ON CERTAIN CRITERIA FOR OPTIMUM ESTIMATION

by

Bernard Ostle

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Mathematical Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1949

TABLE OF CONTENTS

I. INTRODUCTION

One of the most frequent problems arising in statistics is that of estimating the values of the parameters of a population by means of a sample from that parent population. We will wish, naturally, to obtain estimates which are "best" in some acceptable sense. Intuitively, "best" would seem to say that we wish to find an estimate such that it will be as close as possible to the true parameter value with as large a probability as possible. The introduction of the concept of probability into our requirements for a best estimate is necessary because the estimate is a function of the sample values, and is hence a random variable. It will not, then, coincide with the true parameter value except by chance and it will have a distribution which, we hope, will be centered in some sense at the true value. We will restrict ourselves to the consideration of random samples only, since if the sample is not random and nothing precise is known about the nature of the bias operating when it was chosen, very little can be inferred from it about the parent population.

Let us attempt to make more precise our statement that a "best" estimate will be one for which the probability that it will be as close as possible to the true value will be as large as possible. We assume that the form of the parent population is known except for the values of certain parameters. For the purposes of exposition we shall restrict ourselves, primarily, to the case of one unknown parameter. Following the terminology introduced by Pitman (37), we distinguish between a method of estimation, called an <u>estimator</u>, and the value to which it gives rise in particular cases, called an estimate. The merit of an estimator is judged by the population of estimates to which it gives rise.[a] Several writers[b] have advanced criteria by which we may provide good estimators. Since the formulation of criteria of this nature depend on considerations other than mathematical requirements, we should expect to find more than one criterion set forth in the literature.[c] It is the purpose of this work to set out the relationships which have been shown to exist between these several criteria and to add some new relationships which do not appear to have been discussed previously.

A. Desirable Properties of Good Estimators

Before setting down the criteria referred to above, it is necessary that we enumerate the properties which seem desirable in a good estimator.

(1) An estimator t_n, based on a sample of n values, will be said to be a consistent estimator of θ if, for any positive ϵ and δ, however small, there is some N such that

$$P\left(\left|t_n - \theta\right| < \epsilon\right) > 1 - \delta \qquad \text{for } n > N. \tag{1}$$

t_n is said to converge in probability to θ. Thus t is a consistent estimate of θ if it converges to θ in probability as n increases.

a. Kendall, M. G., The Advanced Theory of Statistics, London, Charles Griffin and Co. Ltd. 1946, Vol. II, p. 2.
b. References 4, 10, 29, 36, 41, 42, in the Bibliography.
c. Neyman, J. Lectures and Conferences on Mathematical Statistics, Washington, D. C., The Graduate School of the United States Department of Agriculture. 1937, pp. 127-142.
Wald, A., On the Principles of Statistical Inference, Notre Dame, Indiana, University of Notre Dame. 1942.

(2) The property of consistency being a limiting property as the sample size tends to infinity, it is evident that we need some similar property for finite n. This is realised by requiring that the mean value of t, for all n, shall be θ. That is, we define what is called an unbiased estimator by the relation

$$E_\theta (t) = \theta \tag{2}$$

where $E_\theta (t)$ denotes the expected value of t under the assumption that θ is the true value of the parameter.

(3) In a great many estimation problems it is possible to construct estimators $t(x_1, x_2, \ldots, x_n)$ such that $\sqrt{n}\,(t-\theta)$ has a normal distribution with zero mean in the limit as the sample size n increases. Confining our attention to this class of estimators, there may be one or more estimators which will have a limiting variance which is smaller than the limiting variances of the other estimators. The estimators which have the smallest limiting variance are called asymptotically efficient estimators of θ. It is to be observed that asymptotically efficient estimators are necessarily consistent.

(4) It is convenient to distinguish between an efficient estimator and an asymptotically efficient estimator. Cramér[a] has proved that under certain conditions the variance of t, where t is an unbiased estimator of θ, can never be less than $\left[n\,E\left(\frac{\partial \ln f}{\partial \theta}\right)^2\right]^{-1}$ where n is the fixed size of sample. If this minimum value is achieved, t is said to be an efficient estimator of θ. We note that an efficient estimator exists only under

a. Cramér, H., Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946, p. 480.

rather restrictive conditions, whereas the existence of an asymptotically efficient estimator can be proved as soon as certain general regularity conditions are satisfied.

(5) If $(x_1, x_2, \ldots, x_n)$ is a sample from a population with distribution $f(x;\theta)$ and if $t(x_1, x_2, \ldots, x_n)$ is an estimator of θ such that the likelihood function is expressible in the form

$$L(x_1, x_2, \ldots, x_n;\theta) = L_2(x_1, x_2, \ldots, x_n) \; L_1(t;\theta) \tag{3}$$

where L_1 does not contain the x's otherwise than in the form t and L_2 is independent of θ, t is said to be a <u>sufficient</u> estimator for θ.

We might note the following theorem, a proof of which is furnished by Dugué (8), relating to sufficient estimators:

If t_1 is sufficient and t_2 is any other estimator of θ (not a function of t_1) the joint distribution of t_1 and t_2 may be put in the form

$$dF = f_1(t_1;\theta) \quad f_2(t_2;t_1) \, dt_1 \, dt_2 \tag{4}$$

where f_2 does not contain θ. From (4) it follows that for any given t_1 the distribution of t_2 is equal to $f_2(t_2;t_1)\,dt_2$; i.e., is independent of θ. Consequently, if we know t_1, the probability of any range of values of t_2 is the same for all θ. The distribution of t_2 given t_1, therefore, can provide information about t_1 but not about θ, so that any information provided by t_2 is of no use. It is in this sense that t_1 may be said to contain all the information in the sample regarding the parameter θ.

We note that sufficiency, unlike consistency and asymptotic efficiency, is not merely an asymptotic property. As we have defined it, a sufficient estimator, if it exists at all, is unique except that if t satisfies (3) any function of t will obviously satisfy the same relation. It is necessary, therefore, to choose from all such functions one which gives a consistent estimator or, if possible, an unbiased estimator. If we choose such a function as our estimator of the parameter in question, it follows that this particular sufficient estimator will be an asymptotically efficient estimator provided that an asymptotically efficient estimator exists. If we could always find sufficient estimators the problem of estimation would be greatly simplified, but unfortunately sufficiency is the exception rather than the rule.

B. Principles for Obtaining Estimators

Probably the most important method of providing estimators is that of maximum likelihood which was advanced by R. A. Fisher (10). Incidentally, it was Fisher who formulated the criteria we have listed above as being desirable properties of good estimators.

If the frequency function of the parent population is $f(x;\theta)$, the likelihood function of a sample of size n is defined to be

$$L = \prod_{i=1}^{n} f(x_i;\theta). \tag{5}$$

The principle of maximum likelihood says that if we find a function $t(x_1, x_2, \ldots x_n)$ which maximizes L for variations in θ, providing such a function exists, then t is called a maximum likelihood estimator of θ. That is, we solve the following equation

$$\frac{\partial L}{\partial \theta} = 0 , \quad \frac{\partial^2 L}{\partial \theta^2} < 0 \tag{6}$$

or what is equivalent (since L is positive), and often more convenient, we look for a solution of

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = \frac{\partial}{\partial \theta} \ln L = 0 \tag{7}$$

There is no general argument which will show that maximum likelihood estimators are the best possible estimators.[a] The principle of maximum likelihood will be a good one if it leads to results which conform to our preconceived ideas as to desirable properties of estimators. After examining the properties of maximum likelihood estimators, it will be apparent that Fisher's principle is a useful one. The important properties of maximum likelihood estimators are[b]:

(i) Maximum likelihood estimators have the true parameter values near the centers of their distributions, but in general they are not unbiased.

(ii) If (a) the frequency function $f(x;\theta)$ is continuous in x throughout its range, and (b) if $f(x;\theta)$ is continuous and monotonic is θ in some θ-interval containing the true value of θ and for all x in some x-interval, then the maximum likelihood estimator of θ, say t, is consistent. A similar property can be stated for the case of discontinuous variates.

(iii) If t is a maximum likelihood estimator for θ and if $u(\theta)$ is any single valued function of θ, then $u(t)$ is a maximum likelihood estimator of $u(\theta)$.

---

a. Mood, A. M., Theory of Statistics, Iowa State College, 1948.
b. References 4, 24, 28, 44 in the Bibliography.

(iv) If (a) the frequency function $f(x;\theta)$ is continuous in x throughout its range, and (b) in a θ-interval containing the true value of θ, $\frac{\partial f}{\partial \theta}$ approaches a continuous function of θ as x tends to infinity, and $\frac{\partial f}{\partial \theta}$ does not vanish in some interval, then the maximum likelihood estimator t tends to normality for large n.

(v) If t be any estimator of θ, the range of $f(x;\theta)$ is independent of θ, and in large samples t is distributed normally about mean θ_0 (the true value of θ) with finite variance, then a maximum likelihood estimator (if one exists) is an asymptotically efficient estimator.

(vi) If a sufficient estimator exists, it is a function of the maximum likelihood estimator. We remember that not all parameters have sufficient estimators, but if a sufficient estimator does exist, it can be shown that the maximum likelihood estimator will be a sufficient estimator.

A second criterion which is often used to obtain estimators is to consider the class of all unbiased estimators and, in particular, the element of this class which has <u>minimum variance</u>. This is accomplished by finding a value of $t = t(x_1,\ldots,x_n)$ such that

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (t-\theta)^2 \; L \prod_{i=1}^{n} dx_i = \text{minimum} \quad , \tag{8}$$

where $L = \prod_{i=1}^{n} f(x_i;\theta)$, subject to the restriction that

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} t \; L \prod_{i=1}^{n} dx_i = \theta \tag{9}$$

or

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} t \frac{\partial L}{\partial \theta} \prod_{i=1}^{n} dx_i = 1 \tag{10}$$

provided that the range of x is independent of θ or that $f(x;\theta)$ vanishes at any extreme which depends on θ. This is equivalent to finding the unconstrained maximum of

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \left((t-\theta)^2 - 2\lambda t \frac{\partial L}{\partial \theta} \right) \prod_{i=1}^{n} dx_i \tag{11}$$

where λ is an unspecified parameter which may depend on θ but not on the x's. A solution exists if we can express $\frac{\partial \ln L}{\partial \theta}$ in the form

$$\frac{\partial \ln L}{\partial \theta} = \frac{t-\theta}{\lambda} . \tag{12}$$

We now note another principle which has been suggested for providing estimators. The data are grouped into cells with expected frequency typified by μ_j and observed frequency by m_j, then the function

$$\chi^2 = \sum \frac{(m_j - \mu_j)^2}{\mu_j} \tag{13}$$

$$= \sum \frac{m_j^2}{\mu_j} - n$$

where

$$n = \sum \mu_j = \sum m_j \tag{14}$$

is minimized with respect to the μ_j. This method is called the method of <u>minimum chi-square</u>.

