ECONOMETRICA

VOLUME 16 JANUARY, 1948 NUMBER 1

CONSISTENT ESTIMATES BASED ON PARTIALLY CONSISTENT OBSERVATIONS*

By J. NEYMAN AND ELIZABETH L. SCOTT

CONTENTS

1. INTRODUCTION

THE present paper contains general definitions and some results relating to a category of problems that appear to be very frequent and important in many fields of applications, but that thus far have escaped systematic study.

The problems that we have in mind may be labeled as the problems of *consistent estimates based on partially consistent observations*. Let x_i stand for a (possibly multivariate) random variable and assume that the variables of the sequence $x_1, x_2, \cdots, x_n, \cdots$ are mutually independent. Assume further that the probability laws of the x_i contain some unknown parameters $\Theta_1, \Theta_2, \cdots$ and consider the problem of consistent estimates of these parameters, that is to say, the problem of determining such functions $T_k(x_1, \cdots, x_n)$ of the random variables that, whatever be $\epsilon > 0$, the probability that

$$(1) \qquad | T_k - \Theta_k | > \epsilon$$

tends to zero as n is indefinitely increased.

The recent book of H. Cramér [1][1] summarizes and gives some very elegant developments relating to the important case where the

* Paper presented at the Annual Meeting of the Institute of Mathematical Statistics in Atlantic City, N. J., Jan. 25, 1947.

[1] Numbers in square brackets refer to list of references at the end.

number of unknown parameters involved in the probability laws of the observable random variables is finite and where each parameter appears in all (or at least an infinite number) of the probability laws. In this case it could be said roughly that the observable variables $\{x_i\}$ are "*consistent.*" In fact, the knowledge of any one of them tends to contribute something to our knowledge of the totality of the unknown parameters $\Theta_1, \Theta_2, \cdots$. This case is obviously very important and for this reason has attracted attention since Gauss. In particular, we have the almost universal method of obtaining estimates that are not only consistent but also asymptotically efficient, in the sense that (1) they tend to be normally distributed as the number of observations grows and (2) their variances tend to zero at least as fast as those of any other estimates. Of course, the method in question is that of maximum likelihood.

Although cases of consistent variables are very important, it is interesting that other cases occur frequently in applications. Of these we will consider especially cases of variables that we shall describe as *partially consistent.* This is the situation where the set of unknown parameters involved in the totality of probability laws of the random variables $\{x_i\}$ is infinite and can be split into two parts. The first part is composed of a *finite* number of parameters, say $\Theta_1, \Theta_2, \cdots, \Theta_\nu$, each of which appears in the probability laws of an *infinity* of random variables of the sequence $\{x_i\}$. In respect to these parameters $\Theta_1, \Theta_2, \cdots, \Theta_\nu$ the sequence $\{x_i\}$ is, then, consistent. The second part of the set of unknown parameters is infinite and is composed of parameters ξ_m each of which appears in the probability law of only a finite number of the random variables considered. Thus, in the simplest case, ξ_m may appear only in the probability law of x_m, for $m=1, 2, \cdots$.

Alternatively ξ_1 may appear in the probability laws of $x_1, x_2, \cdots, x_{m_1}$; ξ_2 in the probability laws of $x_{m_1+1}, \cdots, x_{m_1+m_2}$; etc. Since the random variables x_k are allowed to be multivariate, it will be seen that this particular case does not differ essentially from the first.

Really more complicated cases occur when the parameters $\xi_1, \xi_2, \cdots$ appear in several probability laws in varying combinations, for example, ξ_1 and ξ_2 in the probability law of x_1; ξ_2 and ξ_3 in the probability law of x_2; etc.

Since it is convenient to have labels for the conceptions studied, the parameters, $\Theta_1, \Theta_2, \cdots, \Theta_\nu$ each appearing in an infinity of probability laws of the observable random variables will be called *structural.* All the other parameters, $\xi_1, \xi_2, \cdots$, an infinity of them, will be called *incidental.*

This first study will be limited to the case where there is a one-to-one correspondence between the observable random variables x_i and the incidental parameters ξ_i so that the probability law of x_i depends on ξ_i but not on any other incidental parameter.

As to the structural parameters, it will be assumed that the probability law of each x_i depends on some of these parameters, but not necessarily on all of them. Nevertheless, for convenience in formulae, whenever necessary the probability law of x_i will be written with all the Θ's appearing as its arguments. To be definite, we will assume that the observable random variables are all continuous and denote by $p_i(x_i | \Theta_1, \cdots, \Theta_\nu, \xi_i)$ the probability density function of x_i.

2. EXAMPLES

Before proceeding to results, it may be useful to indicate the generality of the problem by presenting a few examples.

EXAMPLE (1). Let α be some physical constant such as the radial velocity of a star or the velocity of light.[2] Assume that s series of measurements are to be made and let x_{ij} stand for the result of the jth measurement of the ith series ($i=1, 2, \cdots, s$; $j=1, 2, \cdots, n_i$). We will assume that the measurements follow the normal law with the same mean α and an unknown standard error σ_i which may and probably does vary from one series of observations to another. Thus the probability density function of x_{ij} is

$$(2) \qquad p_{ij}(x_{ij} \mid \alpha, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-(x_{ij}-\alpha)^2/2\sigma_i^2}.$$

This is exactly the case when α stands for the radial velocity of a star and the x_{ij} are its measurements obtained from n_i different spectral lines on the ith plate. With the velocity of light the situation is similar. This is also the situation in all cases where it is desired to combine measurements of physical quantities, made in different laboratories, by different experimenters, etc.

In order to bring this example into correspondence with the above description of the general situation, notice that the unknown parameter α appears in all the probability laws of the observable variables. This, then, is the structural parameter with respect to which the observable random variables are consistent.

In addition to the structural parameter α, each of the probability laws (2) depends on another unknown parameter σ_i. However, there are only n_i probability laws depending on a particular σ_i. Thus each $\sigma_i (i=1, 2, \cdots)$ is an incidental parameter. If we use the letter x_i with just one subscript to denote the whole ith series of measurements

$$(3) \qquad x_i = \{x_{i1}, x_{i2}, \cdots, x_{in_i}\},$$

[2] Here we presume that the velocity of light is a constant.

then the correspondence between the example and the general description will be established completely.

EXAMPLE (2). Consider again an increasing sequence of s series of measurements $x_{ij}(i=1, 2, \cdots, s \to \infty; j=1, 2, \cdots, n_i)$. Assume again that all the measurements are mutually independent and follow a normal law. However, this time it will be assumed as given that the precision of measurements does not change from one series to another and is characterized by a single unknown standard error σ. On the other hand, it will be assumed that the quantity measured in the ith series, say α_i, is unknown and does not necessarily equal α_k, that corresponding to the kth series. Thus the probability law of the x_{ij} can be written as

$$p_{ij}(x_{ij} \mid \alpha_i, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_{ij}-\alpha_i)^2/2\sigma^2}. \tag{4}$$

Here we have a partially consistent system of observable random variables with just one structural parameter and the incidental parameters α_i. Situations of this kind occur when the same apparatus is used to make limited amounts of measurements of different quantities and it is desired to combine all those measurements to obtain a single estimate of the standard error σ.

Essentially, Example (2) is very similar to Example (1). The reason for quoting both will be apparent at a later stage when we will discuss certain phenomena relating to some current methods of obtaining estimates. Needless to say, both the above examples are of a very common and familiar type.

EXAMPLE (3). This example will be considered in some detail in a separate paper by one of the authors. Let ξ and η be two physical quantities assuming varying values and known to be connected by an equation of some specified type with a few unknown parameters. For example, it may be known that η is a linear function of ξ

$$\eta = \alpha + \beta\xi \tag{5}$$

where α and β are unknown.

Suppose that in some increasing series of s instances, measurements of the corresponding values of ξ and η are made. The true unknown values of ξ and η in the ith instance will be denoted by ξ_i and η_i and the corresponding measurements respectively by

$$\begin{array}{l} x_{i1}, x_{i2}, \cdots, x_{im_i} \\ y_{i1}, y_{i2}, \cdots, y_{in_i}. \end{array} \tag{6}$$

It is further assumed that the measurements are normally distributed about the true means ξ_i and η_i respectively and that the precision of measurement of ξ is always the same σ_1 and that of η is also always the same σ_2, both constants σ_1 and σ_2 being unknown. The probability density function relating to the ith series of observations will be written as, say

$$(7) \quad p_i = \left(\frac{1}{\sigma_1\sqrt{2\pi}}\right)^{m_i} e^{-\sum_{j=1}^{m_i}(x_{ij}-\xi_i)^2/2\sigma_1^2} \left(\frac{1}{\sigma_2\sqrt{2\pi}}\right)^{n_i} e^{-\sum_{j=1}^{n_i}(y_{ij}-\alpha-\beta\xi_i)^2/2\sigma_2}$$

for $i=1, 2, \cdots, s, \cdots$. It frequently happens that each of the series of measurements is very short, so that both m_i and n_i are small numbers. On the other hand, the number s of independent series is large and, in principle, can be increased without limit.

Situations of the above kind relate to the familiar problem of "fitting a straight line when both variables are subject to errors." The problem is old and was studied by a number of authors from Adcock in 1877 to Abraham Wald [2] who gives an extensive bibliography. Recently, the problem also was studied by F. H. Seares [3], [4], [5].

Needless to say, the situation described in Example (3) is the simplest of a broad category of similar situations. The more complicated ones may involve a third variable ζ_i (or more). Also the precision of measurements may vary from one series to another.

