{n-1}(\chi^2)d\chi^2)]-\right.$$
$$\left.-[1-2\,(1-\int_0^{n+1} f_{n+1}(\chi^2)\,d\chi^2)]\frac{n-1}{n+1}\right\} \quad [77]$$

which is an expression for mean absolute error of linearly sufficient A-estimate of σ^2 in normal population with unknown mean.

Next

$$E\left|s^2-\sigma^2\right|=\sigma^2\left\{[2\,(1-\int_0^{n} f_{n+1}(\chi^2)\,d\chi^2)-1]\frac{n-1}{n}+\right.$$
$$\left.+[1-2\,(1-\int_0^{n} f_{n-1}(\chi^2)\,d\chi^2)]\right\} \quad [78]$$

which is an expression for mean absolute error of sample variance used as an estimate of normal population variance.

Next

$$E\left|s_0^2-\sigma^2\right|=2\,\sigma^2[(1-\int_0^{n} f_{n+2}(\chi^2)\,d\chi^2)-(1-\int_0^{n} f_n(\chi^2)\,d\chi^2)] \quad [79]$$

which is an expression for mean absolute error of unbiased efficient estimate of σ^2 in normal population.

And finally

$$E\left|\frac{n}{n+2}\cdot s_0^2-\sigma^2\right|=\sigma^2\left\{[1-2\,(1-\int_0^{n+2} f_n(\chi^2)\,d\chi^2)]-\right.$$
$$\left.-[1-2\,(1-\int_0^{n+2} f_{n+2}(\chi^2)\,d\chi^2)]\frac{n}{n+2}\right\} \quad [80]$$

which is an expression for mean absolute error of linearly efficient A-estimate of σ^2 in normal population.

In the above formulas the expressions in parentheses could be evaluated without using the χ^2 table, by Fisher's formula[1]) for even ν

$$1-\int_0^{\chi_0^2} f_\nu(\chi^2)\,d\chi^2=e^{-\frac{\chi_0^2}{2}}\left(1+\frac{\chi_0^2}{2}+\frac{\left(\frac{\chi_0^2}{2}\right)^2}{2!}+\frac{\left(\frac{\chi_0^2}{2}\right)^3}{3!}+\ldots+\frac{\left(\frac{\chi_0^2}{2}\right)^{\frac{\nu-2}{2}}}{\left(\frac{\nu-2}{2}\right)!}\right) \quad [81]$$

However, for the four B-estimates and for the two unbiased A-estimates the simpler formulas can be derived for mean absolute error of estimate. For this purpose we shall write the reduction formula

[1]) R. A. Fisher „The mathematical distributions used in the common tests of significance". Econometrica, vol. 3, 1935, pp. 353-65.

$$\int (\chi^2)^{\frac{k}{2}} e^{-\frac{\chi^2}{2}} d\chi^2 = k \int (\chi^2)^{\frac{k-2}{2}} e^{-\frac{\chi^2}{2}} d\chi^2 - 2(\chi^2)^{\frac{k}{2}} e^{-\frac{\chi^2}{2}} \qquad [82]$$

in the form

$$2^{\frac{k+2}{2}} \Gamma\left(\tfrac{k+2}{2}\right) \int f_{k+2}(\chi^2)\, d\chi^2 = k\, 2^{\frac{k}{2}} \Gamma\left(\tfrac{k}{2}\right) \int f_k(\chi^2)\, d\chi^2 - 2(\chi^2)^{\frac{k}{2}} e^{-\frac{\chi^2}{2}}.$$

Solving for $\int f_k(\chi^2)\, d\chi^2$ and reducing gamma functions, we have

$$\int f_k(\chi^2)\, d\chi^2 = \int f_{k+2}(\chi^2)\, d\chi^2 + \frac{(\chi^2)^{\frac{k}{2}} e^{-\frac{\chi^2}{2}}}{k\, 2^{\frac{k-2}{2}}\, \Gamma\left(\tfrac{k}{2}\right)}.$$

Passing to definite integrals, we obtain the relation

$$\int_0^{\chi_0^2} f_k(\chi^2)\, d\chi^2 = \int_0^{\chi_0^2} f_{k+2}(\chi^2)\, d\chi^2 + \frac{(\chi_0^2)^{\frac{k}{2}} e^{-\frac{\chi_0^2}{2}}}{k\, 2^{\frac{k-2}{2}}\, \Gamma\left(\tfrac{k}{2}\right)} \qquad [83]$$

which expresses distribution function of χ^2 with $\nu=k$ in terms of distribution function of χ^2 with $\nu=k+2$.

Now, substituting $n-1$ for k and $\frac{n}{A}$ for χ_0^2, we shall have

$$\int_0^{\frac{n}{A}} f_{n-1}(\chi^2)\, d\chi^2 = \int_0^{\frac{n}{A}} f_{n+1}(\chi^2)\, d\chi^2 + \frac{\left(\frac{n}{A}\right)^{\frac{n-1}{2}} e^{-\frac{n}{2A}}}{2^{\frac{n-3}{2}} (n-1)\, \Gamma\left(\frac{n-1}{2}\right)}. \qquad [84]$$

When $\frac{n}{A} = Me(\chi^2)_{\nu=n-1}$, the last relation gives

$$\int_0^{Me(\chi^2)_{\nu=n-1}} f_{n+1}(\chi^2)\, d\chi^2 = \frac{1}{2} - \frac{Me(\chi^2)_{\nu=n-1}^{\frac{n-1}{2}} \cdot e^{-\frac{Me(\chi^2)_{\nu=n-1}}{2}}}{2^{\frac{n-3}{2}} (n-1)\, \Gamma\left(\frac{n-1}{2}\right)}. \qquad [85]$$

Substituting this into [72], we obtain after some reductions

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n-1}} s^2 - \sigma^2\right| = \sigma^2 \cdot \frac{Me(\chi^2)_{\nu=n-1}^{\frac{n-3}{2}} \cdot e^{-\frac{Me(\chi^2)_{\nu=n-1}}{2}}}{2^{\frac{n-5}{2}}\, \Gamma\left(\frac{n-1}{2}\right)}. \qquad [86]$$

mean absolute error of Pitman's estimate of σ in normal population with unknown mean

When $\frac{n}{A} = Me(\chi^2)_{\nu=n+1}$, the [84] becomes

$$\int_0^{Me(\chi^2)_{\nu=n+1}} f_{n-1}(\chi^2)\, d\chi^2 = \frac{1}{2} + \frac{Me(\chi^2)_{\nu=n+1}^{\frac{n-1}{2}} \cdot e^{-\frac{Me(\chi^2)_{\nu=n+1}}{2}}}{2^{\frac{n-3}{2}} (n-1)\, \Gamma\left(\frac{n-1}{2}\right)}. \qquad [87]$$

Substituting this into [73], we obtain

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n+1}}\cdot s^2-\sigma^2\right| = \sigma^2\,\frac{Me(\chi^2)_{\nu=n+1}^{\frac{n-1}{2}}\cdot e^{-\frac{Me(\chi^2)_{\nu=n+1}}{2}}}{2^{\frac{n-5}{2}}(n-1)\,\Gamma\left(\frac{n-1}{2}\right)}\,. \qquad [88]$$

mean absolute error of our B-estimate of σ^2 in normal population with unknown mean

It can be seen that mean absolute error of Pitman's estimate, $\overset{(\smile)}{\sigma}{}^2$, given in [86] is equal to mean absolute error of our estimate, $\overset{(\cdot)}{\sigma}{}^2$, (given in [88]), when the latter is based on 2 less observations.