If $h(\theta)\,d\theta$ is the prior probability of θ then, according to Bayes' theorem, the posterior probability of θ is given by

$$P\,(\theta \mid x_1,\ldots,x_n) = L(x_1,\ldots,x_n;\theta)\;h(\theta)d\theta. \tag{15}$$

One then determines the "most probable" value of θ by maximizing $P(\theta \mid x_1,\ldots,x_n)$ if we assume $h(\theta)$ to be known. The principles of inference with which we have been concerned do not require the notion of the probability of θ and, even if they did, would not give any guide to the nature of the function $h(\theta)$. One usually assumes that the existence of $h(\theta)$ denotes a prior measure of belief, but, in order to find the most probable value of θ, some further assumption as to its values (comparable to Bayes' postulate that for a finite range $h(\theta)$ is constant) is necessary. The method outlined above is generally referred to as the <u>method of "inverse" probability</u>.

Another method of estimating unknown parameters which has been greatly used in the past is the <u>method of moments</u>. This consists of fitting distributions by identifying lower moments. Fisher (10) has shown that the method of moments when applied in fitting Pearsonian curves has an efficiency exceeding 80 per cent only in the restricted region for which β_2 lies between the limits 2.65 and 3.42 and for which β_1 does not exceed 0.1. Hence this method does not, in general, provide the asymptotically most efficient estimators of unknown parameters.

The next method for providing estimators of parameters which we shall mention is the familiar <u>method of least squares</u>. The most important case occurring in statistical theory concerns regression equations. As the name suggests, there is an analogy to the method of minimum chi-square, which also minimizes a sum of squares to provide estimators of the population parameters under consideration.

Pitman (36) has suggested the principle of closeness as one which has an intuitive appeal when we consider obtaining estimators for an unknown parameter. If t_1 and t_2 are two estimators of θ, then t_1 is a closer estimator of θ than is t_2 if

$$P\left(|t_1-\theta| < |t_2-\theta|\right) > \frac{1}{2} \tag{16}$$

If t_1 is closer than any other estimator, t, then it is said to be the closest estimator of θ. The criterion of closeness means that if we make use of a closest estimator then our error will be smaller, more than fifty per cent of the time, than if we used any other estimator. We should also note that (16) must hold for all permissible values of θ for the definition to be of value.

Wald (41) has considered a much more general problem of which estimation is one part. Wald's formulation of the problem is such that $\overline{\theta}(E)$, where $E = (x_1, x_2, \ldots, x_n)$; i.e., the sample point, is a best estimate of θ relative to a given weight function $W(\theta, \overline{\theta})$ if the system of regions determined by $\overline{\theta}(E)$ is a best system of regions[a] relative to the weight function considered. Corresponding to any estimate $\theta(E)$, we define a risk function

$$r(\theta) = \int_M W(\theta, \theta(E))\, p(E|\theta)\, dE \tag{17}$$

and by choosing a $\theta(E)$ to minimize "the maximum risk for variation in θ" (if such a choice is possible) we obtain a best estimate as defined by Wald. Under certain general conditions, among which is one that the value of the

---

a. For a definition of a "best system of regions" see Wald (41), p. 303.

weight function shall depend only on $u = |\theta - \bar{\theta}|$, it has been shown that if a maximum likelihood estimator $\hat{\theta}(E)$ exists then $\hat{\theta}(E)$ is a best estimate in the sense defined above.

II. RELATIONS BETWEEN SEVERAL PRINCIPLES FOR OBTAINING ESTIMATORS

A. Review of the Literature

It has been shown by Fisher (11, 12) that for large samples the minimum chi-square estimator tends to the maximum likelihood estimator. Kendall (24) states that on theoretical grounds there seems no reason to use minimum chi-square instead of maximum likelihood. The method has some practical value, however, where the maximum likelihood equations are difficult to solve. The method of inverse probability for providing estimators may also be shown to be equivalent (for large n) to the method of maximum likelihood.

If in the method of least squares we assume the residuals to be normally and independently distributed with variance σ^2, the method is equivalent to that of maximum likelihood. If such an assumption is not made, the two methods may give different results and then the justification for applying the least squares principle becomes more or less empirical.

Aitken and Silverstone (1), starting with postulates of (i) unbiased or consistent estimators, and (ii) minimum variance of the estimators, have set forth fairly simple conditions under which maximum likelihood provides an estimator actually possessing minimum variance, even though the sample is finite and the distribution of the estimator is not normal. This is similar to the result stated by Cramér following the assumption of the existence of an efficient estimator, the difference being some added requirements in the assumptions necessary for the existence of an efficient estimator. It is noted that the types of distribution functions which are amenable to the postulates of unbiased estimator and minimum variance also admit sufficient statistics for the estimation of the parameters being considered.

The relationship between the criteria of minimum variance for unbiased estimators and closeness was treated in a special case by Geary (17). Considering two estimators, t_1 and t_2, of θ to have a joint normal distribution, he demonstrated that the concepts of closeness and minimum variance were equivalent.

B. Statement of the Problem

What we shall consider here is a more general situation than that analyzed by Geary. Taking any two estimators, t_1 and t_2, of the parameter θ, subject only to the restriction that their expected values be equal to θ, that is, $E(t_1) = E(t_2) = \theta$, certain conditions under which the closer estimator has the smaller variance are determined. Further, we obtain rather general conditions on the form of the joint distribution function of t_1 and t_2 so that the concepts of closeness and smaller variance will be equivalent. Also, since no definition of "asymptotic closeness" appears to have been considered, we attempt such a definition and compare the concept with those of asymptotic smaller variance and asymptotic efficiency. An extension is made to the case of biased estimators and a generalization of the concept of closeness is examined in the multi-parameter situation.

III. THE EQUIVALENCE OF CLOSENESS AND MINIMUM VARIANCE

A. Continuous Distribution Functions

In this section we will be concerned only with those variates which admit a probability density function. We may represent the probability density function of a random variate W by f(w), where f(w) satisfies the following conditions:

(i) $f(w) \geqslant 0$ and is continuous for $-\infty \leqslant w \leqslant \infty$

(ii) $\displaystyle\int_{-\infty}^{\infty} f(w)\,dw = 1$

(iii) $\displaystyle P(W \leqslant w) = \int_{-\infty}^{w} f(w)\,dw = F(w)$

Now, given that t_1 and t_2 are two estimators of the parameter θ, such that $E(t_1) = E(t_2) = \theta$, we may consider without loss of generality the variates X and Y, where $X = t_1 - \theta$, and $Y = t_2 - \theta$. Obviously, $E(X) = E(Y) = 0$. Let t_1 and t_2 have variance σ_1^2 and σ_2^2 respectively.

Let

$$Q = P\left(|t_1-\theta| < |t_2-\theta|\right) = P\left(|X| < |Y|\right). \tag{1}$$

Let us consider the variate $Z = kX$, $k > 0$.

Define

$$Q(k) = P\left(|Z| < |Y|\right) = P\left(|kX| < |Y|\right)$$
$$= P\left(|X| < |Y/k|\right) = \int_{-\infty}^{\infty}\int_{-|y/k|}^{|y/k|} f(x,y)\;dx\,dy \tag{2}$$

where $f(x,y)$ is the joint probability density function of X and Y.

We note that as $k \to 0$, $Q(k) \to 1$ and that as $k \to \infty$, $Q(k) \to 0$. We see, then, that if we define $Q(0) = 1$ and $Q(\infty) = 0$, $Q(k)$ is such that $0 \leqslant Q(k) \leqslant 1$. This is as expected since $Q(k)$ was defined to be a probability. We now prove a lemma which is necessary for further work.

Lemma.

If X and Y are two random variates such that:

(i) $E(X) = E(Y) = 0$

(ii) $f(x,y)$ is the joint probability density function of X and Y

(iii) $f(x,y) > 0$ for $x = y = 0$

then $Q(k) = P(|kX| < |Y|)$, where $k > 0$, is:

(a) continuous with respect to k

(b) $0 \leqslant Q(k) \leqslant 1$

(c) strictly monotone decreasing with increasing k.

Proof:

We shall first establish the property of continuity. We say that $Q(k)$ is continuous if, for any given $\epsilon > 0$ there exists a $\delta > 0$ such that for $|k-k_o| < \delta$, $|Q(k)-Q(k_o)| < \epsilon$.

Now

$$Q(k)-Q(k_o) = \int_{-\infty}^{\infty} \int_{-|y/k|}^{|y/k|} f(x,y)\,dx\,dy - \int_{-\infty}^{\infty} \int_{-|y/k_o|}^{|y/k_o|} f(x,y)\,dx\,dy \;:\; k > k_o \quad (3)$$

$$= 1 - \int_{-\infty}^{\infty} \int_{-k|x|}^{k|x|} f(x,y)\, dy\, dx - \left[1 - \int_{-\infty}^{\infty} \int_{-k_o|x|}^{k_o|x|} f(x,y)\, dy\, dx \right] \tag{4}$$

$$= \int_{-\infty}^{\infty} \int_{-k_o|x|}^{k_o|x|} f(x,y)\, dy\, dx - \int_{-\infty}^{\infty} \int_{-k|x|}^{k|x|} f(x,y)\, dy\, dx \tag{5}$$

$$= \int_{-\infty}^{\infty} \left[\int_{k|x|}^{k_o|x|} f(x,y)\, dy + \int_{-k_o|x|}^{-k|x|} f(x,y)\, dy \right] dx \tag{6}$$

$$= \int_{-\infty}^{\infty} h(x)\, dx \tag{7}$$

where $h(x)$ is defined by the equation:

$$h(x) = \int_{k|x|}^{k_o|x|} f(x,y)\, dy + \int_{-k_o|x|}^{-k|x|} f(x,y)\, dy\ . \tag{8}$$

Since $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$, we can choose R so that

$$\int \int_{x^2+y^2>R^2} f(x,y)\, dx\, dy < \epsilon/2.$$

Also, since $f(x,y)$ is bounded, we have $f(x,y) \leqslant M$.

