SOME INFERENCE THEOREMS IN STOCHASTIC PROCESSES[1]

BY M. M. RAO

Communicated by William Feller, August 13, 1962

1. **Introduction.** Let $\{X(t),\ t\in T\}$ be a stochastic process with $\{(\Omega,\ \mathfrak{B},\ P_\alpha),\alpha\in A\}$, as the family of underlying probability spaces. Here T is an interval on the real line and A is a subset of the euclidian k-space, indexing the probability measures on $\mathfrak{B}$. If $A=\bigcup_{i=1}^{m}A_i$ $(A_i\cap A_j=\varnothing,\ i\neq j)$, then a major problem in the inference theory of stochastic processes is to decide, on the basis of one realization of the process at the t-points on a subset of T, the correct subfamily P_α, $\alpha\in A_i$. Another problem is to estimate α in some optimal way. The purpose of the present paper is to report some further developments on these problems (cf. [3; 7]), and in particular to present the results that are valid without assuming the processes to be stationary, Markovian, or the like. A new feature here is to introduce Wald's theory [9] in the present general set up, and also to include "explosive" processes [6]. Of course the study of special processes is of interest and it is then possible to use special techniques too (cf., e.g., [3, pp. 233–247; 5]), but they are not considered here.

2. **The testing problem.** Let H_i denote the hypothesis that $\alpha\in A_i$, $i=1,\ \cdots,\ m$. For nontriviality of the testing problem the distinctness of the hypotheses must be assumed. If P_α is a probability measure for $\alpha\in A$, then, following [1], the hypotheses H_i are said to be *distinct* if there exists a set E in $\mathfrak{B}$, such that for all $\alpha_i\in A_i$ and all $\alpha_j\in A_j$ $(i\neq j)$, it is true that $P_{\alpha_i}(E)\neq P_{\alpha_j}(E)$, $i,\ j=1,\ \cdots,\ m$. Now suppose that the finite dimensional distributions of the process are absolutely continuous relative to the Lebesgue measure, μ, with densities $f_{t_1,\ldots,t_n}(x_1,\ \cdots,\ x_n;\ \alpha)$, or $f_n(x,\ \alpha)$, depending on $\alpha\in A$, where $t_1<\ \cdots\ <t_n$ are in T. Here α (scalar or vector) is assumed not to depend on n. (See, however, Theorem 2 below.) Also the A_i are closed bounded and connected subsets. (This last assumption is not essential. It simplifies the formulations.)

The following regularity conditions are imposed on $f_n(x,\ \alpha)$:

1. For each x, $f_n(x,\ \alpha)$ is a continuous function of α.

2. If $\mathfrak{A}$ is a σ-field of Borel sets of A, then $f_n(x,\ \alpha)$ is jointly measurable relative to $\mathfrak{B}\times\mathfrak{A}$, and that the carriers of $f_n(x,\ \alpha)$ remain invariant for $\alpha\in A$.

[1] This work was supported in part under the grant NSF-G 14832.

3. If $\xi(\alpha)$ is any probability measure (prior distribution) on $\mathfrak{A}$ assigning positive measure to each A_i, then

(*) $f_{n+1}(x, \alpha)\int_A W_i(\alpha)f_n(x, \alpha)d\xi(\alpha) - f_n(x, \alpha)\int_A W_i(\alpha)f_{n+1}(x, \alpha)d\xi(\alpha)$ is either non-negative or nonpositive for all $n(\geqq n_0$, large) and $\alpha \in A_i$, where $W_i(\alpha)$ is a bounded measurable function (weight function) on A such that $W_i(\alpha) = 0$ if $\alpha \in A_i$ and positive otherwise.

DEFINITION. Let $\delta_n^i(x)$ be the probability that H_i is chosen when $(x_1, \cdots, x_n)$ is observed, where $\sum_{i=1}^m \delta_n^i(x) = 1$. The vector function $\delta_n(x) = (\delta_n^1, \cdots, \delta_n^m)$ is called a decision function. The function $\delta_n(x)$ is said to be a *Bayes solution* (relative to $\xi(\alpha)$ and $W_i(\alpha)$), to the testing problem if, and only if, the following is true:

For all x and $n(\geqq n_0)$, $\delta_n^i(x) = 1$ whenever

$$t_n^i < \min\,(t_n^j, j \neq i, j = 1, \cdots, m),\ t_n^k = \int_A W_k(\alpha)f_n(x, \alpha)d\xi(\alpha).$$

The following result states that such a solution is possible in the general case also. More precisely,

THEOREM 1. *Let* $\{X(t), t \in T\}$ *be a real separable stochastic process without fixed points of discontinuity* $[P_\alpha, \alpha \in A]$ *and with finite dimensional density functions* $[\mu]$, $f_n(x, \alpha)$ *satisfying Conditions* 1–3 *above. Then, relative to any prior distribution* $\xi(\alpha)$ *on* $\mathfrak{A}$ *and weight functions* $W_i(\alpha)$ *satisfying Condition* 3, *there exists an essentially unique Bayes solution for testing the distinct hypotheses* $H_i, i = 1, \cdots, m$, *based on a set of* n $(\geqq n_0$ *large) observations on the process at* t_i *of* D, *a dense denumerable subset of* T. *Moreover the class of Bayes solutions is an essentially complete class.*

REMARK. Condition 3 has content only if A_i have more than one point. If each A_i is a one point set, then taking $\xi(\alpha)$ as a discrete measure concentrating symmetrically on the points of A, it is seen that (*) is always satisfied being identically zero.

To prove the theorem, one considers the stochastic variables[2] $Y_n^{ij} = t_n^i(X)/t_n^j(X)$, $i \neq j$ (t_n^i are defined above) and shows that under the given conditions, for every fixed sequence, $\{Y_n^{ij}, n \geqq 1\}$ forms a sub (or super) martingale according as the one or the other inequality obtains in (*). Then by an application of the corresponding theorem (cf. [2, p. 354]) one shows that $\{Y_n^{ij}\}$ has a limit and is independent of the sequence involved. Then by a detailed analysis of the limit, via Andersen-Jessen Theorem and the assumption of the distinctness of

[2] As usual, the stochastic variables are denoted by capitals and the values assumed by them by the corresponding small letters.

hypotheses, all the other conclusions of the theorem are established. If $m=2$, and A_1, A_2 are one point sets, Y_n^{ij} reduces to the likelihood ratio [2, p. 93 and p. 348]. (Cf. also [7].)

In the above result it was assumed that, in $f_n(x, \alpha)$, α does not depend on n. This may be relaxed. For simplicity the case $m=2$ will be considered with W taking only (0, 1) values.

CONDITION 3′. If $\xi_n(\alpha)$ is a probability measure on $\mathfrak{A}$, which assigns positive measure to both A_1 and A_2 (which are now subsets of the space of bounded measurable or continuous functions on T), and which satisfies the compatibility conditions (i.e., $\xi_n(\alpha_1, \cdots, \alpha_{n-1}, \infty) = \xi_{n-1}(\alpha_1, \cdots, \alpha_{n-1})$, etc.) then

$$f_{n+1}(x, \alpha) \int_{A_1} f_n(x, \alpha) d\xi_n(\alpha) - f_n(x, \alpha) \int_{A_1} f_{n+1}(x, \alpha) d\xi_{n+1}(\alpha)$$

is either non-negative or nonpositive.

Now the following result can be stated:

THEOREM 2. *If the Conditions* 1, 2, 3′ *are assumed instead of* 1, 2, 3 *of Theorem* 1, *and the rest of the hypothesis holds, then also there exists an essentially unique Bayes solution for the testing problem as in that result.*

A great deal of the work on second order processes, [3], can be unified and slightly extended [7], using the results on Karhunen representation and the martingale theory [2]. In the second order case, Hilbert space methods are also available but they are used in estimation problems more conveniently than in the testing problem.

3. **Estimation problems.** Let the index set A be a subset of the euclidian k-space, and the family $\{P_\alpha, \alpha \in A\}$ be dominated by a fixed σ-finite measure λ on $\mathfrak{B}$, with densities $\{f(\omega, \alpha)\}$. If $\hat{\alpha}_n(\omega)$ is an estimator of α, the problem concerns its consistency and efficiency properties. It is first noted that, by extending a method given in [3, p. 230], if $f^2(\omega, \alpha)$ is λ-integrable for each $\alpha \in A$, it is possible to generate infinitely many nontrivial (even unbiased) estimators $\hat{\alpha}(\omega)$ of α. To choose optimal estimators one considers their efficacy relative to a risk function (e.g., variance). The following general result giving a lower bound is useful for that purpose. Let $W_\alpha(t)$ be a symmetric (in each component) convex, non-negative function of $t=(t_1, \cdots, t_k)$ depending on α also, which is jointly measurable in t and α, and such that $W_\alpha(0)=0$.