$$\epsilon\,(\overset{(\smile)}{\sigma}{}^2_n) = \epsilon\,(\overset{(\cdot)}{\sigma}{}^2_{n-2}) \qquad [89]$$

Putting now $n+1$ for n in [85] and then substituting into [74], we obtain

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n}}\cdot s_0^2-\sigma^2\right| = \sigma^2\,\frac{Me(\chi^2)_{\nu=n}^{\frac{n-2}{2}}\cdot e^{-\frac{Me(\chi^2)_{\nu=n}}{2}}}{2^{\frac{n-4}{2}}\,\Gamma\left(\frac{n}{2}\right)}\,. \qquad [90]$$

mean absolute error of Pitman's estimate of σ^2 in normal population with known mean

Similarly, putting $n+1$ for n in [87] and then substituting into [75], we obtain

$$E\left|\frac{n}{Me(\chi^2)_{\nu=n+2}}\cdot s_0^2-\sigma^2\right| = \sigma^2\,\frac{Me(\chi^2)_{\nu=n+2}^{\frac{n}{2}}\cdot e^{-\frac{Me(\chi^2)_{\nu=n+2}}{2}}}{2^{\frac{n-4}{2}}\,n\,\Gamma\left(\frac{n}{2}\right)}\,. \qquad [91]$$

mean absolute error of our B-estimate of σ^2 in normal population with known mean

Again it can be seen from [90] and [91] that mean absolute error of Pitman's estimate, $\overset{(\sim)}{\sigma}{}^2$, is equal to mean absolute error of our estimate, $\overset{(\bullet)}{\sigma}{}^2$, when the latter is based on 2 less observations

$$\epsilon\,(\overset{(\sim)}{\sigma}{}^2_n) = \epsilon\,(\overset{(\bullet)}{\sigma}{}^2_{n-2})\,. \qquad [92]$$

Thus the taking of Pitman's estimates instead of ours results in a loss of 2 observations out of n.

Comparing [86] with [90] and [88] with [91] it can be also seen that mean absolute errors of estimates based on s^2 with n observations are equal to mean absolute errors of the corresponding estimates based on s_0^2 with $n-1$ observations

$$\epsilon\,(\overset{(\smile)}{\sigma}{}^2_n) = \epsilon\,(\overset{(\sim)}{\sigma}{}^2_{n-1}),\quad \epsilon\,(\overset{(\cdot)}{\sigma}{}^2_n) = \epsilon\,(\overset{(\bullet)}{\sigma}{}^2_{n-1})\,. \qquad [93]$$

Bringing together relations [89], [92], and [93], we can write

$$\epsilon\,(\overset{(\smile)}{\sigma}{}^2_n) = \epsilon\,(\overset{(\sim)}{\sigma}{}^2_{n-1}) = \epsilon\,(\overset{(\cdot)}{\sigma}{}^2_{n-2}) = \epsilon\,(\overset{(\bullet)}{\sigma}{}^2_{n-3}) \qquad [94]$$

which shows the order of B-estimates in respect to their efficiency from the point of view of B-criterion. It is interesting to note that the same order is preserved from the point of view of A-criterion for the corresponding A-estimates of σ^2, $\tilde{\sigma}^2 = \frac{n}{n-1} \cdot s^2$, $\breve{\sigma}^2 = s_0^2$, $\dot{\sigma}^2 = \frac{n}{n+1} \cdot s^2$, and $\mathring{\sigma}^2 = \frac{n}{n+2} \cdot s_0^2$

$$\mathcal{E}^2 (\breve{\sigma}^2_n) = \mathcal{E}^2 \tilde{\sigma}(^2_{n-1}) = \mathcal{E}^2 (\dot{\sigma}^2_{n-2}) = \mathcal{E}^2 (\mathring{\sigma}^2_{n-3}), \qquad [95]$$

so that the linear efficiencies in B-sense of the above B-estimates are proportional to the linear efficiencies in A-sense of the corresponding A-estimates.

Putting in [83] $n-1$ for k and $n-1$ for χ_0^2, and substituting into [70] with $A = \frac{n}{n-1}$, we shall obtain

$$E\left|\frac{n}{n-1} \cdot s^2 - \sigma^2\right| = \sigma^2 \frac{(n-1)^{\frac{n-3}{2}} e^{-\frac{n-1}{2}}}{2^{\frac{n-5}{2}} \Gamma\left(\frac{n-1}{2}\right)} \cdot \qquad [96]$$

mean absolute error of unbiased sufficient joint estimate of σ^2 in normal population

Putting in [83] n for k and n for χ_0^2, and substituting into [71] with $A = 1$, we obtain

$$E\left|s_0^2 - \sigma^2\right| = \sigma^2 \frac{n^{\frac{n-2}{2}} e^{\frac{-n}{2}}}{2^{\frac{n-4}{2}} \Gamma\left(\frac{n}{2}\right)} \cdot \qquad [97]$$

mean absolute error of unbiased efficient estimate of σ^2 in normal population

Again it can be seen from [96] and [97], and from [77] and [80] that mean absolute error of unbiased sufficient joint estimate of σ^2 based on n observations is equal to mean absolute error of unbiased efficient estimate based on $n-1$ observations, and similarly for linearly sufficient and efficient estimates

$$\epsilon (\breve{\sigma}^2_n) = \epsilon (\tilde{\sigma}^2_{n-1}), \; \epsilon (\mathring{\sigma}^2_n) = \epsilon (\mathring{\sigma}^2_{n-1}) . \qquad [98]$$

Fig. 1 gives mean absolute errors (encircled on continuous lines) and mean square root errors (on broken lines) of estimates of σ^2 in normal population, with $n = 1, 2, 4, 8, 16$, relatively to the estimated parameter, for the nine estimates arranged according to the rank of their linear efficiency in B-sense. The relativity of the notion of efficiency is clearly brought out, especially in the case of s^2 and s_0^2, the first of which is more efficient than the other in

A-sense and less in B-sense. As it has been said, in cases of divergent results obtained by the two criteria the ultimate decision should belong to B-criterion rather than to A-criterion. When a comparative appraisal of two or more statistics is sought, either criterion can be used for most of determinations, the discrepancies being comparatively small and extending only to estimates lying close to the controversial „best estimates", with the rest of determinations consistent with both criteria.

The use of B-criterion instead of A-criterion illustrates also relativity of our notion of bias. While an unbiased estimate in A-sense is any a^* satisfying $E(a^*) = a$, an unbiased estimate in B-sense could be conceived as any a^* satisfying $Me(a^*) = a$. Clearly, the unbiased estimate in one sense will be biased in the other, unless sampling distribution of the statistic in question is symmetrical. But, from our point of view, the presence or absence of bias in both senses is from à priori considerations quite irrelevant to the qualifications of good estimate. As we have seen, the latter generally turn out to be negatively biased (in both senses), whether chosen by A- or by B-criterion.

The main point of issue is not a dilemma of A- versus B-criterion, but exact solutions by either. None of the criteria in general use do justice to this point. But on the other hand, in more difficult cases of estimation, neither A- nor B-criterion provides ready means for determining best functional form of estimate. In this respect the method of Maximum Likelihood has a unique advantage. As long as we confine ourselves to linearly biased estimates (which seems to suffice for most practical purposes), we can improve our estimates by combining Maximum Likelihood solutions with the principle of optimum bias.