Therefore

$$\int_{x^2+y^2<R^2} h(x)\,dx \leqslant 4M \int_{x^2+y^2<R^2} dA = 4M\left[\frac{1}{2}R^2\alpha\right] = 2MR^2\alpha \tag{9}$$

$$\text{and } k_o < \left|\frac{y}{x}\right| < k$$

where α = principal value of $\tan^{-1}\left(\frac{k-k_o}{1+k\ k_o}\right)$.

Then

$$\left|Q(k)-Q(k_o)\right| \leqslant \left|\int_{x^2+y^2<R^2} h(x)\,dx\right| + \left|\int_{x^2+y^2>R^2} h(x)\,dx\right| < 2MR^2\alpha+\epsilon/2 \tag{10}$$

Now we can choose $\delta < \tan\frac{\epsilon}{4MR^2}$ so that for k in $|k-k_o| < \delta$,

$$\frac{k-k_o}{1+k\ k_o} < |k-k_o| < \delta < \tan\frac{\epsilon}{4MR^2} \tag{11}$$

and since $\tan x$ is a monotone increasing function of x in the neighbourhood of zero, we have

$$\tan^{-1}\left(\frac{k-k_o}{1+k\ k_o}\right) < \frac{\epsilon}{4MR^2} \tag{12}$$

which gives us

$$\left|Q(k)-Q(k_o)\right| < \epsilon \qquad \text{for } |k-k_o| < \delta.$$

A similar proof follows if we consider $k < k_o$.

Now consider two different values of k, say k_1 and k_2 ($k_1 < k_2$).

$$Q(k_1) - Q(k_2) = \int_{-\infty}^{\infty}\int_{-|y/k_1|}^{|y/k_1|} f(x,y)\,dx\,dy - \int_{-\infty}^{\infty}\int_{-|y/k_2|}^{|y/k_2|} f(x,y)\,dx\,dy \tag{13}$$

$$= \int_{-\infty}^{\infty}\left[\int_{-|y/k_1|}^{-|y/k_2|} f(x,y)\,dx + \int_{|y/k_2|}^{|y/k_1|} f(x,y)\,dx\right] dy\ . \tag{14}$$

We have seen that $Q(k_1) - Q(k_2) \to 0$ as $k_2 \to k_1$, and from (14) it follows that $Q(k_1) \geqslant Q(k_2)$ as long as $k_1 < k_2$.

We note, from (14) that $Q(k_1) = Q(k_2)$ for $k_1 \neq k_2$ if and only if $f(x,y) = 0$ for

$$|y/k_2| < |x| < |y/k_1| \tag{15}$$

But from condition (iii) we see that $f(x,y)$ cannot equal zero for all x and y satisfying (14). Hence $Q(k_1) > Q(k_2)$ for $k_1 < k_2$.

This completes our proof.

We note that under the conditions of our lemma, we conclude that there exists a unique value of k, say K, such that $Q(K) = 1/2$.

<u>Theorem 1</u>.

If X and Y are two random variates having variances σ_1^2 and σ_2^2 respectively, and the conditions of the preceding lemma are satisfied, then:

(i) K is determined uniquely.

(ii) If $K < \sigma_2/\sigma_1$, minimum variance is equivalent to closeness for kX and Y, for all $k > 0$, except in the interval $K \leqslant k < \sigma_2/\sigma_1$.

(iii) If $K > \sigma_2/\sigma_1$, minimum variance is equivalent to closeness for kX and Y, for all $k > 0$, except in the interval $\sigma_2/\sigma_1 < k \leqslant K$.

(iv) $K = \sigma_2/\sigma_1$ is a necessary and sufficient condition for the equivalence of closeness and minimum variance for kX and Y for all $k > 0$.

<u>Proof</u>:

Since X and Y have variances σ_1^2 and σ_2^2 respectively, $Z = kX$ will have variance $k^2\sigma_1^2$. It is convenient to separate our proof into two distinct cases: (i) when $K < \sigma_2/\sigma_1$, and (ii) when $K > \sigma_2/\sigma_1$.

Case I

(a) If $0 < k < K$, then $Q(k) > 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(b) If $K \leqslant k < \sigma_2/\sigma_1$, then $Q(k) \leqslant 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(c) If $k > \sigma_2/\sigma_1$, then $Q(k) < 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

Case II

(a) If $0 < k < \sigma_2/\sigma_1$, then $Q(k) > 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(b) If $\sigma_2/\sigma_1 < k \leqslant K$, then $Q(k) \geqslant 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

(c) If $k > K$, then $Q(k) < 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

The preceding conditions make it clear that kX or Y is closer to the true parameter value zero according to which has the smaller variance except in the intervals $K \leqslant k < \sigma_2/\sigma_1$ (Case I) and $\sigma_2/\sigma_1 < k \leqslant K$ (Case II). In the first of these, kX has the smaller variance but Y is closer while in the latter, Y has the smaller variance but kX is closer. It is evident, then, that $K = \sigma_2/\sigma_1$ is a necessary and sufficient condition for the equivalence of closeness and minimum variance for kX and Y for all $k > 0$. This completes our proof.

Let us now examine the standardised variates $S = X/\sigma_1$ and $T = Y/\sigma_2$. Obviously $E(S) = E(T) = 0$ and further, S and T each have variance one. Consider

$$P\left(|X| < |Y|\right) = P\left(|\sigma_1 S| < |\sigma_2 T|\right) = P\left(|\sigma_1/\sigma_2 S| < |T|\right)$$

$$= Q\left(\sigma_1/\sigma_2\right) \tag{16}$$

where $Q(k)$ is now defined by

$$Q(k) = P\left(|kS| < |T|\right) \qquad \text{for } k > 0. \tag{17}$$

As before, if K is the value of k such that $Q(K) = 1/2$ then we may consider the following:

<u>Case I</u> $\qquad K < \sigma_T/\sigma_S = 1$

(a) If $0 < k < K$, then $Q(k) > 1/2$ and $k^2 < 1$

(b) If $K \leqslant k < 1$, then $Q(k) \leqslant 1/2$ and $k^2 < 1$

(c) If $k > 1$, then $Q(k) < 1/2$ and $k^2 > 1$

<u>Case II</u> $\qquad K > \sigma_T/\sigma_S = 1$

(a) If $0 < k < 1$, then $Q(k) > 1/2$ and $k^2 < 1$

(b) If $1 < k \leqslant K$, then $Q(k) \geqslant 1/2$ and $k^2 > 1$

(c) If $k > K$, then $Q(k) < 1/2$ and $k^2 > 1$

Therefore we see that kS or T is closer to the true parameter value zero according to which has the smaller variance, except in the intervals $K \leqslant k < 1$ (Case I) and $1 < k \leqslant K$ (Case II), and hence it is obvious that $K = 1$ is a necessary and sufficient condition for the equivalence of closeness and minimum variance for kS and T for all $k > 0$.

This says that a necessary and sufficient condition for smaller and closeness to be equivalent for kS and T is

$$Q(1) = P\,(|S| < |T|) = 1/2 \tag{18}$$

Hence from (16) we have obtained a necessary and sufficient condition for the equivalence of smaller variance and closeness for X and Y which, in standardized units, are jointly distributed like S and T. We have, therefore, proved the following theorem:

<u>Theorem 2</u>.

If X and Y are two random variates having variances σ_1^2 and σ_2^2 respectively and if the conditions of our lemma are satisfied and, further, if $S = X/\sigma_1$ and $T = Y/\sigma_2$, then a necessary and sufficient condition for closeness and minimum variance to be equivalent for X and Y is that $Q(1) = P\,(|S| < |T|) = 1/2$.

B. Discrete Distribution Functions

A random variate X is said to be of the discrete type, or to possess a discrete distribution, if the total mass of the distribution is concentrated in discrete mass points and if, moreover, any finite interval contains at most a finite number of the mass points.[a] The set of all mass points is finite or denumerable. The distribution of X is completely described by stating that, for every i, we have the probability:

$$P\,(X = x_i) = p_i \qquad ;\ i=1,2,\ldots \tag{1}$$

where

$$\sum_i p_i = 1. \tag{2}$$

The cumulative distribution function is then given by

$$F(x) = P\,(X \leqslant x) = \sum_{x_i \leqslant x} p_i \tag{3}$$

We shall extend the results of the preceding section, taking note of certain modifications made necessary by the discrete nature of the distribution. Consider random variables X and Y, where $E(X) = E(Y) = 0$ and

a. Cramér, H., <u>Mathematical Methods of Statistics</u>, Princeton, Princeton University Press, 1946, p. 168.

X and Y have variances σ_1^2 and σ_2^2 respectively. Let $F(x,y)$ be defined by the equation:

$$F(x,y) = P \quad (X \leqslant x,\ Y \leqslant y) \tag{4}$$

Hence

$$Q = P\ (|X| < |Y|) = \iint_{|x| < |y|} dF(x,y) \tag{5}$$

where the integral is taken to be a Lebesgue-Stieltjes integral.

The methods used in obtaining results for continuous distributions indicate the line of procedure to be followed here. As before, define

$$Q(k) = P\ (|kX| < |Y|) \qquad ,\ k > 0. \tag{6}$$

Consider two values of k, say k_1 and k_2 ($k_1 < k_2$). We see that

$$Q(k_1) = P\ (|k_1X| < |Y|) = P\ (|k_2X| < |Y|) + P\ (|k_1X| < |Y| \leqslant |k_2X|)$$

$$= Q(k_2) + P\ (|k_1X| < |Y| \leqslant |k_2X|) \tag{7}$$

Hence $Q(k)$ is a monotone decreasing function of k. However, since $Q(k)$ may be discontinuous (if we talk in terms of general distribution functions), it is not strictly monotone decreasing with increasing k and therefore it does not follow that there exists a unique value of k, say K, such that $Q(K) = 1/2$. Such a value of k may exist or there may be some interval in which $Q(k) = 1/2$. Whichever of these two situations holds, there exist two points L and R defined as follows:

L = upper bound of k such that $Q(k) > 1/2$

R = lower bound of k such that $Q(k) < 1/2$.