As things stand, the sequence of series of observations (6) is consistent with respect to the parameters, α, β, σ_1, and σ_2 which are structural. Besides, there is a system of incidental parameters $\xi_1, \xi_2, \cdots, \xi_s$, each corresponding to a separate series of observations.

Situations of the kind just described occur frequently in physics and astronomy where, for example ξ_i and $\eta_i=\alpha+\beta\xi_i$ may stand for some characteristics of the ith star. It is our impression, however, that the same theoretical model, perhaps with several more structural and incidental parameters, is likely to be applicable to some economic problems. Some sentences in the paper of Wald just quoted seem to indicate that this would be his opinion also. Similar suggestions were made by Ragnar Frisch [6].

In this respect two questions seem to be particularly interesting. First there is the question whether it is reasonable to distinguish between, say, the "true demand" (or some other economic conception) whose measure at a given moment is given by an incidental parameter ξ_i, and the "measured demand," expressed by some random variable x_{ij}. Another, probably less important, question is whether more than one independent measurement $x_{ij}(j=1, 2, \cdots, n_i)$ of the same "true demand" ξ_i is possible. This latter question is connected with the known fact that, apart from the situation considered by Wald which may not be of general occurrence, the presence of at least two measurements in each series appears to be necessary for the possibility of consistent estimates of all the structural parameters.

A paper by W. G. Cochran suggests that a similar set-up may have interesting applications in treating agricultural experiments [7].

EXAMPLE (4). The examples above have a certain common feature. In each case we have a sequence $\{x_i\}$ of possibly multivariate independent variables with the probability laws $p_i(x_i | \Theta_1, \Theta_2, \cdots, \Theta_\nu)$, which depend on *all* of the structural parameters $\Theta_1, \Theta_2, \cdots, \Theta_\nu$ involved in the problem. Example (4) is intended to exhibit a practical problem in which this is not necessarily the case. Certain important problems of astronomy, connected with the study of the dynamics of the galaxy, lead to the following set-up.

With the ith star of a sequence there are connected several, say three, numbers a_i, b_i, c_i which may be taken as known without error. For example, they may be known functions of the angular coordinates of the star which are determined with great accuracy. Next, there are two quantities ξ_i and η_i which are not measurable with anything like the precision of the numbers a_i, b_i, c_i.

For example, ξ_i may mean the distance from the star to the sun and η_i the expectation of the radial velocity of the same star when the component of the sun's motion has been removed. It may be taken as given that the above quantities are related by the equation

$$\eta_i = \xi_i(\alpha a_i + \beta b_i + \gamma c_i), \tag{8}$$

where α, β, and γ are unknown coefficients, independent of the particular star and relevant to the theory of the dynamics of the galaxy.

Now, instead of the true value of ξ_i we have its measurements $x_{i1}, x_{i2}, \cdots, x_{im_i}$ which are subject to large errors. The quantity η_i is not directly measurable either. Instead, we have measurements y_{ij} of the quantity, say Y_i, representing the true radial velocity of the ith star. The variability of the y_{ij}, caused by the errors of measurement, is appreciable. Moreover, Y_i is itself considered as a normal random variable with its true mean equal to η_i and standard error σ independent of i. It is seen that, in this example, each additional star brings in several new observable random variables depending on a new parameter ξ_i. Furthermore, the probability law of these additional variables is generally somewhat different from the previous laws, owing to the changes in the constants a_i, b_i, and c_i. For example, a_i may be equal to zero for some stars so that these will contribute little to the estimation of α.

Modern work by the econometric school suggests that similar set-ups may be applicable in studies of economic phenomena.

3. METHOD OF MAXIMUM LIKELIHOOD APPLIED TO PARTIALLY CONSISTENT OBSERVATIONS

The method of maximum likelihood appears so strongly appealing that, when confronted with a new problem, one tends to apply maximum likelihood almost automatically. It must be remembered, however, that the various attractive properties of the method were proved only for consistent series of observations and that, therefore, an automatic extension to other cases may lead to erroneous results. These are illustrated by the following examples. Naturally, if consistent estimates based on partially consistent observations are possible at all, we may expect them only for structural parameters, because the incidental parameters appear in the probability laws of only a finite number of observable variables. Therefore, the problem of consistency will be considered in relation to the structural parameters only.

(1) Maximum-likelihood estimates of the structural parameters relating to a partially consistent series of observations need not be consistent.

To prove this proposition consider the above Example (2). Easy algebra shows that the maximum-likelihood estimates of all the parameters involved are

$$\hat{\xi}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = x_{i.} \quad \text{(say)} \tag{9}$$

and

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{s} \sum_{j=1}^{n_i} (x_{ij} - x_{i.})^2}{\sum_{i=1}^{s} n_i}, \tag{10}$$

or, if we put for simplicity $n_i = n = \text{const.}$,

$$\hat{\sigma}^2 = \frac{1}{s} \sum_{i=1}^{s} S_i^2, \tag{11}$$

where

$$S_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - x_{i.})^2. \tag{12}$$

As is well known, the expectation

$$\mathcal{E}(S_i^2) = \frac{n-1}{n} \sigma^2. \tag{13}$$

Since for every $i=1, 2, \cdots, s, \cdots$ the sample variance S_i^2 follows the same law, the arithmetic mean representing $\hat{\sigma}^2$ tends in probability to the common expectation (13) and thus the maximum-likelihood estimate $\hat{\sigma}^2$ is not consistent as $s \to \infty$,

$$P \lim_{s \to \infty} \hat{\sigma}^2 = \frac{n-1}{n} \sigma^2. \tag{14}$$

If n is very small, say $n=2$, the underestimate of the parameter σ^2 is considerable.

It may be said that the situation is trivial and the bias in the estimate could easily be corrected by using, say

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^{s} \sum_{j=1}^{n_i} (x_{ij} - x_{i.})^2}{\sum_{i=1}^{s} (n_i - 1)} \cdot \tag{15}$$

This is undoubtedly true but beside the point. It will be observed that (15) is not the maximum-likelihood estimate of σ^2 and that the bias in the latter, given by (10), does not tend to zero as the series of partially consistent observation is increased. This is just the circumstance which the example is meant to illustrate.

(*2*) *Even if the maximum-likelihood estimate of a structural parameter is consistent, if the series of observations is only partially consistent, the maximum-likelihood estimate need not possess the property of asymptotic efficiency.*

To illustrate this point we will consider Example (1) in more detail [7]. The probability density function of the totality of observable variables is the product of expressions (2) for $j=1, 2, \cdots, n_i$ and $i=1, 2, \cdots, s$. Taking the logarithm of this product and differentiating with respect to α and with respect to σ_i we get, say

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^{s} \frac{n_i(x_{i\cdot} - \alpha)}{\sigma_i^2} \tag{16}$$

and

$$\frac{\partial L}{\partial \sigma_i} = -\frac{n_i}{\sigma_i} + \frac{n_i v_i}{\sigma_i^3}, \tag{17}$$

where

$$x_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \tag{18}$$

(19) $$n_i v_i = \sum_{j=1}^{n_i} (x_{ij} - \alpha)^2 = n_i[S_i^2 + (x_{i.} - \alpha)^2],$$

(20) $$n_i S_i^2 = \sum_{j=1}^{n_i} (x_{ij} - x_{i.})^2.$$

The maximum-likelihood equations are then

(21) $$\sigma_i^2 = S_i^2 + (x_{i.} - \hat{\alpha})^2$$

and, say,

(22) $$\sum_{i=1}^{s} F_i(x_i, \hat{\alpha}) \equiv \sum_{i=1}^{s} \frac{n_i(x_{i.} - \hat{\alpha})}{S_i^2 + (x_{i.} - \hat{\alpha})^2} = 0.$$

We will be primarily interested in equation (22) or, rather in a little more general equation (23) obtained from (22) by substituting an arbitrary number w_i instead of n_i. Thus, we will consider the estimate a_s of α obtained from the equation, say,

(23) $$\sum_{i=1}^{s} \Phi_i(x_{i.}, a_s) \equiv \sum_{i=1}^{s} w_i \frac{(x_{i.} - a_s)}{S_i^2 + (x_{i.} - a_s)^2} = 0.$$

It is easy to see that the estimate a_s tends in probability to α provided some mild conditions are satisfied. To prove this, one has only to follow the reasoning of Cramér ([1], p. 500) leading to the proof of the consistency of the maximum-likelihood estimates when the observable variables are consistent. To establish the consistency of a_s as $s \to \infty$ it is sufficient to show

(a) That

(24) $$\mathcal{E}[\Phi_i(x_{i.}, \alpha)] \underset{\alpha}{\equiv} 0 \qquad \text{for } i = 1, 2, \cdots;$$

(b) That the variance $\sigma_{\Phi s}^2$ of

(25) $$\frac{1}{s} \sum_{i=1}^{s} \Phi_i(x_{i.}, \alpha)$$

tends to zero as $s \to \infty$;

(c) That there exists a number $A > 0$ such that, as $s \to \infty$,

(26) $$\lim_{s \to \infty} P\left\{\frac{1}{s}\left|\sum_{i=1}^{s} \Phi_i'(x_{i.}, \alpha)\right| > A\right\} = 1,$$

where Φ_i' denotes the derivative of Φ_i with respect to α;

(d) Finally, that there exist some function $H_i(x_i)$ independent of α, such that, say

$$(27)\qquad |\Phi_i''(x_{i\cdot}, \alpha)| \equiv \left|\frac{\partial^2 \Phi_i}{\partial \alpha^2}\right| < H_i(x_i)$$

and such that

$$(28)\qquad \lim_{s\to\infty} P\left\{\frac{1}{s}\sum_{i=1}^{s} \mathcal{E}[H_i(x_{i\cdot})] < M\right\} = 1,$$

the number M being independent of all the parameters involved.