Table I gives numerical values of coefficients of s^2 and s_0^2 in eight etimates of σ^2 in normal population — for n varying from 1 to 30.

TABLE I

n	Coefficients of s^2				Coefficients of s_0^2			
	By A-criterion		By B-criterion[1])		By A-criterion		By B-criterion[1])	
	weak	strong	weak	strong	weak	strong	weak	strong
	$\frac{n}{n-1}$	$\frac{n}{n+1}$	$\frac{n}{Me(\chi^2)}$		1	$\frac{n}{n+2}$	$\frac{n}{Me(\chi^2)}$	
			$\nu=n-1$	$\nu=n+1$			$\nu=n$	$\nu=n+2$
1	—	—	—	—	1.000	*0.333*	2.198	**0.423**
2	2.000	*0.667*	4.396	**0.845**	1.000	*0.500*	1.443	**0.596**
3	1.500	*0.750*	2.165	**0.894**	1.000	*0.600*	1.268	**0.689**
4	1.333	*0.800*	1.691	**0.919**	1.000	*0.667*	1.192	**0.748**
5	1.250	*0.833*	1.489	**0.935**	1.000	*0.714*	1.149	**0.788**
6	1.200	*0.857*	1.379	**0.945**	1.000	*0.750*	1.122	**0.817**
7	1.167	*0.875*	1.309	**0.953**	1.000	*0.778*	1.103	**0.839**
8	1.143	*0.889*	1.261	**0.959**	1.000	*0.800*	1.089	**0.856**
9	1.125	*0.900*	1.225	**0.963**	1.000	*0.818*	1.079	**0.870**
10	1.111	*0.909*	1.199	**0.967**	1.000	*0.833*	1.070	**0.882**
11	1.100	*0.917*	1.177	**0.970**	1.000	*0.846*	1.064	**0.891**
12	1.091	*0.923*	1.160	**0.972**	1.000	*0.857*	1.058	**0.900**
13	1.083	*0.929*	1.146	**0.975**	1.000	*0.867*	1.053	**0.907**
14	1.077	*0.933*	1.135	**0.976**	1.000	*0.875*	1.050	**0.913**
15	1.071	*0.938*	1.125	**0.978**	1.000	*0.882*	1.046	**0.918**
16	1.067	*0.941*	1.116	**0.979**	1.000	*0.889*	1.043	**0.923**
17	1.062	*0.944*	1.108	**0.981**	1.000	*0.895*	1.041	**0.927**
18	1.059	*0.947*	1.102	**0.982**	1.000	*0.900*	1.038	**0.931**
19	1.056	*0.950*	1.096	**0.983**	1.000	*0.905*	1.036	**0.934**
20	1.053	*0.952*	1.091	**0.983**	1.000	*0.909*	1.034	**0.937**
21	1.050	*0.955*	1.086	**0.984**	1.000	*0.913*	1.033	**0.940**
22	1.048	*0.957*	1.032	**0.985**	1.000	*0.917*	1.031	**0.943**
23	1.045	*0.958*	1.080	**0.986**	1.000	*0.920*	1.030	**0.945**
24	1.043	*0.960*	1.074	**0.986**	1.000	*0.923*	1.028	**0.947**
25	1.042	*0.962*	1.071	**0.987**	1.000	*0.926*	1.027	**0.949**
26	1.040	*0.963*	1.068	**0.987**	1.000	*0.929*	1.026	**0.951**
27	1.038	*0.964*	1.066	**0.988**	1.000	*0.931*	1.025	**0.953**
28	1.037	*0.966*	1.063	**0.988**	1.000	*0.933*	1.024	**0.954**
29	1.036	*0.967*	1.061	**0.989**	1.000	*0.935*	1.023	**0.956**
30	1.034	*0.968*	1.059	**0.989**	1.000	*0.938*	1.023	**0.957**

[1]) Based on exact values of $Me(\chi^2)$ taken from Chi-Square Table at $P = 0.50$.

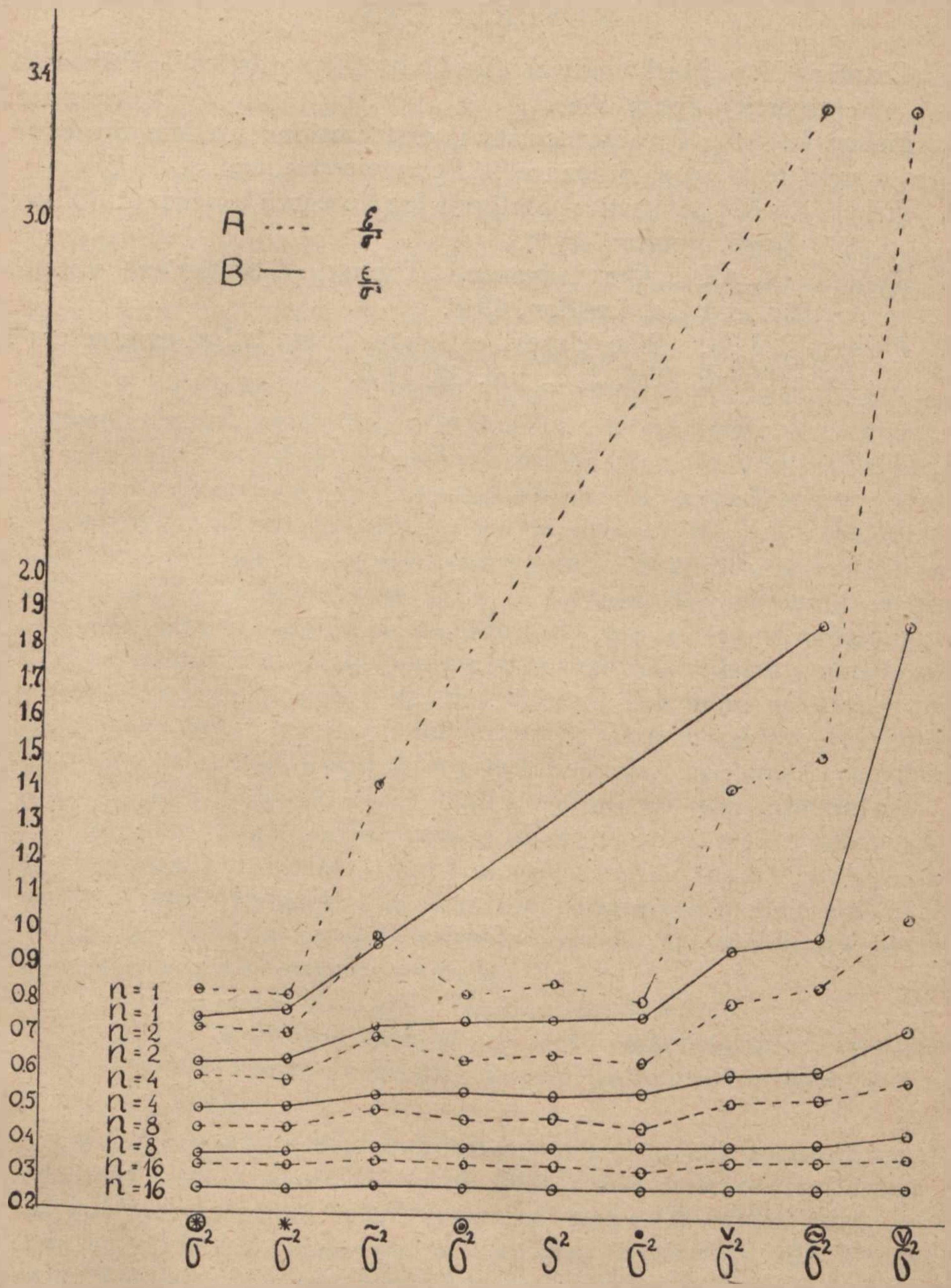
A ---- $\frac{\varepsilon}{\sigma^2}$
B — $\frac{\varepsilon}{\sigma^2}$
n = 1
n = 1
n = 2
n = 2
n = 4
n = 4
n = 8
n = 8
n = 16
n = 16
$\overset{\circledast}{\sigma}{}^2$ $\overset{*}{\sigma}{}^2$ $\tilde{\sigma}^2$ $\overset{\odot}{\sigma}{}^2$ s^2 $\dot{\sigma}^2$ $\check{\sigma}^2$

Fig. 1.

