The Annals of Mathematical Statistics
1969, Vol. 40, No. 3, 1021–1032

ON FIXED PRECISION ESTIMATION IN TIME SERIES[1]

BY J. R. BLUM AND JUDAH ROSENBLATT

University of New Mexico and Case Western Reserve University

0. Introduction. To the best of the authors' knowledge almost all of the work that has been done until the present on confidence intervals and confidence sets of fixed precision has either concerned independent (usually identically distributed) observations, or has been asymptotic in character.

In this paper we treat the problem of fixed length confidence intervals for the parameters of a discrete m-dependent stationary Gaussian process. Our main result is somewhat depressing; namely, that if m is unknown (i.e., the possible distributions consist of all m-dependent such processes for all m) such estimation is impossible. In fact it is impossible in a rather small subclass of these processes.

In this area there are, however, quite a few surprises. For example, the authors had conjectured that the main difficulty would arise in attempting to distinguish a case of independent observations with large mean and small variance from the case of 0 mean highly correlated observations with large variance. In both cases one would see a large first observation followed by a number of observations close by, and it appeared difficult to arrive at a stopping rule in which one could distinguish these two cases.

Our intuition appeared to be justified when we were able to show (Theorem 1) that for one class in which independence-large mean-small variance and high dependence-large variance cases were both included, there is no J-stage scheme for fixed length confidence interval estimation of the mean whose last $J-1$ sample sizes are determined by differences of values observed in previous stages. Recall that in Stein's two sample scheme the second sample size is determined by the first stage sample variance

$$k^{-1}\textstyle\sum_{i=1}^{k} (X_i - \bar{X}_k)^2 = k^{-1} \sum_{i=1}^{k} (k^{-1} \sum_{j=1}^{k} [X_i - X_j])^2$$

which is a function of differences. In the general case of a stationary Gaussian process the variance of the sample mean (the usual estimator of the mean) is a function of the variances and covariances. We would expect that here too the actual sample sizes "should" be functions of previously observed differences, since sample covariances are also determined by differences. However our intuition does not hold up here, for we show in Theorem 2 that there is a two-stage scheme (whose second sample size is determined not by differences of 1st sample size values alone) for this problem.

In the next section we show that this case is really the exception and that if the nature of the dependence is not sufficiently well known, we are defeated. Over all, the results of this paper imply that making strong statistical inference

Received 6 February 1968; revised 18 December 1968.
[1] Research supported by the National Science Foundation, Grant GP-5217.

in stochastic processes is likely to require a rather precise knowledge of the nature of the dependence of the data being observed. In a sense, when one does not know how much information per observation will come from successive readings, one cannot find a stopping rule. (Using "information" nontechnically, we feel that we obtain the same amount of information per observation in the independent case, more information from the first observation in dependent cases.)

In the final section we give some sufficient conditions of the nature discussed in the preceding paragraph permitting a fixed length confidence interval for μ the mean of a Gaussian tail trivial process (which includes the m-dependent case).

1. A case in which one can distinguish large variance, high correlation from small variance independence large mean.

In this section we assume that

$$\cdots, X_{-1}, X_0, X_1, \cdots \tag{1.1}$$

are independent $N(0, 1)$ random variables, that m is an unknown positive integer, and that $\sigma \geqq 0$ and μ is real, both unknown. We are permitted to observe as many of the random variables $Y_1, Y_2, \cdots$ as desired where

$$Y_n = \mu + \sigma m^{-\frac{1}{2}} \textstyle\sum_{j=1}^{m} X_{n-j}. \tag{1.2}$$

(Though the number of Y_j's we can observe need not be known in advance of observing them, the sampling scheme must terminate with probability one under all possible choices of m, μ, σ.) As mentioned, it would appear hard to distinguish the case $m = 1$, large μ, small σ^2 (independence) from the case large m, $\mu = 0$, large σ^2.

(1.3) THEOREM 1. *If J is any given positive integer, there is no J-stage scheme for observing the process $Y_1, Y_2, \cdots$ which can yield a fixed length, fixed confidence interval for μ if the sample sizes of stages $2, 3, \cdots, J$ are functions of differences of Y_i's observed on previous stages.*

PROOF. It is apparent that all differences of Y_i's can be expressed in terms of

$$V_n = Y_{n+1} - Y_n = \sigma m^{-\frac{1}{2}} X_n - \sigma m^{-\frac{1}{2}} X_{n-m} = \sigma m^{-\frac{1}{2}}(X_n - X_{n-m}).$$