We have

$$(29)\qquad \Phi_i'(x_{i\cdot}, \alpha) = w_i\left\{\frac{-1}{v_i} + \frac{2(x_{i\cdot} - \alpha)^2}{v_i^2}\right\},$$

$$(30)\qquad \Phi_i''(x_{i\cdot}, \alpha) = w_i\left\{-\frac{6(x_{i\cdot} - \alpha)}{v_i^2} + \frac{8(x_{i\cdot} - \alpha)^3}{v_i^3}\right\}.$$

Denote by $M_{k,m}^{(i)}$ the expectation of the product $[(x_{i\cdot} - \alpha)^k v^m]$.

We find by easy algebra

$$(31)\qquad M_{k,m}^{(i)} = 0 \text{ for all odd values of } k = 1, 3, 5, \cdots$$

and

$$(32)\qquad M_{2k,m}^{(i)} = \frac{\Gamma(\frac{1}{2} + k)}{\Gamma(\frac{1}{2})} \frac{\Gamma\left(\frac{n_i}{2} + k + m\right)}{\Gamma\left(\frac{n_i}{2} + k\right)} \left(\frac{2\sigma_i^2}{n_i}\right)^{k+m}$$

for all values of the constants k and m for which the symbols on the right-hand side make sense.

By use of (31) it follows that condition (a) is satisfied. The variance $\sigma_{\Phi_s}^2$ is found to be

$$(33)\qquad \sigma_{\Phi_s}^2 = \frac{1}{s^2}\sum_{i=1}^{s} w_i^2 M_{2,-2}^{(1)} = \frac{1}{s^2}\sum_{i=1}^{s} w_i^2 \frac{1}{(n_i - 2)\sigma_i^4}.$$

It is seen that as $s \to \infty$, the variance $\sigma_{\Phi_s}^2$ will tend to zero provided that w_i and the σ_i^2 satisfy but very mild conditions. For example, it would be sufficient to assume that the expressions

$$(34)\qquad \frac{w_i^2}{(n_i - 2)\sigma_i^2}$$

remain bounded or that they do not increase too fast. Really this restriction would apply to the ratios w_i^2/σ_i^2.

Proceeding to condition (c), we find the expectation of $\Phi_i'(x_i, \alpha)$

$$\mathcal{E}[\Phi_i'(x_{i.}, \alpha)] = w_i\{-M_{0,-1} + 2M_{2,-2}\} = -\frac{w_i}{\sigma_i^2}. \tag{35}$$

Similarly, the variance of the arithmetic mean

$$\frac{1}{s}\sum_{i=1}^{s} \Phi_i'(x_{i.}, \alpha),$$

say $\sigma^2_{\Phi' s}$ is found to be

$$\begin{aligned} \sigma^2_{\Phi' s} &= \frac{1}{s^2}\sum_{i-1}^{s} w_i^2\left\{M_{0,-2} - 4M_{2,-3} + 4M_{4,-4} - \frac{1}{\sigma_i^4}\right\} \\ &= \frac{1}{s^2}\sum_{i=1}^{s} w_i^2 \frac{2(n_i^2 + 4n - 8)}{(n_i^2 - 4)(n_i - 4)\sigma_i^4}. \end{aligned} \tag{36}$$

If the terms of the sum in the right-hand side do not increase too fast [condition similar to that already assumed in connection with (34)], then the variance (36) will tend to zero as $s \to \infty$ and the average

$$\frac{1}{s}\sum_{i=1}^{s}\left(\Phi_i' + \frac{w_i}{\sigma_i^2}\right) \tag{37}$$

will tend in probability to zero. Thus if the average

$$\frac{1}{s}\sum_{i=1}^{s}\frac{w_i}{\sigma_i^2} \tag{38}$$

is bounded from zero as $s \to \infty$, then condition (c) will be satisfied.

As to condition (d), we notice that the two terms in brackets (30) are necessarily of different signs. Thus the absolute value of the total does not exceed the greater of the absolute values of the particular terms. Further

$$\frac{|x_{i.} - \alpha|}{v_i^2} = \frac{|x_{i.} - \alpha|}{\{S_i^2 + (x_{i.} - \alpha)^2\}^{1/2}} \frac{1}{\{S_i^2 + (x_{i.} - \alpha)^2\}^{3/2}} \leqq \frac{1}{S_i^3} \tag{39}$$

and similarly

$$\frac{|x_{i.} - \alpha|^3}{v_i^3} \leqq \frac{1}{S_i^3}. \tag{40}$$

Thus

$$|\Phi''(x_{i.}, \alpha)| \leqq \frac{8w_i}{S_i^3}. \tag{41}$$

Since

$$\mathcal{E}(S_i^k) = c\int_0^\infty S_i^{n_i+k-2}e^{-n_iS_i^2/2\sigma_i^2}dS_i$$

(42)
$$= \frac{\Gamma\left(\frac{n_i+k-1}{2}\right)}{\Gamma\left(\frac{n_i-1}{2}\right)}\left(\frac{2\sigma_i^2}{n_i}\right)^{\frac{1}{2}k},$$

thus

(43)
$$\mathcal{E}(S_i^{-3}) = \frac{\Gamma\left(\frac{n_i}{2}-2\right)}{\Gamma\left(\frac{n_i-1}{2}\right)}\left(\frac{n_i}{2\sigma_i^2}\right)^{3/2},$$

(44)
$$\mathcal{E}(S_i^{-6}) = \frac{\Gamma\left(\frac{n_i-1}{2}-3\right)}{\Gamma\left(\frac{n_i-1}{2}\right)}\left(\frac{n_i}{2\sigma_i^2}\right)^3 = \frac{n_i^3}{(n_i-3)(n_i-5)(n_i-7)\sigma_i^6},$$

provided $n_i \geqq 8$.[3]

If the numbers w_i/σ_i^2 do not increase too fast, the n_i's being bounded, then the arithmetic mean

(45)
$$\frac{1}{s}\sum_{i=1}^{s}\mathcal{E}(S_i^{-3})$$

will remain bounded. Moreover, under the same conditions the variance of the mean

(46)
$$\frac{1}{s}\sum_{i=1}^{s}H_i(x_i) \equiv \frac{1}{s}\sum_{i=1}^{s}\frac{8w_i}{S_i^3}$$

will tend to zero. This is sufficient to assert the existence of the number M satisfying condition (28) with $H_i = 8w_i/S_i^3$.

Thus, the solution a_s of equation (23) is a consistent estimate of α. It will be noticed that all the assumptions made reduce to the statement that the precision of measurements, as measured by the reciprocal

[3] This condition can be relaxed but, since the example considered is meant for illustrative purposes only, we thought it useful to treat it in the simplest possible way.

of σ_i, is not negligible, at least in sufficiently many of the series of observations, and that the weights w_i are properly adjusted.

Under the same general conditions, exactly following the reasoning applicable to maximum-likelihood estimates in the case of consistent observations, we can prove that as $s \to \infty$

$$\frac{\sqrt{s}\,(a_s - \alpha)}{\sigma_a} \tag{47}$$

tends to be normally distributed about zero with unit variance, where

$$\sigma_a^2 = \frac{\sum_{i=1}^{s} \mathcal{E}[\Phi_i^2(x_i, \alpha)]}{\left[\sum_{i=1}^{s} \mathcal{E}[\Phi_i'(x_i, \alpha)]\right]^2} = \frac{\sum_{i=1}^{s} \frac{w_i^2}{(n_i - 2)\sigma_i^2}}{\left[\sum_{i=1}^{s} \frac{w_i}{\sigma_i^2}\right]^2} \tag{48}$$

is the asymptotic variance of a_s.

By straightforward algebra it is easy to verify that (48) can be put into the following form which exhibits the dependence of σ_a on the choice of the weights

$$\sigma_a^2 = \frac{1}{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}} + \frac{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}\left(\frac{w_i}{n_i - 2} - U\right)^2}{\left(\sum_{i=1}^{s} \frac{w_i}{\sigma_i^2}\right)^2} \tag{49}$$

with

$$U = \frac{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2} \frac{w_i}{n_i - 2}}{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}} = \frac{\sum_{i=1}^{s} \frac{w_i}{\sigma_i^2}}{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}} \tag{50}$$

representing a weighted mean of the ratios $w_i/(n_i - 2)$.