REFERENCES

1. Cramér, H. „Mathematical Methods of Statistics". Princeton University Press, 1946.
2. Fisher, R. A. „The mathematical distributions used in the common tests of significance". *Econometrica*, vol. 3, 1935.
3. Fisher, R. A. „Statistical Methods for Research Workers". Oliver and Boyd, London, 1948.
4. Kendall, M. G. „The Advanced Theory of Statistics" vol. 2. Griffin and Co, London, 1948.
5. Pitman, E. J. G. „The 'closest' estimates of statistical parameters". *Proc. Camb. Phil. Soc.*, vol. 33, 1937.

STRESZCZENIE

W pracy niniejszej używamy wyrazu „bliskość“ dla oznaczenia skupienia (concentration) wartości losowych jakiejkolwiek charakterystyki próby, a^*, wyrażonej w postaci funkcji spostrzeżeń składających się na próbę ($a^* = a^* (x_1, x_2, \ldots x_n)$) dookoła szacowanego parametru, a. Skupienie to można określić przy pomocy własności odwrotnej — dyspersji — wyrażonej jako średni błąd kwadratowy oceny, $\mathcal{E}^2 (a^*)$, zdefinjowany w [A], albo, w sposób bardziej naturalny jako średni błąd bezwzględny, $\epsilon (a^*)$, zdefinjowany w [B],

Wobec trudności operowania wzorem [B] przyjmujemy naogół wzór [A] za kryterium „dobroci“ danej oceny w porównaniu z innymi ocenami tegoż parametru. Będziemy je nazywali kryterjum A. W łatwiejszych jednak wypadkach będziemy się opierali na kryterium B, t. j. na średnim błędzie bezwzględnym. W wypadkach wyznaczeń rozbieżnych „oceny lepszej“ przez te dwa kryterja, wyznaczenie B będziemy uważali za „ostatecznie lepsze“. Moglibyśmy zawsze mówić o jednej z tych ocen jako o „lepszej w znaczeniu A“, zaś o drugiej jako o „lepszej w znaczeniu B“, ale ponieważ naogół używa się przeważnie wzoru [A], ograniczenia tego nie będziemy akcentowali wyraźnie, rozumiejąc je tylko domyślnie. Tak więc, mówiąc o „ocenach lepszych“, będziemy mieli na myśli takie charakterystyki próby, których średni błąd kwadratowy jest mniejszy, zaś przez „oceny najlepsze“, takie, których średni błąd kwadratowy jest najmniejszy.

Sformułowana powyżej zasada najmniejszego błędu oceny różni się od zasady zalecanej przez R. A. Fisher'a, który ogranicza oceny „dobre“ wyłącznie do ocen nieobciążonych błędem stałym (lub w skrócie, do ocen nieobciążonych). Zasada ta, znajdująca obecnie powszechne zastosowanie, wymaga

1. aby ocena była nieobciążona (unbiased),
2. aby z pośród wszystkich ocen nieobciążonych wybierać, taką, której zmienność losowa (sampling variance) jest najmniejsza.

Ponieważ średni błąd kwadratowy oceny można rozłożyć na dwa składniki według wzoru [1], gdzie $D^2 (a^*)$ jest zmiennością losową a^*, zaś $b (a^*)$ obciążeniem, zdefinjowanym przez [2], widać więc, że zasada Fisherowska opiera się na słabszym kryterjum, niż kryterjum najmniejszego średniego błędu kwadratowego, gdyż zamiast wymagać aby całe wyrażenie [1] było minimum wymaga aby każdy z jego dwuch składników osobno był minimum ($b^2(a^*) = 0$ i $D^2(a^*) =$ minimum). Pod tym względem zasada ta jest pokrewna z zasadą najmniejszych kwadratów, jak też z zasadą „największego podobieństwa" (Maximum Likelihood), gdyż obie te zasady opierają się na słabszych kryterjach, niż kryterjum najmniejszego średniego błędu kwadratowego.

Istotnie, zasada najmniejszych kwadratów za najlepszą ocenę uważa taką, od której suma kwadratów odchyleń wartości spostrzeżonych jest najmniejsza. Dopóki znaczenie „wartości spostrzeżonej", ma implikować, że jest ona oceną nieobciążoną „wartości prawdziwej", zasada najmniejszych kwadratów ogranicza się *ex definitione* do ocen nieobciążonych. Rozciągając tę zasadę do wartości losowych charakterystyki próby, α^*, będziemy musieli minimalizować $E(\alpha^*-\alpha)^2$ ze względu na α, zamiast ze względu na α^*, jakby to było wymagane z punktu widzenia zasady najmniejszego średniego błędu kwadratowego. Rozwiązanie, $\alpha = E(\alpha^*)$, ilustruje właściwe znaczenie zasady najmniejszych kwadratów, które polega na przyporządkowaniu optymalnej funkcji parametrowej, $E(\alpha^*)$, danej charakterystyce próby do oceny, nie zaś na wyznaczeniu odpowiedniej charakterystyki próby jako oceny danego parametru. Dla przykładu, $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$, zgodnie z tą interpretacją, powinno się uważać za lepiej nadające się do oceny $E(s^2) = \frac{(n-1)}{n}\sigma^2$ niż, powiedzmy, do oceny σ^2, lub jakiejkolwiek innej funkcji parametrowej. Podobnie, $\frac{n}{n-1}\cdot s^2$ powinno się uważać za lepiej nadające się do oceny $E\left(\frac{n}{n-1}s^2\right) = \sigma^2$ niż do oceny jakiejkolwiek innej funkcji parametrowej. Rzecz jasna, że nie jest to równoważne z wydaniem sądu, że $\frac{n}{n-1}\cdot s^2$ jest oceną lepszą σ^2 niż jest nią jakakolwiek inna charakterystyka próby.

Zasada największego podobieństwa żąda, aby za ocenę najlepszą obrać taką, dla której gęstość prawdopodobieństwa w punkcie próby $(x_1, x_2, \ldots x_n)$, względnie prawdopodobieństwo otrzymania danej próby (zależnie od tego czy mamy do czynienia ze zmienną ciągłą czy nieciągłą), jest maximum *ze względu na* α.

Dla zmiennej ciągłej: $L(x_1, x_2, \ldots x_n) = f(x_1; \alpha)\cdot f(x_2; \alpha) \ldots f(x_n; \alpha) =$ $=$ maximum.