Given any positive integer k and any real $\sigma^* > 0$, choose σ and m so that $V_1, \cdots, V_k$ are independent $N(0, \sigma^*)$ by letting $2\sigma m^{-\frac{1}{2}} = \sigma^*$ and $m > k$. Now when $V_1, V_2, \cdots$ are independent $N(0, \sigma^*)$, for any given J-stage sampling scheme, given $\epsilon_1, \cdots, \epsilon_J \varepsilon (0, 1)$ we can find $b_1, b_2, \cdots, b_J$ such that if $N_1, N_2, \cdots, N_J$ are the sample sizes for the various stages,

$$\begin{aligned} P(N_1 \leqq b_1) &\geqq 1 - \epsilon_1, \\ P(N_2 \leqq b_2 \mid N_1 \leqq b_1) &\geqq 1 - \epsilon_2, \\ &\vdots \\ P(N_J \leqq b_J \mid N_1 \leqq b_1, \cdots, N_{J-1} \leqq b_{J-1}) &\geqq 1 - \epsilon_J. \end{aligned}$$

(Proof is inductive—to see the first step, note that if for all b

$$P(N_2 \leqq b \mid N_1 \leqq b_1) < 1 - \epsilon_2,$$

then $P(N_2 < \infty \mid N_1 \leqq b_1) \leqq 1 - \epsilon_2$, contradicting the definition of a J-stage scheme for $J \geqq 2$.) Hence

$$P(N_1 + \cdots + N_J \leqq b_1 + \cdots + b_J) \geqq (1 - \epsilon_1) \cdots (1 - \epsilon_J).$$

If we now choose $\epsilon_1, \cdots, \epsilon_J$ so that $(1 - \epsilon_1) \cdots (1 - \epsilon_J)$ is near 1 and $k > b_1 + \cdots + b_J$, we can guarantee that with high probability $N_1 + \cdots + N_J \leqq b_1 + \cdots + b_J \equiv B_J$, still retaining the freedom to choose σ and m as large as desired, with $\sigma m^{-\frac{1}{2}} = \sigma^* 2^{-\frac{1}{2}}$. Thus if $m > B_J$ (the value which under the sampling scheme is a high probability upper bound for sample size when the differences are independent $N(0, \sigma^*)$), then a total sample size exceeding B_J is not likely. All we now need to show is that the sample size needed for given length and confidence estimation of μ goes to ∞ with σ where $\sigma^* > 0$ is fixed and $m^{\frac{1}{2}} = 2^{\frac{1}{2}}\sigma/\sigma^*$. To see this, note that any procedure which can be carried out using $Y_1, Y_2 - Y_1, \cdots, Y_{B_J} - Y_{B_J-1}$ can certainly be accomplished using

$$\begin{aligned}
Q_1 &= \mu + \sigma m^{-\frac{1}{2}} \textstyle\sum_{j=1}^{m+1-B_J} X_{1-J}, \\
Q_2 &= \sigma m^{-\frac{1}{2}} X_2, \\
&\vdots \\
Q_{B_J-1} &= \sigma m^{-\frac{1}{2}} X_{B_J-1}, \\
Q_2^* &= \sigma m^{-\frac{1}{2}} X_{2-m}, \\
&\vdots \\
Q_{B_J-1}^* &= \sigma m^{-\frac{1}{2}} X_{B_J-1-m}.
\end{aligned}$$

For m large enough $(m > B_J)$ all of these random variables are independent. Further, $Q_2, \cdots, Q_{B_J-1}, Q_2^*, \cdots, Q_{B_J-1}^*$ can only give us information about $\sigma m^{-\frac{1}{2}}$ since their distribution is completely determined by $\sigma m^{-\frac{1}{2}}$ for $m > B_J$. Thus for $m > B_J$ we may dispense with $Q_2, \cdots, Q_{B_J-1}, Q_2^*, \cdots, Q_{B_J-1}^*$ if we assume $\sigma m^{-\frac{1}{2}} = \sigma^* 2^{-\frac{1}{2}}$ known. But Var $Q_1 = \sigma^2$, and since there is no restriction on σ^2 from above, it follows from Dantzig's work [3], that we cannot obtain a fixed length confidence interval based on Q_1 for μ, even assuming $\sigma m^{-\frac{1}{2}}$ known.

Hence we cannot obtain such an interval from $Q_1, Q_2, \cdots, Q_{B_J-1}, Q_2^*, \cdots, Q_{B_J-1}^*$ and *a fortiori* from $Y_1, Y_2 - Y_1, \cdots, Y_{B_J} - Y_{B_J-1}$, i.e., from the J-stage sample $Y_1, Y_2, \cdots, Y_{B_J}$. This proves the asserted result. Q.E.D.