It is obvious now that the precision of the estimate a_s is greatest when the weights w_i are chosen to be

$$w_i = n_i - 2 \tag{51}$$

in which case the estimate a_s may be denoted as A_s with the asymptotic variance, say

$$\sigma_{A_s}^2 = \frac{1}{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}}. \tag{52}$$

As to the maximum-likelihood estimate $\hat{\alpha}$, its asymptotic variance is obtained from (48) by putting $w_i = n_i$, say,

$$\sigma_\alpha^2 = \frac{\sum_{i=1}^{s} \frac{n_i^2}{(n_i - 2)\sigma_i^2}}{\left(\sum_{i=1}^{s} \frac{n_i}{\sigma_i^2}\right)^2} . \tag{53}$$

Formula (53) can also be written in another way which exhibits its relation to (52), namely

$$\sigma_\alpha^2 = \sigma_{A_s}^2 + \frac{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}\left(\frac{n_i}{n_i - 2} - V\right)^2}{\left(\sum_{i=1}^{s} \frac{n_i}{\sigma_i^2}\right)^2}, \tag{54}$$

where V stands for the weighted average of the ratios $n_i/(n_i - 2)$,

$$V = \frac{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2} \frac{n_i}{n_i - 2}}{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}} . \tag{55}$$

It follows from (54) that unless, for every $i = 1, 2, \cdots,$

$$\frac{n_i}{n_i - 2} = V, \qquad n_i = \frac{2V}{V - 1} = \text{const.}, \tag{56}$$

then the variance σ_α^2 of the maximum-likelihood estimate is always greater than the variance $\sigma_{A_s}^2$. The two coincide when the numbers n_i of observations in particular series are all equal. Leaving aside this latter case, it is interesting to consider the ratio of the two asymptotic variances

$$\frac{\sigma_{A_s}^2}{\sigma_\alpha^2} = \frac{1}{1 + \dfrac{\sum_{i=1}^{s} \frac{n_i - 2}{\sigma_i^2}\left(\frac{n_i}{n_i - 2} - V\right)^2}{\sum_{i=1}^{s} \frac{n_i}{\sigma_i^2}} \dfrac{\sum_{i=1}^{s} \frac{n_i - 2}{n_i} \frac{n_i}{\sigma_i^2}}{\sum_{i=1}^{s} \frac{n_i}{\sigma_i^2}}} . \tag{57}$$

It is easily seen from (57) that, as $s \to \infty$, the ratio $\sigma_{A_s}^2/\sigma_\alpha^2$ need not tend to unity so that, using a familiar phrase, the maximum-likelihood estimate need not tend to "exhaust all the information supplied by the sample" and, in fact, may exhaust less of this information than the alternative consistent estimate A_s.

Some further facts are interesting to notice. The inequality

$$\sigma_\alpha^2 > \sigma_{A_s}^2 \tag{58}$$

holds good wherever not all the numbers n_i are equal and irrespective of the values of the σ_i^2. Thus, it may be said that the estimate A_s is uniformly more efficient than the maximum-likelihood estimate $\hat{\alpha}_s$. However, there exist estimates of α which *for some particular systems of values of the* σ_i^2 are even more efficient than A_s. One such estimate is given by the simple formula

$$\bar{x} = \frac{\sum_{i=1}^{s} n_i x_i}{\sum_{i=1}^{s} n_i}. \tag{59}$$

It is easy to see that the variance $\sigma_{\bar{x}}^2$ is equal to

$$\sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{s} n_i \sigma_i^2}{\left(\sum_{i=1}^{s} n_i\right)^2} \tag{60}$$

whatever be σ_i^2. Assume for a moment that $n_i = n =$ constant and $\sigma_i = \sigma =$ constant. Then

$$\sigma_{A_s}^2 = \frac{\sigma^2}{s(n-2)}, \quad \text{and} \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{sn} < \sigma_{A_s}^2. \tag{61}$$

On the other hand, if all the $n_i = n$ but the σ_i^2 differ, then

$$\sigma_{A_s}^2 = \frac{\sigma_H^2}{sn\left(1 - \frac{2}{n}\right)} \quad \text{and} \quad \sigma_{\bar{x}}^2 = \frac{1}{sn}\,\frac{1}{s}\sum_{i=1}^{s} \sigma_i^2, \tag{62}$$

where σ_H^2 is the harmonic mean of the σ_i^2, defined by the equation

$$\frac{1}{\sigma_H^2} = \frac{1}{s}\sum_{i=1}^{s} \frac{1}{\sigma_i^2}. \tag{63}$$

Since the harmonic mean of any numbers that are not all equal is always less than the arithmetic mean, it follows that, provided n is sufficiently large,

$$\sigma_{A_s}^2 < \sigma_{\bar{x}}^2. \tag{64}$$

Thus, depending on the numbers n_i and σ_i, the variance (60) of the estimate $\bar{x}$ may be either smaller or greater than the variance $\sigma_{A_s}^2$. The application of the elegant inequality of Cramér-Rao [1], [8] shows that, whatever be an *unbiased* estimate of α, its variance cannot be less than the limit

$$\frac{1}{\sum_{i=1}^{s} \frac{n_i}{\sigma_i^2}} . \tag{65}$$

However, the question of whether any unbiased estimate of α or, even, whether any asymptotically consistent estimate of α whose asymptotic variance attains the minimum (55) exists, remains open.

To those accustomed to the working of maximum-likelihood estimates both statements (i) and (ii) must appear striking. The authors must confess that, to their mind, the possible lack of asymptotic efficiency of the maximum-likelihood estimates when they are consistent appears to be more surprising than the possibility of bias.

4. SEARCH FOR A SYSTEMATIC METHOD OF OBTAINING CONSISTENT ESTIMATES

Consider the case of some ν structural parameters $\Theta_1, \Theta_2, \cdots, \Theta_\nu$ where the probability law of each of the successive observable multivariate variables x_i depends on just one incidental parameter ξ_i which does not appear in the other laws.

The possibility, or, at least, one of the possibilities, of obtaining consistent estimates $t_i(x_1, x_2, \cdots, x_s)$ of the parameters Θ_i, $i=1, 2, \cdots, \nu$, depends upon the possibility of determining ν sequences of functions, say

$$F_{si} = F_{si}(x_1, x_2, \cdots, x_s \mid \Theta_1, \Theta_2, \cdots, \Theta_\nu) \qquad \text{for } s = 1, 2, \cdots \tag{66}$$

each depending on $x_1, x_2, \cdots, x_s$ and on $\Theta_1, \Theta_2, \cdots, \Theta_\nu$, but not on the incidental parameters, and having the following properties. Let $t_1, t_2, \cdots, t_\nu$ stand for continuous sure variables. Let one letter X_s stand for $x_1, x_2, \cdots, x_s$, one letter T stand for $t_1, t_2, \cdots, t_\nu$ and, finally, a single letter Θ stand for $\Theta_1, \Theta_2, \cdots, \Theta_\nu$.

(a) As $s \to \infty$, each function F_{si} tends in probability to zero, *irrespective of the values of the parameters* Θ *and* ξ,

$$P \lim_{s\to\infty} F_{si}(X_s \mid \Theta) \equiv 0. \tag{67}$$

To formulate the next conditions, substitute T for Θ and expand the function $F_{si}(X_s | T)$ in a Taylor series about the point $T=\Theta$,

$$(68) \qquad F_{si}(X_s \mid T) = F_{si}(X_s \mid \Theta) + \sum_{k=1}^{\nu} (t_k - \Theta_k) F_{sik} + R_{si},$$

where

$$(69) \qquad F_{sik} = \frac{\partial F_{si}(X_s, \Theta)}{\partial \Theta_k}.$$

Further, assume the following notation,

$$(70) \qquad \rho^2 = \sum_{k=1}^{\nu} (t_k - \Theta_k)^2,$$

$$(71) \qquad \Delta_s = \begin{vmatrix} F_{s11}, & F_{s12}, & \cdots, & F_{s1\nu} \\ F_{s21}, & F_{s22}, & \cdots, & F_{s2\nu} \\ \cdot & \cdot & \cdots & \cdot \\ F_{s\nu 1}, & F_{s\nu 2}, & \cdots, & F_{s\nu\nu} \end{vmatrix}.$$

Finally, let Δ_{ski} be the cofactor of F_{ski} in Δ_s. Obviously, Δ_s and the Δ_{ski} will be functions of X_s and Θ but are independent of t_k.

(b) There exists a number M such that the probability of the simultaneous inequalities $|\Delta_{ski}|/|\Delta_s| < M$ tends to unity as $s \to \infty$, that is,

$$(72) \qquad \lim_{s\to\infty} P\left\{ \prod_{k=1}^{\nu} \prod_{i=1}^{\nu} \left(\left| \frac{\Delta_{ski}}{\Delta_s} \right| < M \right) \right\} = 1.$$

(c) There exists a number $\rho_0 > 0$ such that, whatever be $\epsilon > 0$, the probability of the ν simultaneous inequalities $|R_{si}|/\rho < \epsilon$ ($i=1, 2, \cdots, \nu$) tends to unity:

$$(73) \qquad \lim_{s\to\infty} P\left\{ \prod_{i=1}^{\nu} \left(\frac{|R_{si}|}{\rho} < \epsilon \right) \right\} = 1$$

uniformly in $t_1, t_2, \cdots, t_\nu$ within the region $0 \leq \rho < \rho_0$.

The familiar reasoning that leads to the conclusion that the maximum-likelihood estimates are consistent in the case of consistent observations shows that, if the functions F_{si} satisfy conditions (a), (b), and (c), then there will be a system of solutions $t_{si}, \cdots, t_{s\nu}$ of the equations

$$(74) \qquad F_{si}(X_s \mid t_{s1}, t_{s2}, \cdots, t_{s\nu}) = 0 \qquad \text{for } i = 1, 2, \cdots, \nu$$

and that each implicit function $t_{sk}(X_s)$ will be a consistent estimate of the structural parameter Θ_k. Pursuing the same analogy we may consider the case where, in addition to (a), (b), and (c), the functions F_{si} possess the properties:

(d) The expectations $\mathcal{E}(F_{si}^2)$ exist and, as $s \to \infty$, each ratio $F_{si}(X_s | \Theta)\sqrt{\mathcal{E}[F_{si}^2]}$ tends to be normally distributed about zero.

(e) For each combination of numbers k, $j = 1, 2, \cdots, \nu$, there exists a sequence of numbers $\{B_{skj}\}$ $(s = 1, 2, \cdots)$ such that all the differences

$$\frac{\Delta_{skj}}{\Delta_s} - B_{skj} \tag{75}$$

tend in probability to zero.