Let us consider the following:

<u>Case I</u> $L \leqslant \sigma_2/\sigma_1 \leqslant R \quad (k \neq \sigma_2/\sigma_1)$

(a) If $0 < k < L$, then $Q(k) > 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(b) If $L \leqslant k < \sigma_2/\sigma_1$, then $Q(k) \geqslant 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(c) If $\sigma_2/\sigma_1 < k \leqslant R$, then $Q(k) \leqslant 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

(d) If $k > R$, then $Q(k) < 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

<u>Case II</u> $\sigma_2/\sigma_1 \leqslant L \leqslant R \quad (k \neq \sigma_2/\sigma_1)$

(a) If $0 < k < \sigma_2/\sigma_1$, then $Q(k) > 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(b) If $\sigma_2/\sigma_1 < k \leqslant L$, then $Q(k) \geqslant 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

(c) If $L < k \leqslant R$, then $Q(k) \leqslant 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

(d) If $k > R$, then $Q(k) < 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

<u>Case III</u> $L \leqslant R \leqslant \sigma_2/\sigma_1 \quad (k \neq \sigma_2/\sigma_1)$

(a) If $0 < k < L$, then $Q(k) > 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(b) If $L \leqslant k < R$, then $Q(k) \geqslant 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(c) If $R \leqslant k < \sigma_2/\sigma_1$, then $Q(k) \leqslant 1/2$ and $k^2\sigma_1^2 < \sigma_2^2$

(d) If $k > \sigma_2/\sigma_1$, then $Q(k) < 1/2$ and $k^2\sigma_1^2 > \sigma_2^2$

If $k = \sigma_2/\sigma_1$, then kX and Y each have variance σ_2^2 and we should choose the one which is the closer estimate. No decision is possible unless $L = \sigma_2/\sigma_1 = R$.

Closeness and smaller variance are seen to be equivalent for kX and Y, for all $k > 0$, except in the closed interval bounded by the smallest and largest of the three quantities L, R and σ_2/σ_1. If $L = R = \sigma_2/\sigma_1$, minimum variance and closeness will be equivalent for kX and Y for all $k > 0$.

Suppose $L \neq R$. Then, if $L \leqslant \sigma_2/\sigma_1 \leqslant R$ and $k^2\sigma_1^2 < \sigma_2^2$, we find that $Q(k) \geqslant 1/2$ instead of the more desirable $Q(k) > 1/2$. We shall say that kX is almost closer to the true parameter value zero than is Y whenever $Q(k) \geqslant 1/2$. We will refer to this as the property of <u>quasi-closeness</u>.

Now, if we have

$$L \leqslant \sigma_2/\sigma_1 \leqslant R \tag{8}$$

then it is evident that (8) is a necessary and sufficient condition for the equivalence of smaller variance and quasi-closeness for kX and Y for all $k > 0$. The 'sufficiency' is apparent. To show the 'necessity', we need only to consider the following: if $\sigma_2/\sigma_1 < L$, there will be, by the definition of L, a value of k such that $\sigma_2/\sigma_1 < k < L$ for which kX would be almost closer[a] but Y would have the smaller variance. A similar result follows when $\sigma_2/\sigma_1 > R$. We have thus proven the following theorem:

<u>Theorem 3</u>.

If X and Y are two random variates with zero means and variances σ_1^2 and σ_2^2 respectively, and if we set $Q(k) = P(|kX| < |Y|)$, $k > 0$, then:

(i) kX or Y will be closer to the true parameter value zero according to which possesses the smaller variance except for values of k in the closed interval bounded by the smallest and largest of the three quantities L, R and σ_2/σ_1, where

L = upper bound of k such that $Q(k) > 1/2$

R = lower bound of k such that $Q(k) < 1/2$.

---

a. A quantity is certainly almost closer if it is closer; i.e., $P(|kX| < |Y|) > 1/2$ implies $P(|kX| < |Y|) \geqslant 1/2$.

(ii) $L = R = \sigma_2/\sigma_1$ is a necessary and sufficient condition for the equivalence of closeness and minimum variance for kX and Y for all $k > 0$.

(iii) kX or Y will be almost closer to zero according to which possesses the smaller variance, for all $k > 0$, if and only if $L \leqslant \sigma_2/\sigma_1 \leqslant R$.

Consider now the standardized variates $S = X/\sigma_1$ and $T = Y/\sigma_2$, where $E(S) = E(T) = 0$ and $\text{var}(S) = \text{var}(T) = 1$. We shall define

$$Q(k) = P(|kS| < |T|) \tag{9}$$

and then the following theorem may be shown to hold:

<u>Theorem 4</u>.

If X and Y are two random variates with $E(X) = E(Y) = 0$ and variances σ_1^2 and σ_2^2 respectively, if $S = X/\sigma_1$ and $T = Y/\sigma_2$ and $Q(k) = P(|kS| < |T|)$, then:

(i) $P(|S| < |T|) = P(|T| < |S|)$ is a sufficient condition for minimum variance to imply quasi-closeness for X and Y for all values of σ_1 and σ_2 if $P(|S| = |T|) = 0$.

(ii) If $P(|S| = |T|) = P > 0$, then $P(|S| < |T|) = P(|T| < |S|)$ is a sufficient condition for the equivalence of minimum variance and closeness for X and Y for all σ_1 and σ_2.

<u>Proof</u>:

<u>Case I</u> $\qquad L \leqslant 1 \leqslant R \qquad (k \neq 1)$

(a) If $0 < k < L$, then $Q(k) > 1/2$ and $k^2 < 1$

(b) If $L \leqslant k < 1$, then $Q(k) \geqslant 1/2$ and $k^2 < 1$

(c) If $1 < k \leqslant R$, then $Q(k) \leqslant 1/2$ and $k^2 > 1$

(d) If $k > R$, then $Q(k) < 1/2$ and $k^2 > 1$

<u>Case II</u> $1 \leqslant L \leqslant R \qquad (k \neq 1)$

(a) If $0 < k < 1$, then $Q(k) > 1/2$ and $k^2 < 1$

(b) If $1 < k \leqslant L$, then $Q(k) \geqslant 1/2$ and $k^2 > 1$

(c) If $L < k \leqslant R$, then $Q(k) \leqslant 1/2$ and $k^2 > 1$

(d) If $k > R$, then $Q(k) < 1/2$ and $k^2 > 1$

<u>Case III</u> $L \leqslant R \leqslant 1 \qquad (k \neq 1)$

(a) If $0 < k < L$, then $Q(k) > 1/2$ and $k^2 < 1$

(b) If $L \leqslant k < R$, then $Q(k) \geqslant 1/2$ and $k^2 < 1$

(c) If $R \leqslant k < 1$, then $Q(k) \leqslant 1/2$ and $k^2 < 1$

(d) If $k > 1$, then $Q(k) < 1/2$ and $k^2 > 1$

If $k = 1$, then kS and T each have the same variance and we should choose the one which is the closer estimate. No decision is possible unless $L = R = 1$.

Consider

$$P(|S| < |T|) + P(|S| = |T|) + P(|T| < |S|) = 1. \tag{10}$$

If

$$P(|S| = |T|) = 0 \tag{11}$$

and

$$P(|S| < |T|) = P(|T| < |S|) \tag{12}$$

then we must have $Q(1) = 1/2$, but $Q(1) = 1/2$ implies $L \leqslant 1 \leqslant R$. By Theorem 3 this implies the equivalence of quasi-closeness and minimum variance for kS and T for all $k > 0$ and hence in particular for $k = \sigma_1/\sigma_2$. This says that X or Y will be almost closer to zero according to which has the smaller variance. This shows the sufficiency of (12).

Now, suppose we have

$$P\,(|S| = |T|) = P > 0. \tag{13}$$

Then

$$P\,(|S| < |T|) + P\,(|T| < |S|) = 1-P$$

and hence (12) gives us

$$Q(1) = (1-P)/2 < 1/2 \tag{15}$$

and therefore, by definition, $R \leqslant 1$. Next, consider $Q(1-\delta)$, $0 < \delta < 1$:

$$\begin{aligned} Q(1-\delta) &= P\,(|(1-\delta)S| < |T|) \\ &\geqslant P\,(|S| < |T|) + P\,(|S| = |T|) \\ &= (1-P)/2 + P \\ &= (1+P)/2 \end{aligned}$$

Hence $Q(1-\delta) > 1/2$ and therefore $1-\delta \leqslant L$. This implies $1 \leqslant L$. To show this, assume $L < 1$. Then $L = 1-\epsilon$, where $\epsilon > 0$, and therefore $L < 1-\epsilon + \epsilon/2 = 1-\epsilon/2$. Calling $\epsilon/2 = \delta$ we have $L < 1-\delta$ which contradicts $1-\delta \leqslant L$. Hence we have $1 \leqslant L$. We now have that $1 \leqslant L \leqslant R \leqslant 1$ and therefore $L = R = 1$. By Theorem 3, this proves the 'sufficiency'.

C. Form of the Distribution

Returning to the case where X and Y admit a continuous probability density function $f(x,y)$, let us find conditions on the form of the distribution such that the criteria of closeness and smaller variance are equivalent.

Consider the standardized variates S and T. They will have a joint probability density function, say $g(s,t)$.

Now, Theorem 2 states that under certain general conditions $P(|S| < |T|) = 1/2$ is a necessary and sufficient condition for closeness and minimum variance to be equivalent for all X and Y. This may be expressed as follows:

$$\int_{-\infty}^{\infty}\int_{-|t|}^{|t|} g(s,t)\,ds\,dt = 1/2 = \int_{-\infty}^{\infty}\int_{-|s|}^{|s|} g(s,t)\,dt\,ds \tag{1}$$

$$= \int_{-\infty}^{\infty}\int_{-|t|}^{|t|} g(t,s)\,ds\,dt\ . \tag{2}$$

Therefore

$$\int_{-\infty}^{\infty}\int_{-|t|}^{|t|}\Big[g(s,t) - g(t,s)\Big]\,ds\,dt = 0 \tag{3}$$

is a necessary and sufficient condition for the equivalence of closeness and minimum variance for all X and Y. We have, then, the following theorem:

<u>Theorem 5</u>.