Though the authors have not pursued the subject (due to the next theorem), it is not implausible to conjecture that there is a purely sequential sampling scheme based on the differences $V_1, V_2, \cdots$ since we should be able to wait long enough to see some dependence between the differences.

(1.4) THEOREM 2. *There is a procedure requiring only two stages of observation, the first stage consisting of* Y_1, Y_2 *for fixed length confidence interval estimation of* μ *in* (1.2).

PROOF. From the results of [2] it is seen that we can obtain a $1 - \alpha/4$ probability upper bound, $\sigma_{\alpha/4}$, for σ (of the form $k_{\alpha/4}|Y_1|$). Further since $V_1 = Y_2 - Y_1 = \sigma m^{-\frac{1}{2}}(X_1 - X_{1-m})$ is normal with mean 0 and variance $2\sigma^2 m^{-1}$, from it we can obtain a $1 - \alpha/4$ probability positive lower bound, $\gamma_{\alpha/4}$, for $2\sigma^2 m^{-1}$, unless $Y_2 - Y_1 = 0$. This special case may be easily disposed of as follows: $Y_2 - Y_1 = 0$ happens with nonzero probability if and only if $\sigma = 0$, and in this case with probability 1 we have $Y_1 = \mu$. Thus, when $Y_2 - Y_1 = 0$ we terminate sampling and assert with 100% confidence that $\mu = Y_1$. For the remaining cases $Y_1 - Y_2 \neq 0$, we may combine our previous results to see that with probability at least $1 - \alpha/2$,

$$m \leqq 2\sigma_{\alpha/4}^2 \gamma_{\alpha/4}^{-1} \equiv M_{\alpha/2}.$$

But clearly once we know an upper bound $M_{\alpha/2}$ for m with probability $1 - \alpha/2$, with this probability we can obtain independent observations each with variance less than σ^2. Thus with probability at least $1 - 3\alpha/4$ we can obtain independent observations with variance less than $\sigma_{\alpha/4}^2$. From such observations we can decide on a sample size which will yield a $1 - \alpha/4$, l length confidence interval for μ. Combining all the stages together yields a $1 - \alpha$, l length confidence interval for μ as asserted. Q.E.D.

2. Main nonexistence theorem. In this section, as before, $\cdots X_{-2}, X_{-1}, X_0, X_1, X_2, \cdots$ will denote a doubly infinite sequence of independent normally distributed random variables with mean value 0 and variance 1.

Our object here is to show that the class of all m-dependent stationary Gaussian processes does not admit fixed length fixed confidence intervals for the mean—even allowing purely sequential sampling schemes (which are assumed to terminate with probability 1).

In order to do this we look at the subclass $Y_1, Y_2, \cdots$ defined by

$$Y_n = \mu + X_n + Z_{n,m} \tag{2.1}$$

where

$$\begin{aligned} Z_{n,m} &= 0 && \text{if } m = 0 \\ &= m^{-\frac{1}{2}} \textstyle\sum_{j=1}^{m} X_{n-j} && \text{if } m = 1, 2, \cdots. \end{aligned} \tag{2.2}$$

Note that $Y_1, Y_2, \cdots$ form a stationary m-dependent Gaussian process. It is assumed that we can observe any finite number of values of $Y_1, Y_2, \cdots$ and that μ and m are unknown to the observer. The intuitive reason for our inability to obtain fixed length confidence intervals for μ lies in the fact that for large m the random variables $Z_{n,m}$ vary only slightly with n. Hence if $Z_{1,m} = z$ it appears as if Y_n are iid random variables with mean $\mu + z$ and variance 1. The variation (or complete lack of variation) in $Z_{n,m}$ as n varies that was detectable previously (and permitted determination of a high probability upper bound for m) is masked by the changes which are also taking place in the X_n term. The technical difficulty in the proof arises from the fact that any value z of $Z_{1,m}$ has probability zero; hence use must be made of a *set* of such possible $Z_{1,m}$ values. We now proceed with the rigorous development. Let M be a fixed positive integer and let $P_{M,\mu,m}$ denote the distribution of $Y_1, Y_2, \cdots, Y_M$ corresponding to the values μ and m

in (2.1) and let $P_{M,\mu,m|z}$ denote the conditional distribution of $Y_1, Y_2, \cdots, Y_M$ given $Z_{1,m} = z$ (also corresponding to the values μ and m in (2.1)).