Under conditions (a) through (e) the solutions t_{sk} of the system of ν equations (68) will have the property that, as $s \to \infty$, the joint distribution of

$$\frac{(t_{sk} - \Theta)}{\sigma_{sk}} \qquad (k = 1, 2, \cdots, \nu) \tag{76}$$

with

$$\sigma_{sk}^2 = \sum_{i=1}^{\nu} \sum_{j=1}^{\nu} B_{ski} B_{skj} \mathcal{E}(F_{si} F_{sj}) \tag{77}$$

will tend to the ν-variate normal law with unit variances. In other words, σ_{sk}^2 of (77) will represent the asymptotic variance of t_{sk}.

Should it be possible to determine effectively a class C of groups of functions F_{si}, satisfying conditions (a) through (e), then the next problem to consider would be that of finding that particular system of functions of class C for which the variances (77) are the smallest.

Unfortunately, thus far there is no systematic method known to the authors to determine a broad class of functions satisfying condition (a).

In these circumstances, one might consider abandoning a considerable amount of generality and limiting oneself to special forms of the functions F_{si} to which a sufficiently general theorem guaranteeing that each F_{si} tends to zero in probability may be applicable. The theorem in question is, of course, the central limit theorem. Thus we may consider ν systems of functions $\Phi_{ij}(x_i, \Theta)$, for $j = 1, 2, \cdots, \nu$, each system depending on just one variable x_i of the sequence $\{x_n\}$ and on the structural parameters Θ, but independent of the incidental parameters, such that the expectation

$$\mathcal{E}[\Phi_{ij}(x_i, \Theta)] \equiv 0 \; (i = 1, 2, \cdots, s, \cdots; j = 1, 2, \cdots, \nu), \tag{78}$$

the identity relating to all parameters Θ and ξ_i. Then the weighted mean of s functions Φ_{ij} may be considered as defining one function

(79)
$$F_{sk} = \frac{1}{\sum_{i=1}^{s} w_{ik}} \sum_{i=1}^{s} w_{ik} \Phi_{ik}(x_i, \Theta).$$

A proper selection of the weights w_{ik} may assure compliance with conditions (a) through (e) and, even, reduce the asymptotic variance of the estimate to a minimum.

Unfortunately, there are cases where this particular method does not work. Consider the distribution $p_i(x_i|\Theta, \xi_i)$ of one variable x_i of the sequence $\{x_n\}$, and let $\xi_{i1}, \xi_{i2}, \cdots, \xi_{in}, \cdots$ be a denumerable sequence of particular values of the incidental parameter ξ_i. For example, these may be all rational values of the particular parameter. Condition (72) implies that

(80)
$$\int_{-\infty}^{\infty} \Phi_{ik}(x_i, \Theta) p_i(x_i \mid \Theta, \xi_{ij}) dx_i = 0$$

for all values of $j=1, 2, \cdots, n, \cdots$. In other words, the function $\Phi_{ik}(x_i, \Theta)$ must be orthogonal to each function of the infinite sequence $\{p_i(x_i|\Theta, \xi_{ij}\}$ $(j=1, 2, \cdots)$. Usually the function Φ_{ik} will be sought within some particular class of functions, such as continuous functions, functions integrable with their squares, etc. It is known that certain sequences of functions p_i may be "closed" within a given class, meaning that every function of the class that is orthogonal to every function of the sequence is necessarily equal to zero. It follows that cases are possible, and indeed there are examples known, where the only function satisfying (78) is zero and, of course, could not be used for estimating the Θ's.

The difficulty and, at the same time, the particular interest of the situation is emphasized by the fact that, thus far, there does not seem to exist a systematic method of solving the following problem underlying the method under consideration:

Given a function $p(x|\xi)$ defined and nonnegative for every point x in the n-dimensional Euclidean space R_n and for $a \leqq \xi \leqq b$ such that

(81)
$$\int_{R_n} p(x \mid \xi) dx \underset{\xi}{\equiv} 1,$$

determine all such functions $\Phi(x)$, independent of ξ (and probably subject to some reasonable limitations), for which

(82)
$$\int_{R_n} \Phi(x) p(x \mid \xi) dx \underset{\xi}{\equiv} 0.$$

It is interesting to note that, for certain categories of distributions

$p(x|\xi)$, the existence of functions $\Phi(x)$ satisfying (82) (that is, apart from the trivial $\Phi \equiv 0$) depends on the number of dimensions n. For example, if x stands for a system of n independent normal variables all having the same expectation ξ and the same variance $\sigma=$const., it is known that if $n=1$, the only function Φ satisfying (82) is equal to zero for almost all x. This was proved by Wald [9]. On the other hand, if $n \geqq 2$, there is a great variety of functions Φ satisfying (82).

In the absence of a systematic method of determining functions Φ satisfying (80) the authors can offer the following procedure, which appears to work in a number of important cases.

5. MODIFIED EQUATIONS OF MAXIMUM LIKELIHOOD

Let $p_i(x_i|\Theta, \xi_i)$ stand for the probability density of the ith multivariate random variable of the sequence. Let ν be the number of structural parameters $\Theta_1, \Theta_2, \cdots, \Theta_\nu$ and let p_i depend on just one incidental parameter ξ_i that does not appear in p_k if $k \neq i$.

We will assume that the probability density functions p_i possess continuous partial derivatives with respect to all the parameters up to the second order and allow differentiation under the integral taken over the whole space of the corresponding multivariate random variable x_i. Write

$$\phi_{ik}(x_i \mid \Theta, \xi_i) = \frac{\partial \log p_i}{\partial \theta_k} \qquad (k = 1, 2, \cdots, \nu) \tag{83}$$

and

$$\omega_i(x_i \mid \Theta, \xi_i) = \frac{\partial \log p_i}{\partial \xi_i} \qquad (i = 1, 2, \cdots, s). \tag{84}$$

Substitute real variables t_k for Θ_k and X_i for ξ_i and write the equations

$$\omega_i(x_i \mid T, X_i) = 0, \tag{85}$$

where T stands for $t_1, t_2, \cdots, t_\nu$. Solve each equation (85) for X_i and substitute the solution, say

$$X_i = f_i(x_i, T) \tag{86}$$

into (83), obtaining

$$\phi_{ik}[x_i \mid T, f_i(x_i, T)]. \tag{87}$$

The function thus obtained depends on x_i and on T but not on ξ_i or X_i.

Let

$$E_{ik} = \mathcal{E}\{\phi_{ik}[x_i \mid T, f_i(x_i, T)]\} \tag{88}$$

be the expectation of (87). There is no guarantee that E_{ik} will be independent of ξ_i. However, in a number of important cases E_{ik} is either a constant or else depends on the structural parameters only. In this latter case write $E_{ik}(\Theta)$ for (88) and build the function, say

$$\Phi_{ik}(x_i, T) = \phi_{ik}(x_i \mid T, f_i(x_i, T)) - E_{ik}(T). \tag{89}$$

Next, taking arbitrary weights w_{ik}, consider the weighted mean

$$F_{sk} = \frac{\sum_{i=1}^{s} w_{ik}\Phi_{ik}(x_i, T)}{\sum_{i=1}^{s} w_{ik}} \tag{90}$$

and see whether or not, by a judicious selection of the weights w_{ik}, it is possible to satisfy the conditions (a) through (e) of the preceding section. The whole procedure was illustrated when we dealt in detail with Example (1). If the functions F_{sk} satisfy the conditions stated, then the system of equations

$$F_{sk} = 0 \qquad (k = 1, 2, \cdots, \nu) \tag{91}$$

will define a set of consistent estimates t_{sk} of the structural parameters Θ_k, which will be asymptotically normal and will have asymptotic variances given by (77). However, as already mentioned, an investigation may reveal that E_{ik} depends on ξ_i and, indeed, that no desirable function exists for which the expectation is independent of ξ_i.

In the favorable case, i.e., when E_{ik} is independent of ξ_i, the ν equations (91) may be called the *modified* maximum-likelihood equations for estimating the Θ's. In fact, equations (91) reduce to ordinary maximum-likelihood equations when we put $w_{ik}=1$ and substitute zero for each E_{ik}. Should the expectations E_{ik} be different from zero, then their omission in (91) may lead to inconsistencies in the estimates obtained from these equations. This possibility was illustrated above. Also, one of the examples already discussed shows that the inclusion and appropriate adjustments of the weights w_{ik} may occasionally lead to a decrease in the asymptotic variances of the estimates.

