statistic for W (cf. [5], p. 232, §5). Let f be an unbiased estimate of the function g^2 on Θ. For each $\mu_\theta \in W$, the conditional expectation, $E_\theta(f \mid \cdot)$, of f with respect to t is defined. Since conditional expectations are fully determined by conditional probabilities (although, in general, not as usual integrals. Cf. [4], pp. 48, 49; also [5], p. 230) it follows from the sufficiency of t that there exists a function $E(f \mid \cdot)$, on Γ, with $E_\theta(f \mid \tau) = E(f \mid \tau)$ a.e. (ν_θ) for each $\theta \in \Theta \cdot E^*(f \mid \cdot)$ is again an unbiased estimate of g, and we have

COROLLARY 2. *Let t be a sufficient statistic for the family $W = \{\mu_\theta, \theta \in \Theta\}$; and f, an unbiased estimate of g. For $s \geqq 1$, and each $\theta \in \Theta$,*

$$\int_\Omega | E^*(f \mid \cdot) - g(\theta) |^s \, d\mu_\theta \leqq \int_\Omega | f - g(\theta) |^s \, d\mu_\theta .$$

Equality holds

(i) *for $s = 1$, if and only if* sgn $[f - g(\theta)]$ *is essentially (μ_θ) a function of t;*

(ii) *for $s > 1$, if and only if f is essentially (μ_θ) a function of t.*

REFERENCES

[1] E. W. BARANKIN, "Locally best unbiased estimates", *Annals of Math. Stat.*, Vol. 20 (1949), pp. 477–501.

[2] D. BLACKWELL, "Conditional expectation and unbiased sequential estimation", *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[3] L. M. GRAVES, *The Theory of Functions of Real Variables*, McGraw-Hill, 1946.

[4] A. KOLMOGOROFF, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Vol. 2 (1933).

[5] P. R. HALMOS AND L. J. SAVAGE, "Application of the Radon-Nikodym Theorem to the theory of sufficient statistics", *Annals of Math. Stat.*, Vol. 20 (1949), pp. 225–241.

NOTE ON CONSISTENT ESTIMATES OF THE LINEAR STRUCTURAL RELATION BETWEEN TWO VARIABLES[1]

BY ELIZABETH L. SCOTT

University of California, Berkeley

1. Introduction. The purpose of this note is to present another case in which the structural linear relation between two observable random variables may be consistently estimated. Of the recent papers on this subject I wish to mention the paper by Wald [1], which contains a history of the work done on the problem, and the more recent paper by Housner and Brennan [2]. Also relevant is the important result due to Reiersøl [3], [4].

2. Statement of problem. Assume that the two observable random variables x and y have the structure

[1] Paper prepared with partial support of the Office of Naval Research.

The results summarized were presented in a discussion held at the Cleveland Meeting of the Institute of Mathematical Statistics, December, 1948.

$$\begin{cases} x = \xi + u \\ y = \alpha + \beta\xi + v, \end{cases} \tag{1}$$

where α and β are unknown parameters to be estimated, and ξ, u and v are completely independent random variables. The latter two variables, interpreted as the random errors of measurement, are assumed to vary normally about zero with unknown variances σ_1^2 and σ_2^2, respectively.

An increasing number n of completely independent pairs of simultaneous values of x and y are to be observed

$$(x_i, y_i), \qquad i = 1, 2, \cdots, n, \tag{2}$$

so that each pair (x_i, y_i) corresponds to a value ξ_i of the unobservable random variable ξ which is independent of the value ξ_j of ξ corresponding to any other pair (x_j, y_j), $i \neq j$.

It is well known that if the distribution of ξ is normal then the parameters α, β, σ_1 and σ_2 are unidentifiable. Reiersøl proved [4] that these parameters are identifiable in all other cases. Wald and Housner and Brennan found consistent estimates of these parameters assuming that, although the particular values of ξ are not known exactly, a certain amount of knowledge concerning the values of ξ is available. The present note gives a method for obtaining a consistent estimate of β, which is the key to the problem of estimating the four parameters, for the case where it is known that a specified central moment of the distribution of ξ exists and differs from that of the normal distribution.

Since work on this subject continues, the present brief note deals particularly with the simplest case, when one of the odd central moments of ξ exists and differs from the "normal" value, zero. It will be observed that the hypotheses made here are of entirely different character from those adopted by other writers. The present note postulates knowledge concerning a moment of the distribution of ξ, whereas the papers quoted postulate some knowledge of the particular values assumed by ξ. The method adopted was suggested by a remark made by Neyman [5] in 1936.

3. Preliminary theorems. Let

$$x_. = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad y_. = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3}$$

and let b be an arbitrary real number.