LEMMA 1. *For each fixed positive integer* M

$$\lim_{m\to\infty} \sup_{A \varepsilon B_M, z \varepsilon B, \mu \text{ real}} |P_{M,\mu,m|z}(A) - P_{M,\mu+z,0}(A)| = 0$$

where B_M *is the class of Borel sets in* M *dimensional Euclidean space and* B *is any fixed bounded set of reals.*

PROOF. It is clear that if we can prove that

$$\lim_{m\to\infty} \sup_{A \varepsilon B_M, z \varepsilon B} |P_{M,0,m|z}(A) - P_{M,z,0}(A)| = 0$$

then the asserted result follows, since letting

$$A_\mu = \{(x_1, \cdots, x_M): (x_1 + \mu, \cdots, x_M + \mu) \varepsilon A\}$$

we have $$P_{M,\mu,m|z}(A) = P_{M,0,m|z}(A_\mu)$$

and $$P_{M,\mu+z,0}(A) = P_{M,z,0}(A_\mu).$$

Hence we may assume $\mu = 0$ in the remainder of the proof.

We now make use of the following theorem on the multivariate normal distribution (see [1], p. 29):

If $\begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}$ is a normally distributed vector, with mean $\begin{pmatrix} \mu_{(1)} \\ \mu_{(2)} \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, then the conditional distribution of $X_{(1)}$ given $X_{(2)} = x_{(2)}$ is normal with mean $\mu_{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(x_{(2)} - \mu_{(2)})$ and covariance matrix $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ (provided Σ_{22} is invertible).

In the case being considered,

$$X_{(1)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_M \end{pmatrix}, \qquad X_{(2)} = Z_{1,m}, \qquad \begin{pmatrix} \mu_{(1)} \\ \\ \mu_{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

and hence it can be seen that for $m > M$,

$$(\Sigma_{11})_{ik} = \delta_{ik} + 1 - |k - i| m^{-1} + (1 - \delta_{ik})m^{-\frac{1}{2}}$$

$$\Sigma_{22} = 1, \qquad (\Sigma_{12})_k = 1 - (k - 1)m^{-1}.$$

Hence for $m > M$, the distribution of $\begin{pmatrix} Y_1 \\ \vdots \\ Y_M \end{pmatrix}$ given $Z_{1,m} = z$ is normal with mean

$$\begin{pmatrix} 1 \\ 1 - (1/m) \\ 1 - (2/m) \\ \vdots \\ 1 - [(M - 1)/m] \end{pmatrix} z$$

and covariance matrix whose (i, k) element is

$$\begin{aligned} &\delta_{ik} + 1 - |k - i|\, m^{-1} + (1 - \delta_{ik})m^{-\frac{1}{2}} \\ (2.3) \qquad &\qquad + (1 - (k-1)m^{-1})(1 - (i-1)m^{-1}) \\ &\qquad = \delta_{ik} - |k - i|\, m^{-1} + (k-1)m^{-1} + (i-1)m^{-1} \\ &\qquad - (k-1)(i-1)m^{-2} + (1 - \delta_{ik})m^{-\frac{1}{2}}, \end{aligned}$$

where δ_{ik} is the Kronecker delta.

From this it follows that the density of $\begin{pmatrix} Y_1 \\ \vdots \\ Y_M \end{pmatrix}$ given $Z_{1,m} = z$ approaches the normal density with mean $\begin{pmatrix} z \\ z \\ \vdots \\ z \end{pmatrix}$ and identity covariance matrix, the approach being uniform for all z in any given bounded set of reals, and over any given bounded subset of M dimensional Euclidean space. To see this let us denote the elements of the covariance matrix Γ by Γ_{ij} and denote the cofactors by γ_{ij}. Then

$$\Gamma^{-1} = \operatorname{adj} \Gamma / \det \Gamma = (\gamma_{ij})^T / \det \Gamma.$$

Since $\Gamma = I - A$ where A consists of elements all close to 0 when m is large, the elements of Γ^{-1} are all close to those of I since γ_{ij} are all close to δ_{ij} and $\det \Gamma$ is close to 1.

For z in any bounded set B and $\begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix}$ in any bounded set D the vector

$$\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} - \begin{pmatrix} 1 \\ 1 - 1/m \\ \vdots \\ 1 - [M-1]/m \end{pmatrix} z$$

lies in some bounded set G and is uniformly close to the vector

$$\mathbf{w} = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} z \quad \text{for large } m.$$

Hence we see that for $\begin{pmatrix} x_1 \\ \vdots \\ x_M \end{pmatrix} \varepsilon\, D$ and $z \,\varepsilon\, B$ the value $\mathbf{v}'\Gamma^{-1}\mathbf{v}$ lies in some bounded set H and is uniformly close to $\mathbf{w}'\mathbf{w}$ for large m. Since det Γ is close to 1 for large m, from the uniform continuity of the exponential on bounded sets we have the asserted uniform approach of the conditional density of $\begin{pmatrix} Y_1 \\ \vdots \\ Y_M \end{pmatrix}$.