6. LOWER BOUND OF THE ASYMPTOTIC VARIANCE

Assume that the consistent estimate, t_{sk}, of Θ_k, obtained from the system of equations (91) is unbiased, i.e., such that its expectation is identically equal to Θ_k. In this case the Cramér-Rao inequality [1], [8], [10], gives the lower bound, $\sigma_0^2(s, k)$, of the variance σ_{sk}^2 of t_{sk}. Let

$$\begin{aligned} \lambda_{ijk} &= \lambda_{ikj} = \mathcal{E}(\phi_{ij}\phi_{ik}), \\ \mu_{ij} &= \mathcal{E}(\phi_{ij}\omega_i), \\ \nu_i &= \mathcal{E}(\omega_i^2), \end{aligned} \tag{92}$$

$$(93)\qquad \Lambda_s' = \begin{vmatrix} \sum_{i=1}^{s} \lambda_{i11}, & \sum_{i=1}^{s} \lambda_{i12}, & \cdots, & \sum_{i=1}^{s} \lambda_{i1\nu}, & \mu_{11}, & \mu_{21}, & \cdots, & \mu_{s1} \\ \sum_{i=1}^{s} \lambda_{i21}, & \sum_{i=1}^{s} \lambda_{i22}, & \cdots, & \sum_{i=1}^{s} \lambda_{i2\nu}, & \mu_{12}, & \mu_{22}, & \cdots, & \mu_{s2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sum_{i=1}^{s} \lambda_{i\nu 1}, & \sum_{i=1}^{s} \lambda_{i\nu 2}, & \cdots, & \sum_{i=1}^{s} \lambda_{i\nu\nu}, & \mu_{1\nu}, & \mu_{2\nu}, & \cdots, & \mu_{s\nu} \\ \mu_{11}, & \mu_{12}, & \cdots, & \mu_{1\nu}, & \nu_1, & 0, & \cdots, & 0 \\ \mu_{21}, & \mu_{22}, & \cdots, & \mu_{2\nu}, & 0, & \nu_2, & \cdots, & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mu_{s1}, & \mu_{s2}, & \cdots, & \mu_{s\nu}, & 0, & 0, & \cdots, & \nu_s \end{vmatrix}.$$

Finally, let Λ'_{skl} be the cofactor of the $(k,\cdot l)$ term of Λ_s'. With this notation the lower bound σ_0^2 of the variance of t_{sk} is given by

$$(94)\qquad \sigma_0^2(s, k) = \frac{\Lambda_{skk}'}{\Lambda_s'}\cdot$$

Easy algebra gives

$$(95)\qquad \Lambda_s' = \Lambda_s \prod_{i=1}^{s} \nu_i,$$

$$(96)\qquad \Lambda_{skk}' = \Lambda_{skk} \prod_{i=1}^{s} \nu_i \quad \text{for} \quad k \leqq \nu,$$

where Λ_s is a determinant of νth order whose (kl) element is

$$(97)\qquad \sum_{i=1}^{s} \left(\lambda_{ikl} - \frac{\mu_{ik}\mu_{il}}{\nu_i} \right)$$

and Λ_{skk} its appropriate minor. Thus, finally, the variance σ_{sk}^2 of any unbiased estimated of Θ_k must satisfy the inequality

$$(98)\qquad \sigma_{sk}^2 \geqq \sigma_0^2(s, k) = \Lambda_{skk}/\Lambda_s.$$

Generally, the estimates obtained from equations (91) will not be unbiased, but subject to a milder restriction of consistency. However their asymptotic variance given by (77) ordinarily remains bounded from below by (98).

This is a direct consequence of the following Lemma which generalizes somewhat the Cramér-Rao inequality. The proof applies to random variables satisfying the conditions enumerated in the Lemma and is not restricted to partially consistent systems.

Let one letter X stand for a system of random variables whose probability density function is

$$p(X, \Theta, \xi) = p(X \mid \Theta_1, \cdots, \Theta_\nu, \xi_1, \cdots, \xi_s) \tag{99}$$

depending on $\nu + s$ parameters, $\Theta_1, \Theta_2, \cdots, \Theta_\nu, \xi_1, \cdots, \xi_s$. Although X stands for a system of several random variables the integral of $p(X, \Theta, \xi)$ taken over the whole space, with respect to all the variables will be indicated by

$$\int p(X, \Theta, \xi) dX. \tag{100}$$

Obviously the value of this integral is identically equal to one.

Let $F_i = F_i(X, \Theta)$ $(i = 1, 2, \cdots, \nu)$ be a system of functions of X and $\Theta_1, \Theta_2, \cdots, \Theta_\nu$, independent of $\xi_1, \xi_2, \cdots, \xi_s$ and differentiable with respect to the Θ's. Let

$$F_{ik} = \frac{\partial F_i}{\partial \Theta_k} \qquad (i, k = 1, 2, \cdots, \nu). \tag{101}$$

LEMMA. (1). If $p(X, \Theta, \xi)$ *is differentiable with respect to* $\Theta_{1j}, \cdots, \Theta_\nu$ $\xi_1 \cdots, \xi_s$;

(2) *If the integral* (100) *admits of differentiation under the integral sign;*

(3) *If the expectations of the squares of logarithmic derivatives, say*

$$\phi_i = \frac{\partial \log p}{\partial \Theta_i} \qquad (i = 1, 2, \cdots, \nu) \tag{102}$$

and, say

$$\phi_{\nu+j} = \frac{\partial \log p}{\partial \xi_j} \qquad (j = 1, 2, \cdots, s) \tag{103}$$

all exist, so that

$$\lambda_{ij} = \int \phi_i \phi_j p(X, \Theta, \xi) dX \tag{104}$$

has meaning for $i, j = 1, 2, \cdots, \nu + s$;

(4) *If the expectations of the functions* F_i *exist and are all equal to zero, identically in the* Θ*'s and* ξ*'s, so that*

$$\int F_i p(X, \Theta, \xi) dX \equiv 0, \qquad (i = 1, 2, \cdots, \nu); \tag{105}$$

(5) *If* (105) *can be differentiated under the sign of the integral with respect to the* Θ*'s and* ξ*'s;*

(6) *If the expectations of the derivatives* F_{ik} *exist*

$$\int F_{ik}p(X, \Theta, \xi)dX = g_{ik}, \text{ (say)}; \tag{106}$$

(7) *If the determinant*

$$G = \left| g_{ik} \right|_{i,k=1,2,\cdots,\nu} \neq 0; \tag{107}$$

Then for every $k=1, 2, \cdots, \nu$

$$L(k) = \frac{1}{G^2} \sum_{i=1}^{\nu} \sum_{j=1}^{\nu} G_{ik}G_{jk}\mathcal{E}(F_iF_j) \geqq \frac{\Lambda_{kk}}{\Lambda}, \tag{108}$$

where G_{ik} *denotes the* (ik) *minor of* G, *where* Λ *denotes the determinant*

$$\Lambda = \left| \lambda_{ij} \right|_{i,j=1,2,\cdots,\nu+s}, \tag{109}$$

and Λ_{kk} *denotes the* (kk) *minor of* Λ.

Remark. In order to perceive the connection between this Lemma and the problem of the lower bound σ_0^2 of the variance of a consistent estimate t_{sk} of a structural parameter Θ_k, notice that, in cases described in the previous section, the asymptotic variance of t_{sk}, formula (77), has exactly the form of the left-hand side of (108). If, as frequently occurs, the functions F_{si} in the equation (74) are arithmetic means of type (90), then as $s \to \infty$, the derivatives F_{sik} will tend in probability to these expectations and then the quantity B_{skj} of (75) and (77) will correspond to the ratio G_{kj}/G of (108). Of course, the determinant Λ of (109) is perfectly analogous to Λ' of (95).

Proof of Lemma. Write, for simplicity,

$$u_i = G_{ik}/G, \tag{110}$$

and notice that

$$\sum_{i=1}^{\nu} u_i g_{ij} = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases} \tag{111}$$

The proof of the Lemma is based on the obvious remark that whatever be $v_1, v_2, \cdots, v_{\nu+s}$, the expectation, say

$$\begin{aligned} I &= I(v_1, v_2, \cdots, v_{\nu+s}) \\ &= \int \left(\sum_{i=1}^{\nu} u_iF_i - \sum_{j=1}^{\nu+s} v_j\phi_j \right)^2 p(X, \Theta, \xi)dX \geqq 0. \end{aligned} \tag{112}$$

The equality $I=0$ holds only when

$$\sum_{i=1}^{\nu} u_iF_i = \sum_{j=1}^{\nu+s} v_j\phi_j \tag{113}$$

for almost all X. Thus (112) will remain nonnegative even when we substitute for $v_1, v_2, \cdots, v_{\nu+s}$ the values $v_1^0, v_2^0, \cdots, v_{\nu+s}^0$ that minimize I.

I may be written in the form

$$I = \int \left(\sum_{i=1}^{\nu} u_i F_i \right)^2 p(X, \Theta, \xi) dX - 2 \int \left(\sum_{i=1}^{\nu} u_i F_i \right) \left(\sum_{j=1}^{\nu+s} v_j \phi_j \right) p(X, \Theta, \xi) dX + \int \left(\sum_{j=1}^{\nu+s} v_j \phi_j \right)^2 p(X, \Theta, \xi) dX \geqq 0. \tag{114}$$

The first term in (114) is equal to $L(k)$ in (108). Now we use assumptions (4), (5), and (6) to simplify the second term in (114). Differentiating (105) with respect to Θ_j we obtain

$$\int (F_{ij} + F_i \phi_j) p(X, \Theta, \xi) dX \equiv 0. \tag{115}$$

Hence

$$\int F_i \phi_j p(X, \Theta, \xi) dX = - g_{ij} \quad \text{for} \quad i, j = 1, 2, \cdots, \nu. \tag{116}$$

Also, differentiating (105) with respect to ξ_j, we obtain

$$\int F_i \phi_j p(X, \Theta, \xi) dX \equiv 0 \text{ for } \begin{cases} i = 1, 2, \cdots, \nu, \\ j = \nu + 1, \nu + 2, \cdots, \nu + s. \end{cases} \tag{117}$$

Using (111) we may now write (114) as

$$I = L(k) + 2 v_k + \sum_{i=1}^{\nu+s} \sum_{j=1}^{\nu+s} v_i v_j \lambda_{ij} \geqq 0. \tag{118}$$

The minimizing values v_j^0 are obtained by setting the derivatives of (118) with respect to v_j equal to zero. The system of linear equations obtained in this way can be written as

$$\begin{aligned} \sum_{i=1}^{\nu+s} v_i^0 \lambda_{ij} &= -1 \quad \text{for} \quad j = k, \\ \sum_{i=1}^{\nu+s} v_i^0 \lambda_{ij} &= 0 \qquad \text{for} \quad j \neq k. \end{aligned} \tag{119}$$

Multiplying (119) by v_j^0 and summing for $j = 1, 2, \cdots, \nu+s$, we obtain

(120) $$\sum_{i=1}^{\nu+s} \sum_{j=1}^{\nu+s} v_i^0 v_j^0 \lambda_{ij} = - v_k^0,$$

and it follows that the minimum, say I^0 of I can be written

(121) $$I^0 = L + v_k^0 \geqq 0.$$

But the equations (119) imply

(122) $$v_i^0 = - \frac{\Lambda_{ik}}{\Lambda} \qquad (i = 1, 2, \cdots, \nu + s).$$

Therefore $L(k) - \Lambda_{kk}/\Lambda \geqq 0$ which proves the Lemma.