THEOREM 1: *If μ_3, the third central moment of ξ, exists then the arithmetic mean*

$$F_{n,1}(b) = \frac{1}{n}\sum_{i=1}^{n} [y_i - y_. - b(x_i - x_.)]^3 \tag{4}$$

converges in probability to

$$(\beta - b)^3\mu_3 . \tag{5}$$

PROOF: Simple algebra gives

$$
\begin{aligned}
F_{n,1}(b) &= (\beta - b)^3 \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \xi_.)^3 \\
&+ 3(\beta - b)^2 \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \xi_.)^2 [v_i - v_. - b(u_i - u_.)] \\
&+ 3(\beta - b) \frac{1}{n} \sum_{i=1}^{n} (\xi_i - \xi_.) [v_i - v_. - b(u_i - u_.)]^2 \\
&+ \frac{1}{n} \sum_{i=1}^{n}]v_i - v_. - b(u_i - u_.)]^3.
\end{aligned} \tag{6}
$$

It is obvious that further expansion will express $F_{n,1}(b)$ in terms of averages of the type

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i^p u_i^q v_i^r, \tag{7}$$

with $p + q + r \leqq 3$. Since all the terms over which each average is taken are completely independent, follow the same law and possess finite expectations, the familiar theorem of Khintchine assures that, as n is increased, each average (7) tends in probability to its expectation. Using Slutsky's theorem (see Cramér [6], p. 255), we conclude that $F_{n,1}(b)$ tends in probability to the limit obtained by replacing each average in the expansion (6) by its expectation and then letting $n \to \infty$. The computations are easy and give

$$\lim_{n \to \infty} pF_{n,1}(b) = (\beta - b)^3 \mu_3. \tag{8}$$

Q.E.D.

Let $\{X_n\}$ denote a sequence of observable random variables (multivariate or not) such that the distribution function of X_n depends on the parameters θ_i with $a_i < \theta_i < b_i$, $i = 1, 2, \cdots, s$. Furthermore, let λ denote a real variable and $\{\phi_n(X_n, \lambda)\}$ a sequence of functions of the arguments X_n and λ defined for all possible values of X_n and for all values of λ within the limits $a_1 \leqq \lambda \leqq b_1$.

THEOREM 2: *If the sequence of functions $\{\phi_n(X_n, \lambda)\}$ has the following properties:*

(i) *whatever be the true values $\theta_1', \theta_2', \cdots, \theta_s'$ of the parameters θ_i within the limits $a_i < \theta_i' < b_i$, $i = 1, 2, \cdots, s$, as n is increased, the sequence $\{\phi_n(X_n, \lambda)\}$ tends in probability to a function $f(\lambda, \theta_1')$ of arguments λ and θ_1' only.*

(ii) *whatever be $\delta > 0$, there exist in (a_1, b_1) two numbers λ_1 and λ_2, each differing from θ_1' by less than δ and such that the product $f(\lambda_1, \theta_1') f(\lambda_2, \theta_1')$ is negative,*

(iii) *for every n and every possible value x_n of X_n, the function $\phi_n(x_n, \lambda)$ is continuous with respect to λ for $a_1 \leqq \lambda \leqq b_1$,*

then whatever be $\epsilon > 0$ and $\eta > 0$ there exists a number $N_{\epsilon,\eta}$ such that for $n > N_{\epsilon,\eta}$ the probability that the equation $\phi_n(X_n, \lambda) = 0$ has a root between $\theta_1' - \epsilon$ and $\theta_1' + \epsilon$ exceeds $1 - \eta$.

PROOF: Let $\epsilon > 0$ and $\eta > 0$ be two arbitrarily small numbers. Let λ_1 and λ_2 be two numbers such that $\lambda_i \in (a_1, b_1)$ and $| \theta_1' - \lambda_i | < \epsilon$, $i = 1, 2$, and such that $f(\lambda_1, \theta_1') < 0 < f(\lambda_2, \theta_1')$. Select $N_{\epsilon,\eta}$ so large that for $n > N_{\epsilon,\eta}$ the probability of having simultaneously

$$(9) \qquad |\phi_n(X_n, \lambda_i) - f(\lambda_i, \theta_1')| < \tfrac{1}{2}|f(\lambda_i, \theta_1')| \quad \text{for } i = 1, 2$$

differs from unity by less than η. It is clear that if the inequalities (9) are satisfied for any particular value x_n of X_n, then

$$(10) \qquad \phi_n(x_n, \lambda_1) < 0 < \phi_n(x_n, \lambda_2)$$

and the continuity of $\phi_n(x_n, \lambda)$ for $\lambda \epsilon (a_1, b_1)$ implies that there exists a number $\lambda(x_n)$ between λ_1 and λ_2 such that $\phi_n(x_n, \lambda(x_n)) = 0$. Obviously $|\theta_1' - \lambda(x_n)| < \epsilon$. Thus, whatever be $\epsilon, \eta > 0$, there exists a number $N_{\epsilon,\eta}$ such that the probability that $\phi_n(X_n, \lambda)$ has a root in the interval $(\theta_1' - \epsilon, \theta_1' + \epsilon)$ exceeds $1 - \eta$ provided $n > N_{\epsilon,\eta}$. This proves Theorem 2.