Given two random variates, X and Y, which satisfy the conditions of Theorem 1, consider them in standardized units $S = X/\sigma_1$ and $T = Y/\sigma_2$. Let $g(s,t)$ be the joint probability density function of S and T. Then a necessary and sufficient condition for the equivalence of closeness and minimum variance for X and Y for all values of σ_1 and σ_2 is

$$\int_{-\infty}^{\infty}\int_{-|t|}^{|t|}\Big[g(s,t)-g(t,s)\Big]\,ds\,dt = 0.$$

We note that if $g(s,t)$ is a symmetric function in s and t, then closeness and minimum variance are equivalent for X and Y. The result obtained by Geary[a] is seen to be a special case of the above.

a. Geary, R. C., Comparison of the Concepts of Efficiency and Closeness for Consistent Estimates of a Parameter, Biometrika, 33:123-128, 1944.

IV. ASYMPTOTIC CLOSENESS

A. Continuous Distribution Functions

Efficiency, as defined by Fisher, is an asymptotic property while Pitman's concept of closeness is not. It seems necessary, therefore, to formulate a definition of "asymptotic closeness" and then compare it with the notions of asymptotic smaller variance and asymptotic efficiency.

__Definition.__

Let t_{1n} and t_{2n} be two estimators of the parameter θ, each calculated from a sample of size n and $Q_n = P\left(|t_{1n}-\theta| < |t_{2n}-\theta|\right)$. We will say that t_{1n} is an __asymptotically closer__ estimator of θ than is t_{2n} if

$$\lim_{n \to \infty} Q_n = \lim_{n \to \infty} P\left(|t_{1n}-\theta| < |t_{2n}-\theta|\right)$$

exists and is greater than one-half.

As in the preceding sections, we may consider $X_n = t_{1n}-\theta$ and $Y_n = t_{2n}-\theta$. We may easily prove the following rather obvious fact, that is, that a consistent estimator is asymptotically closer than any non-consistent estimator.

Consider X_n and Y_n, which converge in probability to zero and $\theta^* > 0$ respectively. We may state this as follows:

(i) X_n converges in probability to zero if, for any ϵ and δ, $\epsilon > 0$ and $\delta > 0$, there exists an N such that for all $n > N$,

$$P\left(|X_n| < \epsilon\right) > 1-\delta \tag{1}$$

(ii) Y_n converges in probability to $\theta^* > 0$ if, for any ϵ and δ, $\epsilon > 0$ and $\delta > 0$, there exists an N^* such that for all $n > N^*$,

$$P\left(\left|Y_n-\theta^*\right| < \epsilon\right) > 1-\delta \tag{2}$$

We note that

$$1-Q_n = P\left(\left|X_n\right| \geqslant \left|Y_n\right|\right) \tag{3}$$

and therefore

$$\begin{aligned} 1-Q_n &\leqslant P\left(\left|Y_n\right| \leqslant \left|X_n\right| < \epsilon\right) + P\left(\left|X_n\right| \geqslant \epsilon\right) \\ &\leqslant P\left(\left|X_n\right| \geqslant \epsilon\right) + P\left(\left|Y_n\right| < \epsilon\right) \end{aligned} \tag{4}$$

Now, equation (1) tells us that

$$P\left(\left|X_n\right| \geqslant \epsilon\right) = 1-P\left(\left|X_n\right| < \epsilon\right) < \delta \tag{5}$$

for $n > N$.

For any $\epsilon < \frac{\theta^*}{2}$, there cannot exist a value of Y_n which satisfies both $\left|Y_n\right| < \epsilon$ and $\left|Y_n-\theta^*\right| < \epsilon$. Therefore

$$P\left(\left|Y_n\right| < \epsilon\right) + P\left(\left|Y_n-\theta^*\right| < \epsilon\right) \leqslant 1 \tag{6}$$

and hence

$$P\left(\left|Y_n\right| < \epsilon\right) \leqslant 1-P\left(\left|Y_n-\theta^*\right| < \epsilon\right) < \delta \tag{7}$$

for $n > N^*$.

Using the results of equations (5) and (7), we see that (4) gives us $1-Q_n < 2\delta$ for $n > \max(N, N^*)$, and we may then conclude that $\lim_{n \to \infty} Q_n = 1$. A similar result may be proved for $\theta^* < 0$.

We have thus proven the following theorem:

Theorem 6.

If X_n converges in probability to zero, and Y_n converges in probability to $\theta^* \neq 0$, then $\lim_{n \to \infty} Q_n = 1$, that is, X_n is an asymptotically closer estimate of zero than is Y_n.

Consider now the variates $S_n = \dfrac{t_{1n}-\theta}{\sigma_{1n}}$ and $T_n = \dfrac{t_{2n}-\theta}{\sigma_{2n}}$ where t_{1n} and t_{2n} have variances σ_{1n}^2 and σ_{2n}^2 respectively. Let us assume that S_n and T_n have a joint probability density function $g_n(s,t)$. We may further assume that this sequence of density functions converges uniformly to a joint probability density function $g(s,t)$ with zero means and unit variances.

Define.

$$Q_n(k) = P\left(|kS_n| < |T_n|\right) \qquad ;k > 0. \tag{8}$$

Then

$$Q_n(k) = \int_{-\infty}^{\infty}\int_{-|t/k|}^{|t/k|} g_n(s,t)\,ds\,dt \tag{9}$$

$$= \int_{-\infty}^{\infty}\int_{-|t/k|}^{|t/k|} g(s,t)\,ds\,dt + \int_{-\infty}^{\infty}\int_{-|t/k|}^{|t/k|} \left[g_n(s,t)-g(s,t)\right] ds\,dt \tag{10}$$

where $\displaystyle\int_{-\infty}^{\infty}\int_{-|t/k|}^{|t/k|} \left[g_n(s,t)-g(s,t)\right] ds\,dt = I_n(k)$, say.

We wish to show that $I_n(k)$ converges uniformly to zero, that is, that $Q_n(k)$ converges uniformly to $Q(k)$ where $Q(k)$ is defined by the equation:

$$Q(k) = \int_{-\infty}^{\infty}\int_{-|t/k|}^{|t/k|} g(s,t)\,ds\,dt. \tag{11}$$

We note that $0 \leqslant Q_n(k) \leqslant 1$ for every n, and that as $k \to 0$, $Q_n(k) \to 1$ and that as $k \to \infty$, $Q_n(k) \to 0$. Also, $Q(k)$ satisfies these same conditions. Each $Q_n(k)$; $n=1, 2, 3, \ldots$; and $Q(k)$ are continuous functions of k.

Consider

$$I_n(k) = \int_{-\infty}^{\infty} \int_{-|t/k|}^{|t/k|} \left[g_n(s,t) - g(s,t) \right] ds\, dt. \tag{12}$$

We may construct a square S_R (of side 2R) centered at the origin such that:

$$\text{(i)} \quad \int\int_{S_R} g(s,t)\, ds\, dt < \frac{\epsilon}{4} \tag{13}$$

$$\text{(ii)} \quad \int\int_{S_R} g_n(s,t)\, ds\, dt < \frac{\epsilon}{4} \text{ , for every n.} \tag{14}$$

Then

$$I_n(k) \leqslant \int_{-\infty}^{\infty} \int_{-|t/k|}^{|t/k|} \left| g_n(s,t) - g(s,t) \right| ds\, dt \tag{15}$$

$$< \int_{-R}^{R} \int_{-R}^{R} \left| g_n(s,t) - g(s,t) \right| ds\, dt + \frac{\epsilon}{4} + \frac{\epsilon}{4} \text{ .} \tag{16}$$

Since $g_n(s,t)$ converges uniformly to $g(s,t)$ as $n \to \infty$, we may say that, for $n > N$,

$$\left| g_n(s,t) - g(s,t) \right| < \frac{\epsilon}{8R^2} \tag{17}$$

and hence

$$I_n(k) < \frac{\epsilon}{8R^2} \int_{-R}^{R} \int_{-R}^{R} ds\ dt + \frac{\epsilon}{2} \tag{18}$$

$$< \frac{\epsilon}{8R^2} (4R^2) + \frac{\epsilon}{2} = \epsilon \qquad \text{for } n > N. \tag{19}$$

Therefore $I_n(k)$ converges uniformly to zero and hence $Q_n(k)$ converges uniformly to $Q(k)$.

Let us assume that $\lim_{n \to \infty} \sigma_{1n}/\sigma_{2n} = \sigma < 1$. Then $|\sigma_{1n}/\sigma_{2n} - \sigma_{1m}/\sigma_{2m}| < \epsilon$ for $m, n > N$ and hence

$$\sigma_{1m}/\sigma_{2m} - \epsilon < \sigma_{1n}/\sigma_{2n} < \sigma_{1m}/\sigma_{2m} + \epsilon. \tag{20}$$

Therefore

$$P((\sigma_{1m}/\sigma_{2m} + \epsilon)\ |S_n| < |T_n|) < P((\sigma_{1n}/\sigma_{2n})\ |S_n| < |T_n|)$$

$$< P((\sigma_{1m}/\sigma_{2m} - \epsilon)\ |S_n| < |T_n|) \tag{21}$$

for $m, n > N$.

Keeping m fixed and letting $n \to \infty$, we have

$$P((\sigma_{1m}/\sigma_{2m} + \epsilon)|S| < |T|) < L < P((\sigma_{1m}/\sigma_{2m} - \epsilon)|S| < |T|) \tag{22}$$

where $L = \lim_{n \to \infty} P((\sigma_{1n}/\sigma_{2n})|S_n| < |T_n|)$.

Now, letting $m \to \infty$, we obtain

$$P((\sigma + \epsilon)\ |S| < |T|) < L < P((\sigma - \epsilon)\ |S| < |T|) \tag{23}$$

where $\epsilon > 0$ is any arbitrary number.