Using the above Lemma, we may now formulate a theorem relating to variances of consistent estimates t_{sk} of structural parameters as described in the previous section.

THEOREM 1. *If the functions $F_{si}(X_s | \Theta)$ of* (66) *satisfy the conditions* (a), (b), (c), (d) *and if, as $s \to \infty$, the differences $[F_{sik}(X_s | \Theta) - \mathcal{E}(F_{sik})]$ tend to zero in probability, then the asymptotic variance of the estimate t_{sk} obtained from* (74) *of the parameter Θ_k* $((k = 1, 2, \cdots, \nu)$ *is at least equal to the ratio $\sigma_0^2(s, k) = \Lambda_{kk}/\Lambda$.*

7. CASES OF IMPAIRED AND UNIMPAIRED ASYMPTOTIC EFFICIENCY

As is well known, if the observable random variables are consistent and certain mild restrictions are satisfied, then the lower bound provided by the Cramér-Rao inequality is always attained by the asymptotic variances of the consistent estimates of the parameters. For example, the maximum-likelihood estimates possess this property.

It is interesting to note that in cases of partial consistency of the observable variables, this need not be so and, whatever the functions F_{sk} used to compute the consistent estimates, the asymptotic variances σ_{sk}^2 may always be greater than the limit (98). Here we have to distinguish three cases. First it may happen that not only $\sigma_{sk}^2 > \sigma_0^2(s, k)$ but also the upper limit as $s \to \infty$ of the ratio $\sigma_{sk}^2/\sigma_0^2(s, k)$ may be greater than unity. In cases of this kind we shall say that the asymptotic efficiency is *essentially impaired* by the presence of the incidental parameters. The second case is that where the inequality $\sigma_{sk}^2 > \sigma_0^2(s, k)$ is combined with the equality

(123) $$\lim_{s \to \infty} \frac{\sigma_{sk}^2}{\sigma_0^2(s, k)} = 1.$$

Here we shall say that the asymptotic efficiency is *impaired inessentially*. Finally, it may happen that for all values of $s = 1, 2, \cdots$ we have $\sigma_{sk}^2 = \sigma_0^2(s, k)$. This will be the case of *unimpaired efficiency*.

THEOREM 2. *If the consistent estimates t_{sk} of structural parameters Θ_k $(k = 1, 2, \cdots, \nu)$ are obtained from equations $F_{sk}(X | T) = 0$, where* *the functions F_{sk} satisfy the conditions of Theorem 1, then for t_{sk} to have unimpaired efficiency, it is necessary and sufficient that*

$$\frac{1}{G}\sum_{i=1}^{\nu} G_{ik}F_{si} = -\frac{1}{\Lambda'}\sum_{j=1}^{\nu+s} \Lambda_{jk}'\phi_j \tag{124}$$

for almost all systems of values of the observable random variables.

The proof of Theorem 2 follows immediately from the remark made in the course of proving the Lemma of the preceding section. This is to the effect that, for the value of I of (112) to be equal to zero it is both necessary and sufficient that the equality (113) be satisfied almost everywhere.

THEOREM 3. *For the existence of a system of ν equations $F_{si}(X|T)=0$ $(i=1, 2, \cdots, \nu)$ satisfying the conditions of Theorem 1, whose solutions with respect to T yield the consistent estimate t_{sm} of the parameter Θ_m of unimpaired efficiency, it is necessary and sufficient* (a) *that for almost all X_s it is possible to represent the expression*

$$\sum_{j=1}^{\nu+s} \Lambda_{jm}'\phi_j \tag{125}$$

by a sum of $\nu' \leqq \nu$ terms

$$\sum_{j=1}^{\nu+s} \Lambda_{jm}'\phi_j = \sum_{i=1}^{\nu'} A_i\Psi_i(X_s, \Theta) \tag{126}$$

where the A_i are independent of X_s and where the Ψ_i are functions of the observable variables and the structural parameters but are independent of the incidental parameters and (b) *that $\nu-\nu'$ functions $\Psi_{\nu'+j}(X_s, \Theta)$ can be found such that the system of Ψ_j $(j=1, 2, \cdots, \nu)$ satisfies the conditions of Theorem 1. Then the unimpaired estimate of Θ_m will be obtained from the system of equations $\Psi_j(X_s, T)=0$ $(j=1, 2, \cdots, \nu)$.*

The necessity of the conditions enumerated follows directly from Theorem 2. In fact, if the system of equations $F_{si}(X_s|T)=0$ $(i=1, 2, \cdots, \nu)$ satisfying the conditions of Theorem 1 and yielding an unimpaired estimate of Θ_m exists, then (124) must be satisfied for almost all X_s and thus it must be possible to represent (125) in the form (126) with the function Ψ_i coinciding with the functions $F_{si}(X_s|\Theta)$. In order to prove the sufficiency of the conditions, assume that the system of functions $\Psi_j(j=1, 2, \cdots, \nu)$ exists, satisfying the conditions of Theorem 3. Then

$$\sum_{j=1}^{\nu+s} \Lambda_{jm}'\phi_j = \sum_{j=1}^{\nu} A_j\Psi_j(X_s, \Theta), \tag{127}$$

where some of the A_j may be equal to zero. Let h_{ij} be defined in relation to the functions Ψ_i in the same way as g_{ij} was defined in relation to the function F_i of Theorem 2. Thus

$$h_{ij} = -\int \Psi_i \phi_j p(X_s, \Theta, \xi) dX \qquad (i, j = 1, 2, \cdots, \nu) \tag{128}$$

and further

$$H = | h_{ij} |_{i,j=1,2,\cdots,\nu}, \tag{129}$$

with H_{ij} denoting the minor of the determinant H. It is easy to see that (127) implies

$$A_j = -\frac{\Lambda_s}{H} H_{jm}. \tag{130}$$

In fact, multiplying (127) in turn by ϕ_k ($k=1, 2, \cdots, \nu$) and taking the expectations, we obtain

$$\sum_{j=1}^{\nu} A_j h_{jk} = -\sum_{j=1}^{\nu+s} \Lambda_{sjm}' \lambda_{jk} = 0, \quad k \neq m, \tag{131}$$

and

$$\sum_{j=1}^{\nu} A_j h_{jm} = -\sum_{j=1}^{\nu+s} \Lambda_{sjm}' \lambda_{jm} = -\Lambda_s', \tag{132}$$

and it is seen that the coefficients A_j must have the form (130). But then, by Theorem 2, the asymptotic variance of the estimate of t_{sm} derived from the equation $\Psi_j=0$ must be equal to the lower bound provided by the Cramér-Rao inequality.

Theorem 3 provides easy means of determining equations yielding unimpaired estimates of the structural parameters, if they exist. The procedure is as follows. Write down in a column the expressions of the logarithmic derivatives of the joint distribution of all the variables forming a partially consistent system. Next to this column write the matrix of the Λ_{skl}' for $k=1, 2, \cdots, \nu+s$ and $l=1, 2, \cdots, \nu$. Referring to notation (83), (84), (94), and (95) we have

$$\begin{array}{c|llll|} \sum_{i=1}^{s} \phi_{i1} & \Lambda_{s11}', & \Lambda_{s12}', & \cdots, & \Lambda_{s1\nu}' \\ \sum_{i=1}^{s} \phi_{i2} & \Lambda_{s21}', & \Lambda_{s22}', & \cdots, & \Lambda_{s2\nu}' \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{s} \phi_{i\nu} & \Lambda_{s\nu1}', & \Lambda_{s\nu2}', & \cdots, & \Lambda_{s\nu\nu}' \\ \omega_1 & \Lambda_{s,\nu+1,1}', & \Lambda_{s,\nu+1,2}', & \cdots, & \Lambda_{s,\nu+1,\nu}' \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \omega_s & \Lambda_{s,\nu+s,1}', & \Lambda_{s,\nu+s,2}', & \cdots, & \Lambda_{s,\nu+s,\nu}' \end{array} . \tag{133}$$

Next form the sum of products of the logarithmic derivatives by the corresponding terms of the first column of the matrix. If this sum splits into not more than ν terms of the form $A_j\Psi_j$ where the functions Ψ_j satisfy the conditions of Theorem 3, then it is possible to obtain an estimate of Θ_1 having unimpaired efficiency. Similar trials with columns 2, 3, $\cdots$, ν of the matrix will answer the same question relating to estimates of Θ_2, Θ_2, $\cdots$, Θ_ν.