Now let $f_{M,z,0}$ denote the M dimensional multivariate normal density with mean $\begin{pmatrix} z \\ z \\ \vdots \\ z \end{pmatrix}$ and identity covariance matrix, and let $f_{M,0,m|z}$ denote the M dimensional multivariate normal density with mean

$$\begin{pmatrix} 1 \\ 1 - 1/m \\ 1 - 2/m \\ \vdots \\ 1 - [M - 1]/m \end{pmatrix} z$$

and covariance matrix whose ik element is given by (2.3) ($m > M$ is assumed). Given $\epsilon > 0$, we may choose an M dimensional interval I so that for all z in the bounded set B, $\int_I f_{M,z,0}(u)\, du > 1 - \epsilon/4$. From the previous arguments concerning uniform approach of $f_{M,0,m|z}$ to $f_{M,z,0}$ we see that I can also be chosen so that for all $m > M$

$$\int_I f_{M,0,m|z}(u)\, du > 1 - \epsilon/4.$$

We see that for each event A

$$\begin{aligned} |P_{M,0,m|z}(A) - P_{M,z,0}(A)| &= \left|\int_A f_{M,0,m|z}(u)\, du - \int_A f_{M,z,0}(u)\, du\right| \\ &\leqq \int_A |f_{M,0,m|z}(u) - f_{M,z,0}(u)|\, du \\ &= \left(\int_{A\cap I} + \int_{A\cap I^c}\right) |f_{M,0,m|z}(u) - f_{M,z,0}(u)|\, du \\ &\leqq \int_{A\cap I} |f_{M,0,m|z}(u) - f_{M,z,0}(u)|\, du + \epsilon/2. \end{aligned}$$

If we let $V(I)$ denote the M dimensional volume of I, then using the uniform approach of $f_{M,0,m|z}$ to $f_{M,z,0}$ derived before, we may choose m large enough so that for all z in the arbitrary bounded set B and all u in the bounded M dimensional interval I,

$$|f_{M,0,m|z}(u) - f_{M,z,0}(u)| < \epsilon/2V(I);$$

it then follows that

$$|P_{M,0,m|z}(A) - P_{M,z,0}(A)| < \epsilon. \tag{2.4}$$

Thus we have shown that if B is any bounded set of reals, we may choose m sufficiently large so that for all M dimensional Borel sets A and all z in B (2.4) holds, concluding the proof. Q.E.D.

(2.5) THEOREM 3. *Let $l > 0$ and $0 < \alpha < \frac{1}{3}$ be given. Under the condition that*

(a) $$\int_{-l}^{l} (2\pi)^{-\frac{1}{2}} e^{-t^2/2}\, dt < 1 - 3\alpha,$$

if $Y_n = \mu + X_n + Z_{n,m}$ as in (2.1) there does not exist a sampling plan which terminates with probability one for all real μ and all $m = 0, 1, 2, \cdots$ and which leads to an l length $1 - \alpha$ confidence interval for μ.

PROOF. The proof hinges on the following key Lemma: If there is a sampling scheme whose sample size is denoted by N, which terminates with probability one for all real μ when $m = 0$, then there exists a bounded set $\mathcal{K} \subseteq [-l, l]^c$ having $N(0, 1)$ measure exceeding 2α, and for each positive integer i an associated positive integer b_i, such that for all $\mu \in \mathcal{K}$

(2.6) $$P_{\mu,0}(N \leqq b_i) > 1 - i^{-1},$$

(where the subscript μ, 0 refers to the values of μ and m occurring in (2.1)).

PROOF OF LEMMA. First note that for fixed b, i, $P_{\mu,0}\{N \leqq b\}$ is continuous in μ. Hence $\{\mu \in \mathcal{R}: P_{\mu,0}\{N \leqq b\} > 1 - i^{-1}\}$ is open, hence measurable, as is $\{\mu \in \mathcal{R}: P_{\mu,0}\{N \leqq b\} < 1 - i^{-1}\}$. Thus their complements,

$$\{\mu \in \mathcal{R}: P_{\mu,0}\{N \leqq b\} \leqq 1 - i^{-1}\}$$

and $\{\mu \in \mathcal{R}: P_{\mu,0}\{N \leqq b\} \geqq 1 - i^{-1}\}$ are measurable, as are the intersections of all of the above with any interval. Now choose k_α so that

$$\int_{-k_\alpha}^{-l} (2\pi)^{-\frac{1}{2}} e^{-t^2/2}\, dt + \int_{l}^{k_\alpha} (2\pi)^{-\frac{1}{2}} e^{-t^2/2}\, dt = \tfrac{5}{2}\alpha$$