The result exhibited in equation (23) is obtained as follows: Consider, for example, the left-hand side of the expression:

$$\lim_{m \to \infty} P((\sigma_{1m}/\sigma_{2m} + \epsilon)\,|S| < |T|) = P((\sigma + \epsilon)\,|S| < |T|). \tag{24}$$

or, more explicitly

$$\lim_{m \to \infty} \int_{-\infty}^{\infty} \int_{\frac{-|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}}^{\frac{|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}} g(s,t)\, ds\, dt = \int_{-\infty}^{\infty} \int_{\frac{-|t|}{(\sigma + \epsilon)}}^{\frac{|t|}{(\sigma + \epsilon)}} g(s,t)\, ds\, dt. \tag{25}$$

Since $g(s,t)$ is assumed to be a continuous probability density function, then

$$\int_{\frac{-|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}}^{\frac{|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}} g(s,t)\, ds = h(t,m)\ \text{,say,} \tag{26}$$

is an integrable function for any t, and further, since

$$h(t,m) \leqslant \int_{-\infty}^{\infty} g(s,t)\, ds = g_2(t) \tag{27}$$

where $g_2(t)$ is the marginal probability density function of t (and is hence both bounded and integrable), we may conclude that $h(t,m)$ as defined by equation (26) is both bounded and integrable. Thus we may write:

$$\lim_{m \to \infty} \int_{-\infty}^{\infty} \int_{\frac{-|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}}^{\frac{|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}} g(s,t)\, ds\, dt = \int_{-\infty}^{\infty} \lim_{m \to \infty} \int_{\frac{-|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}}^{\frac{|t|}{(\sigma_{1m}/\sigma_{2m} + \epsilon)}} g(s,t)\, ds\, dt$$

$$= \int_{-\infty}^{\infty} \int_{\frac{-|t|}{(\sigma + \epsilon)}}^{\frac{|t|}{(\sigma + \epsilon)}} g(s,t)\, ds\, dt. \tag{28}$$

Now, since $\epsilon > 0$ is any arbitrary number, we have the conclusion:

$$L = \lim_{n \to \infty} P(\sigma_{1n}/\sigma_{2n} |S_n| < |T_n|) = \lim_{n \to \infty} Q_n(\sigma_{1n}/\sigma_{2n})$$

$$= Q(\sigma). \tag{29}$$

Recalling that we assumed $\sigma < 1$, and since $Q(k)$ is a continuous, strictly monotone decreasing function of k, if we further assume that:

$$\int_{-\infty}^{\infty} \int_{-|t|}^{|t|} \left[g(s,t) - g(t,s) \right] ds\, dt = 0, \tag{30}$$

we have $Q(1) = 1/2$ and hence

$$\lim_{n \to \infty} Q_n(\sigma_{1n}/\sigma_{2n}) = Q(\sigma) > 1/2,\ \sigma < 1. \tag{31}$$

That is,

$$\lim_{n \to \infty} P(|t_{1n} - \theta| < |t_{2n} - \theta|) > 1/2 \tag{32}$$

and hence the variate with the asymptotically smaller variance is asymptotically closer.

Suppose, now, that (32) holds. Then

$$\lim_{n \to \infty} Q_n(\sigma_{1n}/\sigma_{2n}) = P > 1/2 \text{ , that is,}$$

$$\lim_{n \to \infty} \left[Q_n(\sigma_{1n}/\sigma_{2n}) - P\right] = 0, \text{ where } P > 1/2.$$

Let $P = Q(\sigma)$, $\sigma < 1$. Now, for sufficiently large n, $\left| Q_n(\sigma_{1n}/\sigma_{2n}) - Q(\sigma_{1n}/\sigma_{2n}) \right| < \epsilon$ because of the uniform convergence of $Q_n(k)$ to $Q(k)$, and in view of the following equation:

$$Q_n(\sigma_{1n}/\sigma_{2n}) - Q(\sigma) = Q_n(\sigma_{1n}/\sigma_{2n}) - Q(\sigma_{1n}/\sigma_{2n}) + Q(\sigma_{1n}/\sigma_{2n}) - Q(\sigma) \qquad (33)$$

we may conclude that

$$\lim_{n \to \infty} \left[Q(\sigma_{1n}/\sigma_{2n}) - Q(\sigma)\right] = 0. \qquad (34)$$

Since $Q(k)$ is a continuous, strictly monotone decreasing function of k, it follows from (34) that

$$\lim_{n \to \infty} (\sigma_{1n}/\sigma_{2n}) = \sigma < 1. \qquad (35)$$

The above results give us the following theorem:

Theorem 7.

Given two random variates, t_{1n} and t_{2n}, calculated from a sample of size n, with variances σ_{1n}^2 and σ_{2n}^2 respectively, such that:

(i) $S_n = \dfrac{t_{1n}-\theta}{\sigma_{1n}}$ and $T_n = \dfrac{t_{2n}-\theta}{\sigma_{2n}}$ have a joint probability density function $g_n(s,t)$ which converges uniformly, as $n \to \infty$, to a joint probability density function $g(s,t)$ with zero means and unit variances;

(ii) $g(s,t) \neq 0$ at the origin;

(iii) $\displaystyle\int_{-\infty}^{\infty} \int_{-|t|}^{|t|} \left[g(s,t) - g(t,s)\right] ds\, dt = 0;$

then asymptotic closeness and asymptotic smaller variance are equivalent if and only if $\lim_{n \to \infty} \sigma_{1n}/\sigma_{2n}$ exists.

The following theorem on the asymptotically closest estimator follows immediately:

<u>Theorem 8</u>.

If t_n is the asymptotically closest estimator of θ and $\sqrt{n}(t_n-\theta)$ is asymptotically normally distributed with zero mean, then t_n is an asymptotically efficient estimator.

B. Discrete Distribution Functions

We shall consider, in this section, random variates t_{1n} and t_{2n} which admit a discrete distribution function. As before, we will examine the quantity Q_n defined by:

$$Q_n = P(|t_{1n}-\theta| < |t_{2n}-\theta|). \tag{1}$$

<u>Definition</u>

t_{1n} is said to be an <u>asymptotically</u> <u>almost</u> <u>closer</u> estimator of θ than is t_{2n} if

$$\lim_{n \to \infty} Q_n = \lim_{n \to \infty} P(|t_{1n}-\theta| < |t_{2n}-\theta|)$$

exists and is greater than or equal to one-half and for at least one value of θ the inequality holds good.

Consider the variates $S_n = \dfrac{t_{1n}-\theta}{\sigma_{1n}}$ and $T_n = \dfrac{t_{2n}-\theta}{\sigma_{2n}}$, where σ_{1n}^2 and σ_{2n}^2 are the variances of t_{1n} and t_{2n} respectively. Defining

$$Q_n(k) = P(|kS_n| < |T_n|) \quad ; \quad k > 0, \tag{2}$$

and assuming that the sequence $G_n(s,t)$, where $G_n(s,t) = P(S_n \leqslant s,\ T_n \leqslant t)$, converges uniformly to the joint cumulative distribution function $G(s,t)$ in all points of continuity of $G(s,t)$, s and t having zero means and unit variances, we see that $Q_n(k)$ approaches a limiting value $Q(k)$ as $n \to \infty$. This may be shown as follows:

$$Q_n(k) = \iint\limits_{|ks| < |t|} d\, G_n(s,t) \quad ; \; k > 0 \tag{3}$$

and the integral is taken to be a Lebesgue-Stieltjes integral. Therefore

$$Q_n(k) = \iint\limits_{|ks| < |t|} d\, G(s,t) + \iint\limits_{|ks| < |t|} d\left[G_n(s,t) - G(s,t)\right] \tag{4}$$

$$= Q(k) + \iint\limits_{|ks| < |t|} d\left[G_n(s,t) - G(s,t)\right] \tag{5}$$

where $Q(k)$ is defined by the equation

$$Q(k) = \iint\limits_{|ks| < |t|} d\, G(s,t). \tag{6}$$

Now, there exists a square S_R (of side 2R), centered at the origin, such that:

$$\text{(i)} \quad \iint\limits_{\overline{S_R}} d\, G(s,t) < \frac{\epsilon}{4} \tag{7}$$

$$\text{(ii)} \int\int_{\bar{S}_R} d\ G_n(s,t) < \frac{\epsilon}{4}\quad , \text{ for every } n. \tag{8}$$

Therefore

$$\int\int_{|ks| < |t|} d\left[G_n(s,t)-G(s,t)\right] \leqslant \int\int_{S_R} d\left[G_n(s,t)-G(s,t)\right] + \frac{\epsilon}{2}\ . \tag{9}$$

Then, if in S_R we choose N so large that for $n > N$, $\left|G_n(s,t)-G(s,t)\right| < \dfrac{\epsilon}{16R^2}$

and hence $\left|d\left[G_n(s,t)-G(s,t)\right]\right| < 2\left(\dfrac{\epsilon}{16R^2}\right) = \dfrac{\epsilon}{8R^2}$.

we will have:

$$\left|\int\int_{|ks| < |t|} d\left[G_n(s,t)-G(s,t)\right]\right| < \frac{\epsilon}{8R^2}\ (4R^2) + \frac{\epsilon}{2} = \epsilon,\quad n > N. \tag{10}$$

Hence we conclude that $Q_n(k)$ converges uniformly to $Q(k)$ as defined by (6).

Now, if we assume $G(s,t)$ to be a continuous, joint cumulative distribution function and let $g(s,t)$, $g(0,0) \neq 0$, be the continuous, joint probability density function associated with $G(s,t)$, and further assume that

$$\int\int_{|s| < |t|} d\left[G(s,t)-G(t,s)\right] = \int\int_{|s| < |t|} \left[g(s,t)-g(t,s)\right] ds\ dt = 0, \tag{11}$$

then $Q(1) = 1/2$ by virtue of Theorem 5. Hence, if $\lim_{n\to\infty} Q_n(\sigma_{1n}/\sigma_{2n}) = Q(\sigma)$ where $\lim_{n\to\infty} (\sigma_{1n}/\sigma_{2n}) = \sigma < 1$, we have $\lim_{n\to\infty} P(\left|t_{1n}-\theta\right| < \left|t_{2n}-\theta\right|) > 1/2$

and thus the variate with the asymptotically smaller variance is asymptotically closer.