Dla zmiennej nieciągłej: $L(x_1, x_2, \ldots x_n) = p_1(a) \cdot p_2(a) \ldots p_n(a) =$ maximum, gdzie $L(x_1, x_2, \ldots x_n)$ jest tzw. „funkcją podobieństwa“ próby złożonej z n spostrzeżeń; $f(x_1; a)$, $f(x_2; a), \ldots f(x_n; a)$ są gęstościami prawdopodobieństwa dla wartości $x_1, x_2, \ldots x_n$; zaś $p_1, p_2, \ldots p_n$ są prawdopodobieństwami otrzymania wartości $x_1, x_2, \ldots x_n$.

Rozwiązując równanie podobieństwa, $\frac{\partial \log L}{\partial a} = 0$, wyznacza się ocenę a w postaci funkcji n spostrzeżeń próby.

Zasadę największego podobieństwa można uważać za rozciągnięcie zasady, według której wartość najprawdopodobniejszą uważa się za najlepszą ocenę „wartości prawdziwej“. Zasada ta może przemawiać do intuicji, ale uzasadnić się logicznie nie da, „wartość prawdziwa“ bowiem niekoniecznie musi być wartością najprawdopodobniejszą ani być do niej bliższa niż do jakiejkolwiek innej wartości. Wielką natomiast zaletą metody największego podobieństwa jest to, że rozwiązania otrzymane tą drogą wypadają często w postaci funkcji nie oczywistych, które, gdy raz zostają określone, stają się funkcjami podstawowymi do oceny parametrów. Zasadę największego podobieństwa rzadko stosuje się w formie czystej; oceny otrzymane tą metodą koryguje się zwykle na obciążenie. Tak więc w praktycznym stosowaniu zasadę największego podobieństwa łączy się z zasadą ocen nieobciążonych. Zgodnie z przyjętym przez nas kryterjum najmniejszego średniego błędu kwadratowego, rozwiązania otrzymane metodą największego podobieństwa powinno się, jak to zobaczymy dalej, korygować w kierunku odwrotnym, co czyniąc je jeszcze bardziej obciążonymi, zmniejsza średni błąd kwadratowy oceny.

Mając na myśli „dobroć oceny“ w znaczeniu Fisherowskim, będziemy mówili, idąc za C r a m é r'e m: „dobroć oceny nieobciążonej“. W szczególności wyrażenie „ocena wydajna“ (efficient estimate), przez którą się rozumie taką ocenę nieobciążoną, $\tilde{a}$, której średni błąd kwadratowy jest równy wartości granicznej dla ocen nieobciążonych, zastąpimy wyrażeniem „ocena wydajna nieobciążona“.

Przy spełnieniu pewnych ogólnych warunków regularności można zgodnie z obecnym stanem teorji, wyznaczyć minimalną graniczną wartość $\mathcal{E}^2$ dla charakterystyk próby posiadających obciążenie zadanej wielkości. Dwie nierówności określają tę wartość graniczną, jedna dla przypadku o typie ciągłym, druga — dla przypadku o typie nieciągłym ([3] i [4]). Dla ocen nieobciążonych, $\hat{a}$, wartość graniczną $\mathcal{E}^2$ znajdujemy ze wzorów [5] i [6]. Przy spełnieniu pewnych warunków istnieją oceny nieobciążone o zmiennościach równych tym wartościom granicznym. Gdy oceny takie istnieją, można je zawsze wykryć za pomocą metody największego podobieństwa.

Ogólnie, dla ocen posiadających obciążenie równe, powiedzmy, b, wartości graniczne średniego błędu kwadratowego przy ocenie danego parametru oblicza się ze wzoru 7].

Jeżeli obciążenie można wyrazić, jak to zwykle bywa, w postaci funkcji linjowej parametru, $b = a\alpha$, to zgodnie z [7], wartością graniczną $\mathcal{E}^2(\alpha^{\bullet}_{a\alpha})$ powinno być $\mathcal{E}_0^2(\alpha^{\bullet}_{a\alpha}) = D_0^2(\alpha) \cdot (1+a)^2$ ([9]). *Otóż można wykazać, że jeżeli b jest $a\alpha$, to nie istnieje wogóle żadnej oceny obciążonej, której średni błąd kwadratowy byłby równy podanej wyżej wartości granicznej. Nierówności [3] i [4] można wtedy poprawić i znaleść takie obciążenie optymalne, $b_0 = a_0\alpha$, że ocena $(1+a_0)\bar{\alpha}$, gdy istnieje, będzie najlepsza, to znaczy, „najbliższa" ze wszystkich możliwych ocen linjowo obciążonych i nieobciążonych danego parametru. Ocenę tę będziemy nazywali oceną „linjowo wydajną".*

Aby znaleść obciążenie optymalne, minimalizujemy najpierw $\mathcal{E}^2(\alpha^{\bullet}_{a\alpha})$ ze względu na a, otrzymując [12]. Na podstawie [12] możemy już powiedzieć, że dla dowolnej oceny nieobciążonej, $\hat{\alpha}$, da się wyznaczyć optymalną (dla oceny α) funkcję linjową, $(1+a_0)\hat{\alpha}$, jak też dla dowolnej oceny linjowo obciążonej' α^*, optymalną funkcję linjową, $\frac{1+a_0}{1+a}\,\alpha^*$. Funkcje te są podane w [13], zaś średni błąd kwadratowy tych funkcji w [14]. Z [14] można zauważyć, że średni błąd kwadratowy będzie minimum gdy $D^2(\hat{\alpha}) = D_0^2(\alpha)$, t. j., gdy $\hat{\alpha}$ jest $\bar{\alpha}$. Ocena posiadająca ten minimalny graniczny błąd kwadratowy, a więc ocena linjowo wydajna, jest podana w [15], zaś jej średni błąd kwadratowy w [16]. Nieosiągalną wartość graniczną średniego błędu kwadratowego, podaną w [9], można teraz zastąpić osiągalnym minimum ([16]), zaś nierówności [3] i [4] nierównościami [17] i [18].

Gdy ocena linjowo wydajna nie istnieje, można użyć wzoru [12] do wyznaczenia względnie najlepszej oceny linjowej, $\alpha_0^{\bullet}$, pochodzącej od takiej charakterystyki próby, $\alpha^{\bullet}$, dla której średni błąd kwadratowy optymalnej funkcji linjowej jest mniejszy niż dla innych znanych charakterystyk próby ([19]). Formę funkcyjną takiej charakterystyki można nieraz wykryć metodą największego podobieństwa.

Gdy istnieje „wystarczająca łączna" ocena nieobciążona, (sufficient joint unbiased estimate), $\hat{\alpha}$, to najlepszą ocenę linjową z niej otrzymaną będziemy nazywali „linjowo wystarczającą oceną, $\mathring{\alpha}$, [20] przy innych parametrach niewiadomych."

Linjowo wydajną oceną σ^2 w populacji normalnej będzie, zgodnie z [15], ocena podana w [21], zaś linjowo wystarczającą przy μ niewiadomym — ocena podana w [25]. Średnie błędy kwadratowe tych ocen będą, zgodnie z [16], wyrażone wzorami [22] i [24].