(that this can be done follows from the hypothesis). For the fixed positive integer i, look at the measurable set

$$\mathcal{K}_{b,i} = \{\mu \in [-k_\alpha, -l) \cup (l, k_\alpha]: P_{\mu,0}\{N \leqq b\} > 1 - i^{-1}\}.$$

We claim that for sufficiently large b (call this value b_i), the $N(0, 1)$ measure of $\mathcal{K}_{b,i}$ must exceed $\frac{5}{2}\alpha - \alpha 2^{-(i+1)}$. For $\mathcal{K}_{b,i}$ is nondecreasing in b. If as $b \to \infty$, $\mathcal{K}_{b,i}$ does not increase to within $N(0, 1)$ measure $\alpha 2^{-(i+1)}$ of $\frac{5}{2}\alpha$, the $N(0, 1)$ measure of $[-k_\alpha, -l) \cup (l, k_\alpha]$, then because $\mathcal{K}_{b,i}$ is nondecreasing in b there must exist a set of μ values of $N(0, 1)$ measure at least $\alpha 2^{-(i+1)}$ which are not in $\mathcal{K}_{b,i}$ for any b. That is, there would exist a set of μ's of $N(0, 1)$ measure at least $\alpha 2^{-(i+1)}$ such that $P_{\mu,0}\{N \leqq b\} \leqq 1 - i^{-1}$ for all b; i.e. for these μ's $P_{\mu,0}\{N < \infty\} < 1$, a contradiction of the assumption that the sampling scheme terminates with probability one.

Now we let

$$\mathcal{K} = \bigcap_{i=1}^{\infty} \mathcal{K}_{b_i,i}$$

and note that $\mathcal{K}$ has $N(0, 1)$ measure exceeding 2α, and is bounded. It is easily seen that (2.6) is satisfied, proving the key lemma.

We now turn to the proof of the theorem. If the theorem were false then there is a sampling scheme with sample size N which is finite with probability 1 for all μ and m, and a confidence interval of length l and confidence $1 - \alpha$ for μ based on $Y_1, \cdots, Y_N$; that is, there exist functions $I_n(y_1, \cdots, y_n)$ such that

$$P_{\mu,m}(|I_N(Y_1, \cdots, Y_N) - \mu| \leqq l/2) \geqq 1 - \alpha \qquad \text{for all} \quad \mu, m.$$

We call the event $|I_N(Y_1, \cdots, Y_N) - z| \leqq l/2$ "cover z," and

$$|I_N(Y_1, \cdots, Y_N) - z| > l/2,$$

"not cover z." By the contradiction hypothesis,

$$\begin{aligned} \alpha \geqq P_{0,m}(\text{not cover } 0) &\geqq \int_{\mathcal{K}} P_{0,m}(\text{not cover } 0 \mid Z_{1,m} = z)\, dP_{0,m}(Z_{1,m} \leqq z) \\ &\geqq \int_{\mathcal{K}} P_{0,m}(\text{cover } z \mid Z_{1,m} = z)\, dP_{0,m}(Z_{1,m} \leqq z); \\ &\qquad \text{since for} \quad z \varepsilon \mathcal{K}, \qquad \text{not cover } 0 \supseteq \text{cover } z \\ &\geqq \int_{\mathcal{K}} \{P_{0,m}(\text{cover } z, N \leqq b_i \mid Z_{1,m} = z)\, dP_{0,m}(Z_{1,m} \leqq z), \end{aligned}$$

i.e.,

$$\int_{\mathcal{K}} P_{0,m}(\text{cover } z, N \leqq b_i \mid Z_{1,m} = z)\, dP_{0,m}(Z_{1,m} \leqq z) \leqq \alpha. \tag{2.7}$$

Now for any given ϵ and any i that is chosen, there exists by Lemma 1 an integer $m(b_i, \epsilon, \mathcal{K})$ such that if $m \geqq m(b_i, \epsilon, \mathcal{K})$ then for all $z \varepsilon \mathcal{K}$

$$|P_{0,m}(\text{cover } z, N \leqq b_i \mid Z_{1,m} = z) - P_{z,0}(\text{cover } z, N \leqq b_i)| \leqq \epsilon.$$