It is possible that the sum of products, say

$$\sum_{k=1}^{\nu} \Lambda_{skm}' \sum_{i=1}^{s} \phi_{ik} = \sum_{j=1}^{\nu'} A_j\Psi_j \tag{134}$$

with $\nu' < \nu$. Then, even if the ν' functions Ψ_j satisfy the necessary conditions, the problem appears of supplying an additional $\nu - \nu'$ functions Ψ_j such that the whole system of ν equations $\Psi(X_s, T) = 0$ yields consistent estimates of all the structural parameters. Theorem 3 guarantees that, in whatever way these $\nu - \nu'$ additional functions are selected, provided the whole system satisfies the conditions of Theorem 1, the resulting estimate of Θ_m will have unimpaired efficiency. This circumstance indicates the desirability of a system that satisfies condition (134) for as many different values of m as possible. Of course, the source of the additional $\nu - \nu'$ equations is the system of modified maximum-likelihood equations described above. It is interesting that cases exist where one of the several structural parameters can be estimated with unimpaired efficiency, but not the others. A case of this kind is discussed below.

8. ILLUSTRATIONS

The present paper originated from the work of one of the authors, relating to the problem of fitting a straight line when both variables are subject to error. The general statement of this problem is given in Section 2 as Example (3) and the results of the complete solution will be published elsewhere. Further below a particular case of the problem is considered in some detail. The purpose of this analysis is to illustrate some of the points of the preceding section, namely, (i) the process of determining equations yielding an unimpaired estimate of one of the structural parameters and (ii) the fact that, on occasion, one of the several structural parameters may admit an unimpaired estimate, but not the others.

In order to illustrate these facts, a special case of the problem of fitting a straight line is chosen for consideration in this paper. This case is somewhat artificial and of minor practical importance. This, however, does not diminish its illustrative value.

The case considered is that where, in the notation adopted in the description of Example (3), Section 2, $m_i = m$ and $n_i = n$ for $i = 1$, 2, $\cdots$, s and $\sigma_1=\sigma_2=\sigma$. It is assumed that σ^2 and α are unknown structural parameters, but that the coefficient β is a known number.

In attempting to obtain estimates of σ^2 and α we follow the steps outlined above.

First step. Write down the logarithmic derivatives of the probability law (7), with respect to the parameters σ^2, α, and ξ_i:

$$
\begin{aligned}
\phi_{i1} &= \frac{\partial \log p_i}{\partial \sigma^2} = -\frac{m+n}{2\sigma^2} \\
&\quad + \frac{1}{2\sigma^4}\left(\sum_{j=1}^{m}(x_{ij}-\xi_i)^2 + \sum_{j=1}^{n}(y_{ij}-\alpha-\beta\xi_i)^2\right); \\
\phi_{i2} &= \frac{\partial \log p_i}{\partial \alpha} = \frac{n}{\sigma^2}(y_{i\cdot}-\alpha-\beta\xi_i), \\
\omega_i &= \frac{\partial \log p_i}{\partial \xi_i} = \frac{1}{\sigma^2}\left\{m(x_{i\cdot}-\xi_i)+n\beta(y_{i\cdot}-\alpha-\beta\xi_i)\right\}.
\end{aligned}
\tag{135}
$$

Here $x_{i\cdot}$ and $y_{i\cdot}$ are the arithmetic means of the x_{ij} and the y_{ij} respectively.

Second step. Compute the expectations

$$
\lambda_{ijk} = \mathcal{E}(\phi_{ij}\phi_{ik}), \qquad \mu_{ij} = \mathcal{E}(\phi_{ij}\omega_i), \qquad \nu_i = \mathcal{E}(\omega_i^2) \tag{136}
$$

for $j=1$, 2 and $i=1$, 2, $\cdots$, s. Simple algebra gives

$$
\begin{aligned}
\lambda_{i11} &= \frac{m+n}{2\sigma^4}, & \lambda_{i12} &= \lambda_{i21} = 0, & \lambda_{i22} &= \frac{n}{\sigma^2}, \\
\mu_{i1} &= 0, & \mu_{i2} &= \frac{n\beta}{\sigma^2}, & \nu_i &= \frac{m+n\beta^2}{\sigma^2}.
\end{aligned}
\tag{137}
$$

Third step. Write the determinant Λ_s' of (95)

$$
\begin{vmatrix}
s\dfrac{m+n}{2\sigma^2} & 0 & 0 & 0 & \cdots & 0 \\
0 & \dfrac{sn}{\sigma^2} & \dfrac{n\beta}{\sigma^2} & \dfrac{n\beta}{\sigma^2} & \cdots & \dfrac{n\beta}{\sigma^2} \\
0 & \dfrac{n\beta}{\sigma^2} & \dfrac{m+n\beta}{\sigma^2} & 0 & \cdots & 0 \\
0 & \dfrac{n\beta}{\sigma^2} & 0 & \dfrac{m+n\beta}{\sigma^2} & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\
0 & \dfrac{n\beta}{\sigma^2} & 0 & 0 & \cdots & \dfrac{m+n\beta}{\sigma^2}
\end{vmatrix}
\tag{138}
$$

and compute the minors Λ_{sik}'.

Fourth step. Write down the column of the logarithmic derivatives of the joint distribution of all the observable random variables and the $\nu=2$ column matrix of minors Λ_{sik}'.

$$(139)\quad \begin{array}{c|cc|} \sum_{i=1}^{s} \phi_{i1} & \dfrac{smn}{\sigma^2(m+n\beta^2)} & 0 \\ \sum_{i=1}^{s} \phi_{i2} & 0 & s\dfrac{m+n}{2\sigma^4}\left(\dfrac{m+n\beta^2}{\sigma^2}\right)^s \\ \omega_1 & 0 & -s\dfrac{m+n}{2\sigma^4}\dfrac{n\beta}{\sigma^2}\left(\dfrac{m+n\beta^2}{\sigma^2}\right)^{s-1} \\ \cdots & \cdots & \cdots \\ \omega_s & 0 & -s\dfrac{m+n}{2\sigma^4}\dfrac{n\beta}{\sigma^2}\left(\dfrac{m+n\beta^2}{\sigma^2}\right)^{s-1} \end{array}.$$

Multiply the logarithmic derivatives by the corresponding terms of the first column of the matrix and sum the results.

$$(140)\quad \begin{aligned} \frac{smn}{\sigma^2(m+n\beta^2)}\sum_{i=1}^{s}\phi_{i1} = \frac{smn}{\sigma^2(m+n\beta^2)}\Bigg\{&-\frac{s(m+n)}{2\sigma^2} \\ &+\frac{1}{2\sigma^2}\sum_{i=1}^{s}\left[\sum_{j=1}^{m}(x_{ij}-\xi_i)^2+\sum_{j=1}^{n}(y_{ij}-\alpha-\beta\xi_i)^2\right]\Bigg\}. \end{aligned}$$

Multiply the logorithmic derivatives by the second column of the matrix and sum the results.

$$(141)\quad \begin{aligned} &s\frac{m+n}{2\sigma^4}\left(\frac{m+n\beta^2}{\sigma^2}\right)^s\sum_{i=1}^{s}\phi_{i2} \\ &\quad - s\frac{m+n}{2\sigma^4}\frac{n\beta}{\sigma^2}\left(\frac{m+n\beta^2}{\sigma^2}\right)^{s-1}\sum_{i=1}^{s}\omega_i = A(y_{..}-\alpha-\beta x_{..}), \end{aligned}$$

where A is a coefficient depending on σ but independent of the random variables x_{ij} and y_{ij} and where $x_{..}$ and $y_{..}$ are the arithmetic means of all the x_{ij} and all the y_{ij} respectively. It is seen that expression (140) does not split into a sum of type (126) of at most two components. On the other hand, expression (141) is itself a product of a function of the observable variables and of just one unknown structural parameter, namely $y_{..}-\alpha-\beta x_{..}$, by a coefficient A, which does not depend on the random variables. Moreover, the function $y_{..}-\alpha-\beta x_{..}$ has its expectation equal to zero, as $s\to\infty$ it tends to zero in probability and its derivative with respect to α is a constant. The conclusions are: (i) there is no system of equations satisfying the conditions of Theorem 1 which yield an unimpaired estimate of σ^2; (ii) an unimpaired estimate, say α^* of the parameter α is possible, provided we use for estimation the equation

$$y_{..} - \alpha^* - \beta x_{..} = 0 \tag{142}$$

which gives immediately

$$\alpha^* = y_{..} - \beta x_{..}. \tag{143}$$

University of California
Berkeley

REFERENCES

[1] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[2] Abraham Wald, "The Fitting of Straight Lines if Both Variables Are Subject to Error," *Annals of Mathematical Statistics*, Vol. 11, September, 1940, pp. 284–300.
[3] Frederick H. Seares, "Regression Lines and the Functional Relation," *Astrophysical Journal*, Vol. 100, 1944, pp. 255–263.
[4] Frederick H. Seares and Mary C. Joyner, "Relation between Color Index and Effective Wave Length from the Observations of Hertzsprung and Vanderlinden," *Astrophysical Journal*, Vol. 100, 1944, pp. 267–278.
[5] Frederick H. Seares, "Regression Lines and the Functional Relation, II," *Astrophysical Journal*, Vol. 102, 1945, pp. 366–376.
[6] Ragnar Frisch, *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo, 1934.
[7] W. G. Cochran, "Problems Arising in the Analysis of a Series of Similar Experiments," *Journal of the Royal Statistical Society, Supplement*, Vol. 4, 1937, pp. 102–118.
[8] C. Radhakrishna Rao, "Information and Accuracy Attainable in One Estimation of Statistical Parameters," *Bulletin of the Calcutta Mtahematical Society*, Vol. 37, 1945, pp. 81–91.
[9] Abraham Wald, "Note on a Lemma," *Annals of Mathematical Statistics*, Vol. 15, September, 1944, pp. 330–333.
[10] Harald Cramér, "A Contribution to the Theory of Statistical Estimation," *Skandinavisk Aktuarietidskrift*, 1946, pp. 85–94.