Z powyższych rozważań wynika, że Fisherowskiej koncepcji wydajności, opartej o wartość graniczną zmienności dla ocen nieobciążonych, można przeciwstawić inną koncepcję, opartą o wartość graniczną średniego błędu kwadratowego dla ocen linjowo obciążonych. Wydajność w tym znaczeniu będziemy nazywali *wydajnością linjową*. Miarę jej określimy z rozważań następujących:

Jeżeli na podstawie n spostrzeżeń próby obliczamy ocenę jakiegoś parametru, która posiada taki stopień dokładności (bliskości) jaki można osiągnąć przez użycie oceny wydajnej obliczonej tylko z n' spostrzeżeń ($n' < n$), to nie możemy powiedzieć, że robimy *pełny* użytek z dostępnych nam informacji. Użytek jaki robimy z n spostrzeżeń, używając takiej oceny, jest równoważny pełnemu wykorzystaniu tylko n' spostrzeżeń z liczby n. Proporcję $\frac{n'}{n}$ będziemy zatem uważali za miarę wydajności oceny rozważanej. Miara ta będzie zależeć od przyjęcia tego czy innego określenia dokładności (bliskości). Jeżeli średni błąd kwadratowy weźmiemy za podstawę do określenia tej dokładności, to otrzymamy miarę wydajności linjowej w sensie *A*, jeżeli średni błąd bezwzględny, to w sensie *B*. Mówiąc poprostu o wydajności linjowej, będziemy mieli na myśli wydajność w sensie *A*. Tak więc wydajność linjową definjujemy jako $e_l(\alpha^*_n) = \frac{n'}{n}$, gdzie $e_l(\alpha^*_n)$ oznacza wydajność linjową linjowo obciążonej (lub nieobciążonej) oceny, α^*, obliczonej z n spostrzeżeń, zaś n' liczbę spostrzeżeń konieczną i wystarczającą do uzyskania linjowo wydajnej oceny o tej samej dokładności. W celu wyznaczenia n' do miary wydajności w sensie *A*, piszemy równanie [27] i rozwiązujemy je względem n'. Tak np., szacując σ^2 w populacji normalnej, piszemy $\mathcal{E}^2(\sigma^{2*}_n) = \frac{2\sigma^4}{n'+2}$, biorąc zaś $\frac{n}{n-1} \cdot s^2$ jako σ^{2*}_n, otrzymujemy $\frac{2\sigma^4}{n-1} = \frac{2\sigma^4}{n'+2}$, skąd $n' = n-3$. Tak więc miarą wydajności linjowej $\frac{n}{n-1} \cdot s^2$ jako oceny σ^2 jest $e_l\left(\frac{n}{n-1}\cdot s^2\right)=\frac{n-3}{n}$. W podobny sposób otrzymujemy miary wydajności dla ocen $\tilde{\sigma}_n^2=s_0^2$, $\dot{\sigma}_n^2=\frac{n}{n+1}\cdot s^2$ i dla s^2 podane w [32], [33] i [34].

Miary wydajności w znaczeniu Fisherowskim otrzymamy z [26 zastępując w [27] min $\mathcal{E}^2(a_n^*$, linjowe $b)$ przez $D_0^2(\hat{a}_n)$. Wyniki otrzymane będą identyczne z wynikami obliczonymi według Fisherowskiej definicji wydajności, $e(\hat{a}_n)=\frac{D_0^2(\hat{a}_n)}{D^2(\hat{a}_n)}$, jak długo mianowniki w wyrażeniach na $D_0^2(\hat{a}_n)$ będą równe dokładnie n. Z tego właśnie względu analogiczna równość $\frac{n'}{n}$ z $\frac{min\ \mathcal{E}^2(a_n^*\text{ linjowe } b)}{\mathcal{E}^2(a_n^*)}$ na ogół nie istnieje (gdyż np. mianownik $min\ \mathcal{E}^2(\sigma_n^{2*}$ linjowe $b)$ dla populacji normalnej jest $n+2$). Nie możemy zatem wyrazić miary wydajności linjowej w sposób analogiczny do Fisherowskiego, tj. w postaci stosunku granicznej wartości średniego błędu kwadratowego dla ocen linjowo obciążonych do średniego błędu kwadratowego oceny rozważanej. Definicja nasza miary wydajności jako proporcji $\frac{n'}{n}$ jest ogólniejsza od wyżej podanej Fisherowskiej.

Z podanych wzorów można zauważyć, że niektóre z ocen omawianych, które są bardziej wydajne w sensie Fisherowskim, w naszym sensie okazują się mniej wydajnymi. W szczególności s_0^2 jako ocena σ^2 w populacji normalnej posiada z naszego punktu widzenia wydajność linjową mniejszą od wydajności s^2, podczas gdy z punktu widzenia Fishera jest oceną najwydajniejszą (o wydajności równej 1). Pochodzi to stąd, że wydajność Fisherowska jest wydajnością relatywną, uwzględniającą tylko oceny nieobciążone, podczas gdy wydajność linjowa jest absolutna — uwzględniająca wszystkie oceny linjowo obciążone (łącznie z nieobciążonymi). Z naszego punktu widzenia nie widzimy żadnego względu, dla którego należałoby wyłączać oceny obciążone z estymacji jak też nie widzimy żadnego względu specjalnego, z wyjątkiem może wygody, dla którego należałoby oddawać pierszeństwo ocenom nieobciążonym. Średni błąd kwadratowy oceny uwzględnia wszystkie możliwe składniki błędu, włączając w nie i błąd stały, jakim jest obciążenie. Powszechny zwyczaj stosowania poprawki na błąd stały, jaką jest np. mnożenie s^2 przez $\frac{n}{n-1}$, nie da się więc uzasadnić z naszego punktu widzenia. Poprawka wymagana przez kryterjum najmniejszego średniego błędu kwadratowego działa w odwrotnym kierunku: pomnożenia s^2 przez $\frac{n}{n+1}$. Poprawka wartości pojedynczej na błąd stały nie da się logicznie odróżnić od poprawki na błąd zmienny. Jeżeli pierwsza wprowadza jeszcze większy błąd zmienny, trudno się oprzeć wrażeniu, że poprawka taka czyni ocenę jeszcze mniej pewną. Rola błędu stałego uwydatnia się dopiero przy sumowaniu wartości pojedynczych. W takich jednak wypadkach, jak np. przy szacowaniu σ^2 wystarczy tylko posiadać wiadomości co do wielkości prób, z których są obliczone pojedyncze oceny σ^2 aby móc wyznaczyć taką skombinowaną ocenę obciążoną, która będzie „bliższa" parametru szacowanego, niż najlepiej skombinowana ocena nieobciążona.

Istotnie, przypisując wagi, $w_1, w_2, \ldots w_k$, indywidualnym ocenom linjowym α, $\overset{*}{\alpha}_1$, $\overset{*}{\alpha}_2, \ldots \overset{*}{\alpha}_k$, obliczonym z k prób niezależnych o wielkościach $n_1, n_2, \ldots n_k$, takie, aby skombinowana ocena, $\sum_{i=1}^{k} w_i \overset{*}{\alpha}_i$, była możliwie „najbliższa", otrzymamy, minimalizując [36], układ rozwiązań dla w_i podany w [37]. Średni błąd kwadratowy tak skombinowanej oceny obciążonej podany jest w [38]. Gdy $\alpha = \sigma^2$, $\overset{*}{\alpha}_i = s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$, populacja zaś jest normalna, otrzymujemy $w_i = \frac{n_i}{n+2-k}$ [41]. Średni błąd kwadratowy tak skombinowanej oceny obciążonej, $\sum_{i=1}^{k} n_i s_i^2/(n+2-k)$ jest równy $\frac{2\sigma^4}{n+2-k}$ ([42]), a więc jest mniejszy od średniego błędu kwadratowego najlepiej skombinowanej oceny nieobciążonej, $\sum_{i=1}^{k} n_i s_i^2/(n-k)$, który jest równy $\frac{2\sigma^4}{n-k}$ ([43]).