However, if we do not assume $G(s,t)$ to be a continuous, joint cumulative distribution function but allow it to be a step-function, our conclusions will be changed slightly. We may again show that $Q_n(k)$ converges uniformly to a limit function $Q(k)$. Consider

$$P(|t_{1n}-\theta| < |t_{2n}-\theta|) = Q_n(\sigma_{1n}/\sigma_{2n}) \tag{12}$$

and assume that

$$\int_{|s|} \int_{< \, |t|} d\left[G(s,t)-G(t,s)\right] = 0. \tag{13}$$

Then, if we also assume that $\int_{|s|} \int_{= \, |t|} d\,G(s,t) = 0$, we have $L \leqslant 1 \leqslant R$ for $Q(k)$ by virtue of Theorem 5. $L \leqslant 1 \leqslant R$ is a necessary and sufficient condition for the equivalence of quasi-closeness and smaller variance for kS and T for all $k > 0$. Hence, if $\lim_{n \to \infty} Q_n(\sigma_{1n}/\sigma_{2n}) = Q(\sigma)$ where $\lim_{n \to \infty} (\sigma_{1n}/\sigma_{2n}) = \sigma < 1$, then

$$\lim_{n \to \infty} P(|t_{1n}-\theta| < |t_{2n}-\theta|) \geqslant 1/2 \tag{14}$$

and we have the variate with the asymptotically smaller variance being asymptotically almost closer.

If t_{1n} is asymptotically closer to θ than is t_{2n}, then $\lim_{n \to \infty} \sigma_{1n}/\sigma_{2n} = \sigma < 1$. For, if

$$\lim_{n\to\infty} P(|t_{1n}-\theta| < |t_{2n}-\theta|) > 1/2$$

then, since $\lim_{n\to\infty} Q_n(\sigma_{1n}/\sigma_{2n}) = Q(\sigma)$, we have $Q(\sigma) > 1/2$ and hence $\sigma < L \leq 1$.

We thus have the following theorem:

Theorem 9.

Given two random variates, t_{1n} and t_{2n}, calculated from a sample of size n, with variances σ_{1n}^2 and σ_{2n}^2 respectively, such that:

(i) $S_n = \dfrac{t_{1n}-\theta}{\sigma_{1n}}$ and $T_n = \dfrac{t_{2n}-\theta}{\sigma_{2n}}$ have a joint cumulative distribution function $G_n(s,t)$ which converges uniformly to the joint cumulative distribution function $G(s,t)$ in all points of continuity of $G(s,t)$, s and t having zero means and unit variances;

(ii) $\displaystyle\int\int_{|s| < |t|} d\left[G(s,t)-G(t,s)\right] = 0;$

then:

(a) Case I: G(s,t) continuous:

Asymptotic closeness is equivalent to asymptotic smaller variance if and only if $\lim_{n\to\infty} Q_n(\sigma_{1n}/\sigma_{2n}) = Q(\sigma)$ where $\lim_{n\to\infty} \sigma_{1n}/\sigma_{2n}$ exists and if $g(0,0) \neq 0$ where $G(s,t)$ is the continuous, joint probability density function associated with $G(s,t)$;

(b) Case II: G(s,t) discontinuous:

Asymptotic quasi-closeness is equivalent to asymptotic smaller variance if $\lim_{n \to \infty} Q_n(\sigma_{1n}/\sigma_{2n}) = Q(\sigma)$ where $\lim_{n \to \infty} (\sigma_{1n}/\sigma_{2n})$ exists and if

$$\int\int_{|s| = |t|} d\, G(s,t) = 0.$$

V. EXTENSION TO BIASED ESTIMATORS

In the preceding sections, we have restricted ourselves to the consideration of unbiased estimators, that is, we have admitted only those estimators, t, for which $E(t) = \theta$ where θ is the true parameter value. Let us now examine estimators of the type $\tilde{\theta}$, where $E(\tilde{\theta}) = \theta + b(\theta)$, and deal with the mean square deviation $E\ (\tilde{\theta}-\theta)^2$ instead of the variance $E\ (\tilde{\theta}-E(\tilde{\theta}))^2$ In this, $b(\theta)$ represents the bias. Consider the case of two estimators, t_1 and t_2, of the parameter θ such that

$$\begin{aligned} E(t_1) &= \theta + b_1(\theta) \\ E(t_2) &= \theta + b_2(\theta) \end{aligned} \tag{1}$$

and restrict ourselves to the situation where t_1 and t_2 admit a continuous probability density function. As before, it is more convenient to consider the variates $X = t_1-\theta$ and $Y = t_2-\theta$. Let t_1 and t_2 have mean square deviations σ_1^2 and σ_2^2 respectively.

Let us again define

$$Q(k) = P(|kX| < |Y|) \quad , k > 0, \tag{2}$$

and assuming $f(x,y)$ to be the joint probability density function of X and Y, we can show (if $f(0,0) \neq 0$) that $Q(k)$ is a continuous, strictly monotone decreasing function of k such that $0 \leqslant Q(k) \leqslant 1$. This, of course, requires the definition of $Q(0) = 1$ and $Q(\infty) = 0$. Thus we have shown that a proposition, equivalent to the lemma proved in (III), holds good for the case of biased estimators.

If we now proceed as in the case of unbiased estimators, it is evident that all proofs will follow through in a similar manner. Hence we may conclude the following theorem:

Theorem 10.

If X and Y are two random variates which satisfy the following conditions:

(i) $E(X) = b_1(\theta)$; $E(Y) = b_2(\theta)$

(ii) $E(X^2) = \sigma_1^2$; $E(Y^2) = \sigma_2^2$

(iii) $f(x,y)$ is the joint probability density function of X and Y

(iv) $f(0,0) \neq 0$.

and, further, if $S = X/\sigma_1$ and $T = Y/\sigma_2$, then $Q(1) = P(|S| < |T|) = 1/2$ is a necessary and sufficient condition for the equivalence of closeness and minimum mean square deviation for all X and Y.

Corollary

If the conditions of the preceding theorem hold and if $g(s,t)$ is the joint probability density function of S and T, then a necessary and sufficient condition for the equivalence of closeness and minimum mean square deviation for X and Y for all values of σ_1 and σ_2 is

$$\int_{-\infty}^{\infty}\int_{-|t|}^{|t|}\left[g(s,t)-g(t,s)\right]\,ds\,dt = 0.$$

If we now consider discrete random variates rather than those which admit a continuous probability density function, it is easily seen that results similar to those obtained in (III) will be realised in any discussion involving mean square deviations instead of variances.

It is thus seen that under fairly general conditions, we may prove the equivalence of closeness and minimum mean square deviation (or quasi-closeness and minimum mean square deviation), and that when our estimators are unbiased we may replace minimum mean square deviation by minimum variance.

VI. EXTENSION TO THE MULTI-PARAMETER CASE

We have, so far, restricted ourselves to the consideration of the case involving only one unknown parameter. This restriction was forced, to some extent, since Pitman defined his concept of closeness only in the case of one parameter. However, it seems desirable to attempt the formulation of a like property for situations involving two or more parameters. Accordingly, let us consider a random variate W distributed by $f(w; \theta_1, \theta_2, \ldots, \theta_m)$ where $f(w; \theta_1, \theta_2, \ldots, \theta_m)$ satisfies all the requirements for a continuous probability density function and $\theta_1, \ldots, \theta_m$ are unknown parameters.

In order to formulate a concept of closeness in m-dimensions, it is first necessary to introduce a metric comparable to the absolute distance between two points as used in our previous work in one dimension. To this end, let us consider the following real-valued function which may be thought of as a <u>distance</u> <u>function</u> or <u>metric</u>:

$$r = ((X_1-Y_1)^2 + (X_2-Y_2)^2 + \ldots + (X_m-Y_m)^2)^{1/2} \qquad (1)$$

where $(X_1, X_2, \ldots, X_m)$ and $(Y_1, Y_2, \ldots, Y_m)$ represent any two points in our m-dimensional space.

If, now, we let $X = (X_1, X_2, \ldots, X_m)$ and $Y = (Y_1, Y_2, \ldots, Y_m)$ be any two estimators of $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$, we shall say that X is a closer estimator of θ than is Y if

$$P\left(|r_1| < |r_2|\right) > 1/2 \qquad (2)$$

where

$$r_1 = ((X_1-\theta_1)^2 + \ldots + (X_m-\theta_m)^2)^{1/2} \tag{3}$$

and

$$r_2 = ((Y_1-\theta_1)^2 + \ldots + (Y_m-\theta_m)^2)^{1/2}\ , \tag{4}$$

and (2) holds good for all values of θ.

If we represent the joint probability density function of X and Y by $f(x,y;\theta)$ and denote by $\sum_1$ and $\sum_2$ the generalized variances of $X = (X_1,\ldots,X_m)$ and $Y = (Y_1,\ldots,Y_m)$, it is possible to formulate a solution to our problem in a similar manner to that used in the one-dimensional case. In fact, we may effectively reduce the problem to a one-dimensional situation by consideration of the metric (defined above) and the generalized variance.

Applying our criterion as developed in (III), and assuming $\sum_1 < \sum_2$, it follows that the equivalence of closeness (as defined in this section) and minimum generalized variance may be deduced as before.

It is to be noted that our definition of closeness in m-dimensions does not imply that X_i is a closer estimator of θ_i than is Y_i; $i=1,\ldots,m$. We must remember that our definition of closeness for the case of two or more parameters was quite arbitrary, some other definition might easily have been used. For example, we could have required that

$$P(|X_1-\theta_1| < |Y_1-\theta_1| \ ,\ldots, |X_m-\theta_m| < |Y_m-\theta_m|) > 1/2, \tag{5}$$

or had we desired to be very restrictive we might have stipulated

$$P(|X_i-\theta_i| < |Y_i-\theta_i|) > 1/2;\ i=1,2,\ldots,m. \tag{6}$$

Each of the above definitions should be considered on its own merits. The one we have chosen to use lends itself to analysis by the methods developed earlier and therefore appears to be the most desirable definition for our purposes.

VII. SUMMARY

For a continuous bivariate distribution with the two variates possessing the same mean and finite (but unequal) variances, it has been shown that the variate having the smaller variance will be the closer estimator of the mean, except when the ratio of the variances is in a finite interval determined by the distribution. From this it was possible to obtain a necessary and sufficient condition for the equivalence of closeness and minimum variance, namely:

If X and Y are two random variates having variances σ_1^2 and σ_2^2 respectively and if (i) $E(X) = E(Y) = \theta$, (ii) $f(x,y)$ is the joint probability density function of X and Y, and (iii) $f(0,0) \neq 0$ and, further, if $S = X/\sigma_1$ and $T = Y/\sigma_2$, then a necessary and sufficient condition for closeness and minimum variance to be equivalent for X and Y is that $P(|S| < |T|) = 1/2$.