Theorem 2 is treated as a convenient lemma on which to base the proof of the existence of a consistent estimate of the parameter β in (1). It is obvious, however, that this Theorem has an independent interest of its own.

4. Consistent estimates of the structural parameter β. Referring to the general set-up of the problem of estimating the structural parameter β in (1) and using the notation (2) and (3), we prove the following theorems.

THEOREM 3: *If the third central moment μ_3 of ξ exists and differs from zero, then the equation*

$$(11) \qquad F_{n,1}(b) = \frac{1}{n}\sum_{i=1}^{n} [y_i - y. - b(x_i - x.)]^3 = 0$$

has a root $\hat{b}$ which is a consistent estimate of β.

PROOF: According to Theorem 1, whatever be b and β, the stochastic limit of $F_{n,1}(b)$ is $(\beta - b)^3\mu_3$ and changes its sign as b passes through the value β. Theorem 2 implies then that whatever be $\epsilon, \eta > 0$, there exists a number $N_{\epsilon,\eta}$ such that for $n > N_{\epsilon,\eta}$ the probability that at least one of the roots of (11) will lie within $\beta - \epsilon$ and $\beta + \epsilon$ is greater than $1 - \eta$. This proves the theorem.

Generally, let μ_m denote the m^{th} central moment of ξ.

THEOREM 4: *If the distribution of ξ has moments up to and including order $2m + 1$ and if at least one of the first m odd central moments μ_{2k+1} differs from zero, $k = 1, 2, \cdots, m$, then the equation*

$$(12) \qquad F_{n,m}(b) = \frac{1}{n}\sum_{i=1}^{n} [y_i - y. - b(x_i - x.)]^{2m+1} = 0$$

has a root $\hat{b}$ which is a consistent estimate of β.

PROOF: The proof of Theorem 4 exactly follows the lines of that of Theorem 3. Using (1), (2) and (3), we write

$$(13) \qquad F_{n,m}(b) = \sum_{k=0}^{2m+1} C_{2m+1}^{k}(\beta - b)^k \left\{\frac{1}{n}\sum_{i=1}^{n} (\xi_i - \xi.)^k[v_i - v. - b(u_i - u.)]^{2m+1-k}\right\}.$$

It is easily seen that, as $n \to \infty$, $F_{n,m}(b)$ tends in probability to the limit

$$F_{n,m}(b) \underset{n\to\infty}{\xrightarrow[p]{}} (\beta - b)^3\psi(\beta - b), \tag{14}$$

where $\psi(\beta - b)$ is a linear combination of even powers of $(\beta - b)$ with at least one coefficient different from zero. It follows that the stochastic limit of $F_{n,m}(b)$ changes its sign as b passes through β and the proof is completed by reference to Theorem 2.

Note that the stochastic limit of the first derivative of $F_{n,m}(b)$, evaluated at $b = \beta$, is zero, which is unfortunate. Furthermore, the order of contact of $F_{n,m}(b)$ at $b = \beta$ increases with the order of the first odd central moment of ξ which differs from zero. Therefore, the precision of estimating β may be expected to be better when the low odd central moments are not zero. Without narrowing the generality of the case considered, it is difficult to make an evaulation of the precision of the estimates obtained. Thus, for example, the familiar method of evaluating the asymptotic variance requires the knowledge of higher moments of ξ than those considered here. For similar reasons, it is thus far impossible to speak of the relative efficiency of the estimates found. For this purpose it would be necessary to determine first the measure of the precision of the best estimate whose consistency persists independently of the distribution of ξ provided only that at least one odd central moment differs from zero.

Once the consistent estimate $\hat{b}$ of β is obtained, there is no particular difficulty in obtaining consistent estimates of the other parameters.

J. Neyman has pointed out [7] that Theorem 2 may be used as the basis for a very elementary proof of the consistency of maximum likelihood estimates.

REFERENCES

[1] ABRAHAM WALD, "The fitting of straight lines if both variables are subject to error," *Annals of Math. Stat.*, Vol. 11, (1940), p. 284.

[2] G. W. HOUSNER AND J. F. BRENNAN, "The estimation of linear trends," *Annals of Math. Stat.*, Vol. 19 (1948), p. 380.

[3] OLAV REIERSØL, "Confluence analysis by means of lag moments and other methods of confluence analysis," *Econometrica*, Vol. 9 (1941), p. 1.

[4] OLAV REIERSØL, "Identifiability of a linear relation between variables which are subject to error," *Cowles Commission Discussion Papers, Statistics*, No. 337 (1949).

[5] J. NEYMAN, *Jour. Roy. Stat. Soc.*, Vol. 100 (1937), p. 56.

[6] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[7] J. NEYMAN, *First Course in Probability and Statistics*, Vol. 2, forthcoming.