Ważenia powyższego dokonaliśmy, oczywiście, w założeniu, że średnie arytmetyczne prób nie są podane, względnie, że próby były wylosowane z różnych populacji, różniących się między sobą co do średnich arytmetycznych, ale posiadających takie same zmienności. Gdyby średnie arytmetyczne prób poszczególnych były podane, próby zaś losowane były z tej samej populacji, to najlepszą ocenę skombinowaną otrzymalibyśmy, łącząc wszystkie spostrzeżenia razem, tj.

byłaby nią znana już nam wystarczająca ocena linjowa σ^2 przy niewiadomej średniej arytmetycznej populacji, μ, podana w [25], a którą obliczyć można na podstawie danych z k prób w sposób wskazany w [44], gdzie $\bar{x} = \frac{1}{n} \sum_{i=1}^{k} n_i \bar{x}_i$.

Możemy się spotkać z zarzutem, że gdyby nasze oceny linjowe, które są obciążone ujemnie, zostały użyte do sprawdzianów hipotez statystycznych, to na dłuższą metę przesadzilibyśmy „znamienność" (significance) wyników. Zarzut ten mógłby być słuszny, gdyby trzeba było stosować oceny σ, jak to było dawniej, do niedokładnych sprawdzianów statystycznych. Od czasu jednak wprowadzenia dokładnych sprawdzianów „Studenta" nie zachodzi potrzeba szacowania nieznanej σ, wobec czego zarzut powyższy staje się bezprzedmiotowy. Co prawda, wyrażenia na oceny nieobciążone σ^2 wchodzą do tych sprawdzianów dokładnych, ale pierwiastki kwadratowe z tych wyrażeń, które figurują w sprawdzianach, nie są ocenami nieobciążonymi σ. Ważną tu rzeczą jest zauważyć, że *forma funkcyjna jakiegokolwiek sprawdzianu dokładnego może być wyprowadzona zupełnie niezależnie od jakichkolwiek twierdzeń dotyczących estymacji.* Wydaje się wskazanym, nawet jeśli niektórzy autorzy postępują inaczej, nie mieszać zagadnień sprawdzianów statystycznych z zagadnieniem oceny parametru jedną liczbą.

Przejdziemy teraz do wypadku specjalnego, w którym kryterjum B może być zastosowane. Pewnej próby wprowadzenia oceny obciążonej zamiast oceny nieobciążonej dokonał P i t m a n, który oparł swoje kryterjum „bliskości" na wielkościach bezwzględnych zamiast na kwadratach. Pitman uważa ocenę α, α_1^*, za „bliższą" od oceny α_2^*, jeśli prawdopodobieństwo, że $|\alpha_1^* - \alpha| < |\alpha_2^* - \alpha|$ jest większe od $1/2$. Jednakże jego zasada określania oceny „najbliższej" jest analogiczna do zasady najmniejszych kwadratów i podlega tym samym ograniczeniom. Logikę postępowania Pitmana można pokrótce opisać w sposób następujący:

Korzystając z tego, że minimalizacja $\epsilon(\alpha^*) = E|\alpha^* - \alpha|$ ze względu na α daje rozwiązanie $\alpha = Me(\alpha^*)$, gdzie $Me(\alpha^*)$ jest medjaną α^* w rozkładzie losowym (sampling distribution), Pitman obiera za najlepszą ocenę α takie α^*, medjana którego równa się α, np., kładąc $\alpha = \sigma^2$, bierze $\sigma^{2*} = \frac{\sigma^2}{Me(s^2)} \cdot s^2$ ([45]). Postępowanie Pitmana jest analogiczne do postępowania tych, którzy zgodnie z rozwiązaniem najmniejszych kwadratów, $a = E(a^*)$, biorą za najlepszą ocenę a takie a^*, którego nadzieja matematyczna jest równa a, np., kładąc $a = \sigma^2$, biorą $\sigma^{2*} = \frac{n}{n-1} \cdot s^2$. W rzeczywistości wyrażenie Pitmanowskie nadaje się lepiej do oceny σ^2 niż do oceny innego parametru z punktu widzenia kryterjum B, podobnie jak $\frac{n}{n-1} \cdot s^2$ nadaje się lepiej do oceny σ^2 niż do oceny innego parametru z punktu widzenia kryterjum A. Jako ocena σ^2, wyrażenie Pitmanowskie jest dalekie od tego aby było najlepsze spośród innych możliwych ocen. Istotnie, można je otrzymać, minimalizując jeden tylko (drugi) z dwóch składników, na które można rozłożyć $E \mid As^2 - \sigma^2 \mid$, względem A, zgodnie ze wzorem [46], gdzie $F(As^2)$ oznacza funkcję dystrybucyjną (distribution function) zmiennej ewentualnej As^2. W założeniu populacji normalnej spółczynnik przy s^2 w ocenie Pitmanowskiej, mianowicie $\frac{\sigma^2}{Me(s^2)}$, można wyrazić jako funkcję samego tylko n, ponieważ przy tym założeniu $s^2 = \frac{\sigma^2 \chi^2_{\nu=n-1}}{n}$, a więc $\frac{\sigma^2}{Me(s^2)} = \frac{n}{Me(\chi^2)_{\nu=n-1}}$, rozkład zaś χ^2 zależy od jednego tylko parametru, ν. Pozostawałoby jeszcze zatem obliczyć $Me(\chi^2)_{\nu=n-1}$, co też Pitman czyni podając wzór przybliżony [49]. Rzecz jasna, że dokładne wartości $Me(\chi^2)$ można odczytać z tablicy χ^2 dla $\nu = 1, 2, \ldots 30$ przy $P = 0{,}50$.

Jak widać, ocena Pitmanowska jest dodatnio obciążona i z punktu widzenia kryterjum A jeszcze bardziej oddalona od prawdziwej wielkości σ^2 niż ocena nieobciążona, $\frac{n}{n-1} \cdot s^2$.

Najlepszą, z punktu widzenia kryterjum B, ocenę σ^2, mającą postać As^2 można otrzymać przez zminimalizowanie $\epsilon(As^2) = E \mid As^2 - \sigma^2$ ze względu na A. W tym celu wyprowadzamy wpierw ogólną postać pochodnej $E \mid Aa^* - a \mid$ ze względu na A, przyrównanej do zera, tj. pochodnej średniego błędu bezwględnego dowolnej charakterystyki próby użytej do oceny dowolnego parametru (wzór [54]), następnie zaś podstawiamy σ^2 na miejsce a i s^2 na miejsce a^*, otrzymując równanie [55]. W założeniu populacji normalnej dokonujemy zmiany zmiennej podcałkowej z s^2 na χ^2, otrzymując równanie [56]. Korzystając z właściwości funkcji częstotliwości χ^2, wyrażonej w [57], przekształcamy lewą stronę [56] na wyrażenie będące funkcją dystrybucyjną χ^2 o liczbie stopni swobody zwiększonej o 2 (wzór [59]. Z równania tego wynika bezpośrednio, że $A = \frac{n}{Me(\chi^2)_{\nu=n+1}}$. Tak więc najlepszą oceną σ^2 mającą postać As^2 jest $\frac{n}{Me(\chi^2)_{\nu=n+1}} \cdot s^2$, nie zaś Pitmanowskie $\frac{n}{Me(\chi^2)_{\nu=n-1}} \cdot s^2$.