The above results were then modified slightly for the case where the variates admit only a discrete bivariate distribution.

Asymptotic closeness was defined and, under certain general conditions, it was shown that asymptotic smaller variance and asymptotic closeness were equivalent. As an extension of the theorem concerning the equivalence of asymptotic closeness and asymptotic smaller variance, it was shown that the asymptotically closest estimator (if it be normally distributed in the limit) is an asymptotically efficient estimator. With certain modifications, the results were extended to the case of variates admitting only a discrete distribution.

The results were further extended to the case of biased estimators, using the concept of minimum mean square deviation rather than minimum variance.

A definition of closeness was formulated for the multi-parameter situation and a solution, corresponding to that provided in the one-parameter case, was indicated.

VIII. CONCLUSION

The basis for the use of the criterion of closeness of estimators is expressed by Pitman (36, p. 213) in the following way:

> From the standpoint of pure knowledge, a closest estimate might be regarded as a best estimate, since it is likely to be nearer to the true value than any other estimate; but from the practical point of view, what is best depends upon the use we make of our estimates, and ultimately upon how we pay for our mistakes. If the penalty rule were "the devil take the hindmost", closest estimates would be best estimates.

This statement indicates that, with the proper choice of weight function to represent the penalty incurred by making a bad estimate, the concept of closeness as an optimum property should be derivable from Wald's general theory of statistical decision functions. Investigation of this matter is in progress.

Conditions for the equivalence of closeness and minimum variance have been obtained. When these conditions are satisfied, we have a further justification for using minimum variance estimates of an unknown parameter. A convenient method of obtaining the closest estimator of a parameter does not appear to be available for all cases. Pitman (36) has considered the following approach to this problem:

If a sufficient statistic, t, exists for the estimation of θ, where θ is the unknown parameter, it might be expected that the closest estimator of θ should be derivable from t. The median value of t will be a function of θ, say $\psi(\theta)$, and if $\psi(\theta)$ is a monotone function of θ, $\psi^{-1}(t)$ will be a sufficient statistic for estimating θ and it will have θ as its median value.

Hence the closest estimate of θ would be obtainable by equating such a sufficient statistic to its median value.

However, a general proof of the validity of this method for every type of case is not available and further investigation of this is to be desired. Also, some method of obtaining closest estimators when sufficient statistics do not exist is a problem in need of consideration.

IX. BIBLIOGRAPHY

1. Aitken, A. C. and Silverstone, H. On the estimation of statistical parameters. Proc. of the Roy. Soc. of Edinburgh, Section A. 61:186-194. 1941.

2. Bartlett, M. S. A note on the interpretation of quasi-sufficiency. Biometrika. 31:391-392. 1939.

3. Craig, A. T. A note on the best linear estimate. Ann. of Math. Stat. 14:88-90. 1943.

4. Cramér, Harald. Mathematical methods of statistics. Princeton, New Jersey. Princeton Univ. Press. 1946.

5. Doob, J. L. Probability and statistics. Trans. of the Amer. Math. Soc. 36:759-775. 1934.

6. ________ Statistical estimation. Trans. of the Amer. Math. Soc. 39:410-421. 1936.

7. Dugué, D. Sur le maximum de précision des estimations gaussiennes à la limite. Comptes Rendus Des Séances de l'Academie Des Sciences, 202:193-195. 1936.

8. ________ Sur le maximum de précision des lois limites d'estimations. Comptes Rendus Des Séances de l'Academie Des Sciences. 202:452-454.

9. ________ Sur certains modes de covergences de lois d'estimation. Comptes Rendus Des Séances de l'Academie Des Sciences. 202:1732-1734. 1936.

10. Fisher, R. A. On the mathematical foundations of theoretical statistics. Phil. Trans. of the Roy. Soc. of London, Series A. 222:309-368. 1921.

11. ________ The conditions under which χ^2 measures the discrepancy between observation and hypothesis. Jour. of the Roy. Stat. Soc. 87:442-450. 1924.

12. ________ Theory of statistical estimation. Proc. of the Cambridge Phil. Soc. 22:700-725. 1925.

13. ________ Inverse probability and the use of likelihood. Proc. of the Cambridge Phil. Soc. 28:257-261. 1932.

14. ________ The concepts of inverse probability and fiducial probability referring to unknown parameters. Proc. of the Roy. Soc. of London, Series A. 139:343-348. 1933.

15. Fisher, R. A. Two new properties of mathematical likelihood. Proc. of the Roy. Soc. of London, Series A. 144:285-307. 1934.

16. ____________ The logic of inductive inference. Jour. of the Roy. Stat. Soc. 98:39-82. 1935.

17. Geary, R. C. Comparison of the concepts of efficiency and closeness for consistent estimates of a parameter. Biometrika. 33:123-128. 1944.

18. Halmos, P. R. The theory of unbiased estimation. Ann. of Math. Stat. 17:34-43. 1946.

19. Hendricks, W. A. An application of the method of maximum likelihood. Biometrika. 29:154-155. 1937.

20. Hotelling, H. The consistency and ultimate distribution of optimum statistics. Trans. of the Amer. Math. Soc. 32:847-859. 1930.

21. Huzurbazar, V. S. The likelihood equation, consistency and the maxima of the likelihood function. Ann. of Eugenics. 14:185-200. 1948.

22. Jeffreys, H. The use of minimum χ^2 as an approximation to the method of maximum likelihood. Proc. of the Cambridge Phil. Soc. 34:156-157. 1938.

23. ____________ An invariant form for the prior probability in estimation problems. Proc. of the Roy. Soc. of London, Series A. 186:453-461. 1946.

24. Kendall, M. G. The advanced theory of statistics. London. Charles Griffin and Co. Ltd. 1946.

25. Koopman, B. O. On distributions admitting a sufficient statistic. Trans. of the Amer. Math. Soc. 39:399-409. 1936.

26. Koshal, R. S. Maximal likelihood and minimal χ^2 in relation to frequency curves. Ann. of Eugenics. 9:209-231. 1939.

27. Landau, H. G. Note on the variance and best estimates. Ann. of Math. Stat. 15:219-221. 1944.

28. Mood, A. M. Theory of statistics. Unpublished Mimeographed Class Notes. Ames, Iowa. Dept. of Stat. Iowa State College. 1948.

29. Neyman, J. Outline of a theory of statistical estimation based on the classical theory of probability. Phil. Trans. of the Roy. Soc. of London, Series A. 236:333-380. 1937.

30. ____________ Lectures and conferences on mathematical statistics. Wash., D. C. The Graduate School of the U. S. Dept. of Agr. 1937.

31. Neyman, J. and Pearson, E. S. On the use and interpretation of certain test criteria for purposes of statistical inference. Parts I and II. Biometrika. 20A:175-240 and 262-294. 1928.

32. Pearson, E. S. Maximum likelihood and methods of estimation. Biometrika. 29:155-156. 1937.

33. ______________ Note on Professor Pitman's contribution to the theory of estimation. Biometrika. 30:471-474. 1938.

34. Pearson, Karl. Method of moments and method of maximum likelihood. Biometrika. 28:34-59. 1936.

35. Pitman, E. J. G. Sufficient statistics and intrinsic accuracy. Proc. of the Cambridge Phil. Soc. 32:567-579. 1936.

36. ________________ The 'closest' estimates of statistical parameters. Proc. of the Cambridge Phil. Soc. 33:212-222. 1937.

37. ________________ The estimation of the location and scale parameters of a continuous population of any given form. Biometrika. 30:391-421. 1938.

38. Rao, C. R. Minimum variance and the estimation of several parameters. Proc. of the Cambridge Phil. Soc. 43:280-283. 1947.

39. Romanovsky, V. Note on the method of moments. Biometrika. 28:188-190. 1936.

40. Smith, Kirstine. On the 'best' values of the constants in frequency distributions. Biometrika. 11:262-276. 1915.

41. Wald, A. Contributions to the theory of statistical estimation and testing hypotheses. Ann. of Math. Stat. 10:299-326. 1939.

42. ________ On the principles of statistical inference. Notre Dame, Indiana. Univ. of Notre Dame. 1942.

43. Welch, B. L. On the distribution of maximum likelihood estimates. Biometrika. 31:187-190. 1939.

44. Wilks, S. S. Mathematical statistics. Princeton, New Jersey. Princeton Univ. Press. 1946.

X. ACKNOWLEDGEMENT

The writer wishes to thank Professor Oscar Kempthorne for constructive criticisms without which this work could not have been completed and for suggestions in the organization of the material.

He also wishes to express his gratitude to Dr. Gobind Seth of the Department of Statistics and Dr. H. P. Thielman of the Department of Mathematics for many valuable suggestions.

XI. APPENDIX

ALTERNATIVE METHOD OF PROOF RE THEOREM 7

XI. ALTERNATIVE METHOD OF PROOF RE THEOREM 7

Starting with the same assumptions as before, assume that the proof has been carried to the point where we have shown that $Q_n(k)$ converges uniformly to $Q(k)$. We might then proceed as follows:

Consider the expression:

$$Q_n(\sigma_{1n}/\sigma_{2n}) - Q(\sigma) = Q_n(\sigma_{1n}/\sigma_{2n}) - Q(\sigma_{1n}/\sigma_{2n}) + Q(\sigma_{1n}/\sigma_{2n}) - Q(\sigma) \tag{1}$$

Because of the uniform convergence, we have

$$\lim_{n \to \infty} (Q_n(\sigma_{1n}/\sigma_{2n}) - Q(\sigma_{1n}/\sigma_{2n})) = 0. \tag{2}$$

Therefore

$$\lim_{n \to \infty} (Q_n(\sigma_{1n}/\sigma_{2n}) - Q(\sigma)) = \lim_{n \to \infty} (Q(\sigma_{1n}/\sigma_{2n}) - Q(\sigma)). \tag{3}$$

Now, if $\lim\limits_{n \to \infty} \sigma_{1n}/\sigma_{2n} = \sigma < 1$, then

$$\lim_{n \to \infty} (Q(\sigma_{1n}/\sigma_{2n}) - Q(\sigma)) = 0$$

and hence

$$\lim_{n \to \infty} Q_n(\sigma_{1n}/\sigma_{2n}) = Q(\sigma)$$

We have thus reached the result exhibited in the development leading to Theorem 7 (cf. equation (29)).