(Note that the intersection of "cover z" with "$N \leqq b_i$" is determined by $Y_1, \cdots, Y_{b_i}$ making Lemma 1 applicable. Note that we need the full power of Lemma 1 because this inequality must hold for *all* $z \varepsilon \mathcal{K}$, and "cover z" varies with z.) Furthermore by (2.6) we have

$$|P_{z,0}(\text{cover } z) - P_{z,0}(\text{cover } z, N \leqq b_i)| = |P_{z,0}(\text{cover } z, N > b_i)| < i^{-1}.$$

Hence for $z \varepsilon \mathcal{K}$, $m \geqq m(b_i, \epsilon, \mathcal{K})$

$$|P_{0,m}(\text{cover } z, N \leqq b_i \mid Z_{1,m} = z) - P_{z,0}(\text{cover } z)| < \epsilon + i^{-1}.$$

Thus by (2.7) we have for $m \geqq m(b_i, \epsilon, \mathcal{K})$

$$\int_{\mathcal{K}} [P_{z,0}(\text{cover } z) - (\epsilon + i^{-1})]\, dP_{0,m}(Z_{1,m} \leqq z) \leqq \alpha,$$

i.e., for $m \geqq m(b_i, \epsilon, \mathcal{K})$

$$\int_{\mathcal{K}} P_{z,0}(\text{cover } z)\, dP_{0,m}(Z_{1,m} \leqq z) \leqq \alpha + (\epsilon + i^{-1}) P_{0,m}(Z_{1,m} \varepsilon \mathcal{K}). \tag{2.8}$$

But by assumption we have for all m

$$\int_{\mathcal{K}} P_{z,0}(\text{cover } z)\, dP_{0,m}(Z_{1,m} \leqq z) \geqq (1 - \alpha) P_{0,m}(Z_{1,m} \varepsilon \mathcal{K}) \geqq (1 - \alpha) 2\alpha. \tag{2.9}$$

However for $0 < \alpha < \frac{1}{3}$, $\alpha < (1 - \alpha)2\alpha$. Since ϵ and i^{-1} can be chosen arbitrarily (determining of course $m(b_i, \epsilon, \mathcal{K})$), we see that (2.8) and (2.9) contradict each other for sufficiently small $\epsilon + i^{-1}$. Hence there cannot exist a sampling scheme which terminates with probability one and which leads to a $1 - \alpha$ confidence, l length interval for μ, as asserted. Q.E.D.

Before proceeding to the next section we make some final remarks. First, it may still be possible to consistently test $\mu = \mu_0$ versus $\mu = \mu_1$ under (2.1) since here we know that $n^{-1} \sum_{i=1}^{n} Y_i$ must settle down to μ_0 or μ_1, and we might wait to get an estimate of the "drift rate." What cannot be done is to test $\mu \cong 0$ versus $\mu \cong \mu_1$ at arbitrary levels.

We note furthermore that this theorem shows that for $0 < \alpha < \frac{1}{3}$, if $Y_n = \mu + \sigma[\delta_{0m} + (1 - \delta_{0m})\alpha^{-\frac{1}{2}}](X_n + Z_{n,m})$ where $\sigma \geqq 0$, the standard deviation of the Y_j is also unknown, there are no $1 - \alpha$ confidence intervals of any preassigned length for μ.

It is also seen that we cannot with any degree of precision estimate m, the number of steps after which observations become independent in the class of all m dependent stationary Gaussian processes. For if we could do so, then following such estimation we could surely estimate μ of (2.1) in a single stage by obtaining sufficiently many independent observations with variance at most 2, and mean μ. It also appears intuitively that we cannot obtain a fixed length confidence interval for Var Y_n of (2.1); since if we knew Var $Y_n = 1$ we would know $m = 0$ and hence could estimate μ with fixed precision, while if we knew Var $Y_n = 2$, then we could wait until $n^{-1} \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2$ was close enough to 2, the number of terms needed enabling us to estimate m and thus to estimate μ. That is, if we knew Var Y_n it would seem that we could estimate μ precisely. We hope to be able to supply a rigorous proof to support our intuition in a future paper.

The results here have some significance relative to continuous parameter processes—indicating that in certain cases unless sample spacing is close enough, fixed precision inference may be impossible. (The authors hope to return to this topic also in a future paper.)

3. Some sufficient conditions for fixed precision estimation of μ where

$$Y_n = \mu + \sigma \sum_{j=0}^{\infty} a_j X_{n-j}, \sum_{j=0}^{\infty} a_j^2 = 1$$

(the sequence $\cdots, X_{-1}, X_0, X_1, \cdots$ consisting of independent $N(0, 1)$ random variables and the a_j's μ and $\sigma \geqq 0$ being unknown). Here we will restrict consideration to the case in which the sequence $a_0, a_1, \cdots$ is nonincreasing (hence nonnegative).