Analogicznie wyprowadzamy najlepszą, z punktu widzenia kryterjum B, ocenę σ^2 mającą postać As_0^2 podaną w [63] i porównujemy ją z drugą oceną Pitmanowską — mającą postać As_0^2 — (wzór [64]). Widzimy, że obie nasze oceny różnią się od odpowiadających im ocen Pitmanowskich tym, że χ^2 w naszych ocenach jest wzięte z dwoma stopniami swobody więcej niż w ocenach Pitmanowskich. Okoliczność ta powoduje, że nasze oceny są obciążone ujemnie, podczas gdy Pitmanowskie są obciążone dodatnio.

W celu wyprowadzenia wzorów na średnie błędy bezwzględne tych ocen, piszemy, wychodząc z [52], wzory [65] i [66]. Przechodząc do zmiennej χ^2, otrzymujemy [67] i [68]. Korzystając z relacji [69], która wynika z [48] i [58], otrzymujemy [70] i [71], z których to wzorów, przez podstawienie odpowiednich wartości dla A, otrzymujemy wyrażenia na średnie błędy bezwzględne dla wszystkich dziewięciu ocen rozważanych σ^2. Wzory te są specjalnie przystosowane do korzystania z tablicy χ^2. [72], [73], [74], [75], [76], [77], [78], [79] i [80]. Gdyby chodziło o ścisłe obliczenie wyrażeń typu $1 - \int_0^{\chi_0^2} f_\nu(\chi^2)\, d\chi^2$, występujących w tych wzorach, to, w celu uniknięcia związanej z użyciem tablicy χ^2 interpolacji, należałoby korzystać ze wzoru Fishera podanego w [81] dla ν parzystych.

Dla czterech ocen typu B i dla dwóch ocen typu A (nieobciążonych) można wyprowadzić prostsze wzory na średnie błędy bezwzględne ocen. Korzystając ze wzoru rekurencyjnego [82] i właściwości funkcji gamma, wyprowadzamy, przechodząc do całek oznaczonych, relację [83], wyrażającą funkcję dystrybucyjną $\chi^2{}_{\nu=k}$ przy pomocy funkcji dystrybucyjnej $\chi^2{}_{\nu=k+2}$. Dokonując odpowiednich podstawień, otrzymujemy wzory [85] i [87]. Ze wzorów tych wyprowadzamy proste wyrażenia na średnie błędy bezwzględne oceny Pitmanowskiej, $\frac{n}{Me(\chi^2)_{\nu=n-1}} \cdot s^2$ i oceny naszej, $\frac{n}{Me(\chi^2)_{\nu=n+1}} \cdot s^2$ ([86] i [88]). Porównując te wyrażenia, stwierdzamy, że średni błąd bezwzględny oceny Pitmanowskiej jest równy średniemu błędowi bezwzględnemu oceny naszej opartej o 2 mniej spostrzeżenia. W podobny sposób wyprowadzamy wzory na średnie błędy bezwzględne oceny Pitmanowskiej, $\frac{n}{Me(\chi^2)_{\nu=n}} \cdot s_0^2$ i oceny naszej, $\frac{n}{Me(\chi^2)_{\nu=n+2}} \cdot s_0^2$ ([90] i [91]), stwierdzając i tu, że średni błąd bezwzględny oceny Pitmanowskiej jest równy średniemu błędowi bezwzględnemu oceny naszej opartej o 2 mniej spostrzeżenia. ($s_0^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$, gdzie μ jest śr. ar. populacyjną). Zestawiając razem te wyniki oraz porównując [86] z [90] i [88] z [91], otrzymujemy relację [94], porządkującą 4 oceny typu B ze względu na ich wydajność linjową w sensie B. Stwierdzamy analogiczne ustosunkowanie się odpowiadających im ocen typu A ([95]). Zestawiając [94] z [95], stwierdzamy odpowiedniość miar wydajności czterech par ocen z punktu widzenia dwuch kryterjów.

Wyprowadzamy jeszcze dwa proste wzory dalsze na średnie błędy bezwzględne ocen nieobciążonych, $\frac{n}{n-1} \cdot s^2$ i s_0^2 ([96] i [97]). Stwierdzamy przytym relację [98].

Fig. 1 podaje średnie błędy bezwzględne (zaznaczone na linjach ciągłych) i pierwiastki kwadratowe ze średnich błędów kwadratowyh (zaznaczone na linjach przerywanych) ocen σ^2 w populacji normalnej, opartych o $n = 1, 2, 4, 8, 16$, wyrażonych w stosunku do σ^2, dla dziewięciu ocen uporządkowanych według stopnia ich wydajności w sensie B. Względność pojęcia wydajności występuje wyraźnie na tym wykresie. Jak już mówiliśmy na początku, w wypadkach wyników rozbieżnych otrzymanych przy użyciu dwóch kryterjów, ostateczna decyzja powinna być raczej zgodna z kryterjum B, niż z kryterjum A. Gdy chodzi o ocenę porównawczą dwóch albo więcej charakterystyk próby, w większości wypadków możemy stosować którekolwiek z kryterjów. Rozbieżności są stosunkowo małe i dotyczą tylko ocen położonych blisko dwóch alternatywnych „ocen najlepszych". Oceny położone dalej od punktów spornych są spółbytne z obu kryterjami.

Stosowanie kryterjum B zamiast A ilustruje również względność naszego pojęcia o obciążeniu. Podczas gdy oceną nieobciążoną w sensie A jest dowolne a^* spełniające warunek $E(a^*) = a$, za ocenę nieobciążoną w sensie B moglibyśmy uważać dowolne a^* spełniające warunek $Me(\alpha^*) = \alpha$. Rzecz jasna, że ocena nieobciążona w jednym znaczeniu będzie oceną obciążoną w znaczeniu drugim, jeżeli rozkład losowy danej oceny jest asymetryczny. W obu znaczeniach jednak pojęcie obciążenia jest całkiem nieistotne do ustalania kryterjów wyboru dobrych ocen. W wyniku takiego ustalania oceny najlepsze okazują się naogół ujemnie obciążonymi (w obu znaczeniach), niezależnie od tego, czy przyjęliśmy kryterjum A czy B.

Głównym zagadnieniem jest nie dylemat: kryterjum A czy B, tylko dokładne spełnienie wymagań któregokolwiek z nich. Jak dotąd, żadna z zasad ogólnie stosowanych nie spełnia tego wymagania. Z drugiej strony trzeba przyznać że w wypadkach trudniejszych, zasada przyjęta przez nas w jej obu odmianach nie daje sposobu na wyznaczenie optymalnej postaci funkcyjnej oceny. Pod tym względem metoda największego podobieństwa ma zdecydowaną przewagę. Jednakże, jeśli się ograniczyć do ocen linjowo obciążonych, można osiągać dokładne rozwiązania przynajmniej w wypadkach prostych, jak to próbowaliśmy wykazać w niniejszej pracy. W innych wypadkach oceny można ulepszać, kombinując rozwiązania „największego podobieństwa" z zasadą „optymalnego obciążenia".

Tablica I podaje wartości liczbowe spółczynników przy s^2 i s_0^2 w ośmiu ocenach σ^2 dla populacji normalnej — przy n przyjmującym wartości od 1 do 30.

Wielkopolskie Zakłady Graficzne w Poznaniu, Zakład Główny — 1000 — 67/1. 49.