THEOREM 3. *Let the family of density functions* $[\lambda]$, $f(\omega, \alpha)$ *for* $\alpha \in A$ *satisfy the following conditions*:

1. *For almost all* ω, $D_i(\omega, \alpha) = \partial \log f(\omega, \alpha)/\partial\alpha_i$, $i = 1, \cdots, k$ *exists and* $|D_i| < M_i(\omega)$, *where* $M_i(\omega)$ *is* λ*-integrable, for all* $\alpha \in A$.

2. *For at least one* i, $D_i \neq 0$, *on a set of positive* λ*-measure.*

3. *If for some* p $(\geqq 1)$, $W_\alpha^{1/p}(t)$ *is a symmetric convex function similar to* $W_\alpha(\cdot)$ *defined above, then* $E_\alpha(|D_i|^q) < \infty$, $i = 1, \cdots, k$, $\alpha \in A$, *where* E_α *denotes expectation under the* P_α *measure, and* $q = p/(p-1)$.

Then, for any estimator $T(\omega) = (T_1(\omega), \cdots, T_k(\omega))$ *of* $\alpha = (\alpha_1, \cdots, \alpha_k)$ *in* A, *the lower bound for the risk function,*

$$R(T, \alpha) = E_\alpha(W_\alpha(T_1 - \alpha_1, \cdots, T_k - \alpha_k)),$$

is given by

$$R(T, \alpha) \geqq W_\alpha\left(\frac{b_1(\alpha)}{E_\alpha(G)}, \cdots, \frac{b_k(\alpha)}{E_\alpha(G)}\right) \cdot \left\{\sup_p \left[\frac{E_\alpha(G)}{E_\alpha^{1/q}(G^q)}\right]^p\right\}, \qquad p > 1,$$

where $G = \sum_{i=1}^k |D_i|$, *and* $b_i(\alpha) = \sum_{j=1}^k E_\alpha(T_i D_j)$, *and where the supremum is taken over all* p *for which* $W_\alpha^{1/p}(\cdot)$ *is convex. If* $p = 1$, *then* $E^{1/q}(G^q)$ *is taken as the essential supremum of* G, *and with this interpretation the same bound given above is valid in this case also.*

It is useful to note that $b_i(\alpha) = 1$, for all i, if $T(\omega)$ is an unbiased estimator, and then the lower bound does not depend on T so that an estimator whose risk function is smaller, compared with this bound, may be considered optimal. The result is proven after a slight extension of Theorem 7 of [4]. It reduces to a result of [3, p. 248], if $k = 1$ and $W_\alpha(t) = t^2$. Other risk functions can be considered using the results of [4].

The method of maximum likelihood (m.l.) is very useful for estimation problems in processes [3]. The associated questions of consistency and limit distributions of estimators now become very difficult however. The simplest Gaussian process $\{X_n, n \geqq 1\}$ defined by

$$(**) \qquad X_t + \alpha_1 X_{t-1} + \cdots + \alpha_k X_{t-k} = \epsilon_t,$$

where ϵ_t are independent Gaussian (mean zero, variance one and $\epsilon_t = 0$, $t \leqq 0$), leads to the consistency questions of $\hat{\alpha}_i$, for which no known theorem is applicable even for the case $k = 1$, if $|\alpha| \geqq 1$. One important result is in [8], which however gives the solution only for $k = 1$ *and* $|\alpha| < 1$. By direct calculations, using the structure of (**), consistency can be settled in a number of cases even if $k > 1$, [6], but no general result is available. If $k = 1$, such a result can be given as follows:

THEOREM 4. *Let* $\{X_n, n \geqq 1\}$ *be a (discrete) stochastic process whose finite dimensional distributions are absolutely continuous relative to a*

fixed σ-finite measure λ defined on $(\Omega, \mathfrak{B})$ of the process, with densities $f_n(x_1, \cdots, x_n, \alpha)$, or $f_n(x, \alpha)$, depending on a real parameter α, where $f_n(x, \alpha_1) \neq f_n(x, \alpha_2)$, a.e. $[\lambda]$ if $\alpha_1 \neq \alpha_2$. Suppose that $f_n(x, \alpha)$ satisfies the conditions:

1. *$\partial f_n/\partial\alpha$, $\partial^2 f_n/\partial\alpha^2$ exist and, for each x, are continuous functions of $\alpha \in \overline{A}$ (closure of A) where A is a bounded nondegenerate interval. These are dominated by $G(x)$ and $H(x)$, where $E_\alpha(G(X))$ and $E_\alpha(H(X))$ are bounded for all $\alpha \in \overline{A}$.*

2. *$C_n(\alpha) = E_\alpha(\partial \log f_n/\partial\alpha)$ exists and $C_n(\alpha) \to \infty$ as $n \to \infty$ for $\alpha \in A$.*

3. *If $\phi_n(\alpha) = \partial \log f_n/\partial\alpha$, and $\phi_n'(\alpha) = \partial\phi_n/\partial\alpha$, then for a given $\beta > 0$ there exists an M such that $E_\alpha(\text{lub}_{\alpha'} |\phi_n'(\alpha')|/C_n(\alpha)) \leqq M < \infty$ for all $\alpha, \alpha' \in A$, $|\alpha - \alpha'| < \beta$.*

4. *Given $0 < \delta < 1$, there exists an $\epsilon_\delta > 0$, such that for $\alpha \in \overline{A}$, and all n, $\Pr\{|\phi_n'(\alpha)|/C_n(\alpha) \geqq \epsilon_\delta\} \geqq 1 - \delta$.*

Then the m.l. equation $\phi_n(\alpha) = 0$ has a root $\hat{\alpha}_n$ which is a consistent estimator of α (i.e., $\hat{\alpha}_n$ converges in probability to α, as $n \to \infty$).

This theorem is a considerable extension of the main result of [8] as it also covers many "explosive" processes. In particular the process (**) for $k = 1$ is covered by it. The proof can be given on classical lines with some essential changes.

THEOREM 5. *Every consistent m.l. estimator, given in Theorem 4, has the following property: There exist two sequences of random variables $\{W_n\}$ and $\{V_n\}$ on the same probability space $\{\Omega, \mathfrak{B}, P_\alpha\}$ such that*

$$\lim_{n\to\infty} E_\alpha(W_n) = 0 = \lim_{n\to\infty} \Pr\{V_n = 0\},$$

$$\lim_{n\to\infty} E_\alpha(W_n^2) = 1 = \lim_{n\to\infty} E_\alpha(V_n),$$

$$\lim_{n\to\infty} \{\sqrt{C_n(\alpha)}(\hat{\alpha}_n - \alpha) - (W_n/V_n)\} = 0.$$

REMARK. The above property may be called *weak asymptotic efficiency* of the estimators, extending a classical concept [8].

4. Other possibilities. The multidimensional extension of Theorem 4 presents considerable difficulties since the matrix valued random variables converge, in this generality, almost always to singular matrices (cf. [6, p. 216, Remarks 2 and 3]) and the crucial Condition 4 above has to be formulated differently. A multidimensional extension of [8], which is weaker than that of Theorem 4 can be given without too much difficulty. Also, viewing the prediction problem as

an extension of estimation theory some interesting results can be obtained.

The details and some extensions will appear elsewhere.

REFERENCES

1. A. Berger and A. Wald, *On distinct hypotheses*, Ann. Math. Statist. **20** (1949), 104–109.

2. J. L. Doob, *Stochastic processes*, Wiley, New York, 1953.

3. U. Grenander, *Stochastic processes and statistical inference*, Ark. Math. **1** (1950), 195–277.

4. M. M. Rao, *Theory of lower bounds for risk functions in estimation*, Math. Ann. **143** (1961), 379–398.

5. ———, *Theory of order statistics*, Math. Ann. **147** (1962), 298–312.

6. ———, *Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations*, Ann. Math. Statist. **32** (1961), 195–218.

7. ———, *Inference in stochastic processes*, (to appear).

8. A. Wald, *Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic process*, Ann. Math. Statist. **19** (1948), 40–46.

9. ———, *Statistical decision functions*, Wiley, New York, 1950.

CARNEGIE INSTITUTE OF TECHNOLOGY