our results can easily be shown to remain valid if Assumptions 3, 5 and 7 are replaced by the following one:

ASSUMPTION 9. *It is possible to introduce a distance* $\delta(\theta_1, \theta_2)$ *in the space* Ω *such that the following four conditions hold*:

(i) *The distance* $\delta(\theta_1, \theta_2)$ *makes* Ω *to a metric space*

(ii) $\lim_{i=\infty} f(x, \theta_i) = f(x, \theta)$ *if* $\lim_{i=\infty} \theta_i = \theta$ *for any* x *except perhaps on a set which may depend on* θ *(but not on the sequence* θ_i*) and whose probability measure is zero according to the probability distribution corresponding to the true parameter point* θ_0 .

(iii) *If* θ_0 *is a fixed point in* Ω *and* $\lim_{i=\infty} \delta(\theta_i, \theta_0) = \infty$, *then* $\lim_{i=\infty} f(x, \theta_i) = 0$ *for any* x.

(iv) *Any closed and bounded subset of* Ω *is compact.*

REFERENCES

[1] J. L. DOOB, "Probability and statistics," *Trans. Amer. Math. Soc.*, Vol. 36 (1934).
[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.
[3] V. S. HUZURBAZAR, "The likelihood equation, consistency and the maxima of the likelihood function," *Annals of Eugenics*, Vol. 14 (1948).

ON WALD'S PROOF OF THE CONSISTENCY OF THE MAXIMUM LIKELIHOOD ESTIMATE

BY J. WOLFOWITZ

Columbia University

This note is written by way of comment on the pretty and ingenious proof of the consistency of the maximum likelihood estimate which is due to Wald and is printed in the present issue of the *Annals*. The notation of this paper of Wald's will henceforth be assumed unless the contrary is specified.

The consistency of the maximum likelihood estimate is a "weak" rather than a "strong" property, in the technical meaning which these words have in the theory of probability, i.e., it is a property of distribution functions rather than of infinite sequences of observations. Prof. Wald actually proves strong convergence, which is more than consistency. His proof uses the strong law of large numbers, and he remarks that his method "can be extended to establish consistency of the maximum likelihood estimates for certain types of dependent chance variables for which the strong law of large numbers remains valid." Below we shall use Wald's lemmas to give a proof of consistency which employs only the weak law of large numbers. Not only does this proof have the advantage of being expeditious, but it can be extended to a larger class of dependent chance variables.

The consistency of the maximum likelihood estimate follows from the following

THEOREM. *Let* η *and* ϵ *be given, arbitrarily small, positive numbers. Let* $S(\theta_0, \eta)$ *be the open sphere with center* θ_0 *and radius* η, *and let* $\Omega(\eta) = \Omega - S(\theta_0, \eta)$. *Let*

Wald's Assumptions 1–8 hold. There exists a number $h(\eta)$, $0 < h < 1$, *and another positive number* $N(\eta, \epsilon)$ *such that, for any* $n > N(\eta, \epsilon)$,

$$P_0 \left\{ \frac{\sup_{\theta \epsilon \Omega(\eta)} \prod_{i=1}^{n} f(X_i, \theta)}{\prod_{1}^{n} f(X_i, \theta_0)} > h^n \right\} < \epsilon$$

where P_0 *is the probability of the relation in braces according to* $f(x, \theta_0)$.

Proof: Proceed exactly as in the proof of Wald's Theorem 1 and obtain r_0, $\rho_{\theta 1}, \cdots, \rho_{\theta_h}$, so that the set theoretic sum of the open spheres $S(\theta_i, \rho_{\theta_i})$, $i = 1, 2, \cdots, h$, covers the compact set which is the intersection of $\Omega(\eta)$ with the sphere $|\theta| \leq r_0$. Define $T(\theta_i)$, $i = 1, \cdots, h + 1$, as follows:

$$-2T(\theta_i) = E \log f(X, \theta_i, \rho_{\theta_i}) - E \log f(X, \theta_0)$$

$(i = 1, \cdots, h)$

$$-2T(\theta_{h+1}) = E \log \varphi(X, r_0) - E \log f(X, \theta_0).$$

If any of the right members above are infinite let $T(\theta_i)$ be one, say. Thus all $T(\theta_i)$ are positive. Applying the weak law of large numbers we have that, for any i such that $1 \leq i \leq h + 1$, there exists a positive number N_i such that, when $n > N_i$,

$$P_0 \left\{ \frac{\prod_{1}^{n} f(X_i, \theta_i, \rho_{\theta_i})}{\prod_{1}^{n} f(X_i, \theta_0)} > \exp(-nT(\theta_i)) \right\} > \frac{\epsilon}{h+1}$$

$(i = 1, \cdots, h)$

$$P_0 \left\{ \frac{\prod_{1}^{n} \varphi(X_i, r_0)}{\prod_{1}^{n} f(X_i, \theta_0)} > \exp(-nT(\theta_{h+1})) \right\} > \frac{\epsilon}{h+1}.$$

From this the theorem follows immediately, with

$$N(\eta, \epsilon) = \max_i N_i$$

$$h(\eta) = \max_i \exp\{-T(\theta_i)\}.$$

The author is obliged to Prof. Wald for his kindness in making his paper available to the author.