(3.1) THEOREM 4. *Suppose that $\{a_j\}_{j=0}^{\infty}$ is such that $\sum_{j=0}^{\infty} a_j^2 = 1$, and for each positive lower bound for a_0, say α_0, and each $\epsilon > 0$ we can determine an integer $N(\epsilon, \alpha_0)$ satisfying*

$$\sum_{j=0}^{N(\epsilon,\alpha_0)} a_j^2 \geqq 1 - \epsilon. \tag{3.2}$$

Then there exists a two-stage sampling scheme, for a fixed width $(1 - \alpha)$ confidence interval for μ, with only two observations in the first stage.

REMARKS. This theorem includes Theorem 2 of Section 1. It also includes the Gaussian-Markov case treated in [4]. It also applies to cases such as $a_j = c_j/(j + 1)$ where $0 \leqq c_j \leqq C$ (with C known, a_j nonincreasing,

$$\sum_{j=0}^{\infty} a_j^2 = 1).$$

PROOF. We see that $E(k^{-1}\sum_{n=1}^{k} Y_n) = \mu$

$$\begin{aligned}\operatorname{Var}(k^{-1}\textstyle\sum_{n=1}^{k} Y_n) &= \sigma^2 k^{-1} + 2\sigma^2 k^{-2}\textstyle\sum_{q=1}^{k-1}(k-q)\sum_{j=0}^{\infty} a_j a_{j+q}\\ &\leqq \sigma^2 k^{-1}(1 + 2k^{-1}\textstyle\sum_{q=1}^{k-1}(k-q)[\sum_{j=0}^{\infty} a_j^2]^{\frac{1}{2}}[\sum_{j=0}^{\infty} a_{j+q}^2]^{\frac{1}{2}})\\ &\qquad\text{by the Schwartz inequality}\\ &= \sigma^2 k^{-1}(1 + 2k^{-1}\textstyle\sum_{q=1}^{k-1}(k-q)[\sum_{j=0}^{\infty} a_{j+q}^2]^{\frac{1}{2}})\\ &\leqq \sigma^2 k^{-1}(1 + 2k^{-1}\textstyle\sum_{q=1}^{N(\epsilon,\alpha_0)} k + 2k^{-1}\sum_{q=1}^{k-1}(k-q)\epsilon^{\frac{1}{2}})\\ &= \sigma^2 k^{-1}(1 + 2N(\epsilon,\alpha_0)) + \sigma^2(k-1)k^{-1}\epsilon^{\frac{1}{2}}.\end{aligned}$$

That is,

(3.3) $\operatorname{Var}(k^{-1}\sum_{n=1}^{k} Y_n) \leqq \sigma^2 k^{-1}(1 + 2N(\epsilon,\alpha_0)) + \sigma^2(k-1)k^{-1}\epsilon^{\frac{1}{2}}.$

As in Section 1, we can obtain a $1 - \alpha/4$ probability upper bound $\sigma_{\alpha/4}^2$ for σ^2 based on the first observation Y_1. Now $Y_2 - Y_1$ is normal, mean 0,

$$\operatorname{Var}(Y_2 - Y_1) = 2\sigma^2(1 - \textstyle\sum_{j=0}^{\infty} a_j a_{j+1}).$$

From observing $Y_2 - Y_1$ we can obtain a $1 - \alpha/4$ probability lower bound, $\sigma_{1,2,\alpha/4}^2 > 0$, on $\operatorname{Var}(Y_2 - Y_1)$ (unless $Y_2 = Y_1$, in which case we assert $\mu = Y_1$ as in Theorem 2 of Section 1), i.e., with probability at least $1 - \alpha/4$

$$2\sigma^2(1 - \textstyle\sum_{j=0}^{\infty} a_j a_{j+1}) \geqq \sigma_{1,2,\alpha/4}^2.$$

Hence

$$\textstyle\sum_{j=0}^{\infty} a_j a_{j+1} \leqq 1 - \sigma_{1,2,\alpha/4}^2(2\sigma^2)^{-1} \leqq 1 - \sigma_{1,2,\alpha/4}^2(2\sigma_{\alpha/4}^2)^{-1} \equiv 1 - \gamma_\alpha$$

with probability at least $1 - \alpha/2$. Then

$$\textstyle\sum_{j=0}^{\infty} a_j^2 - \sum_{j=0}^{\infty} a_j a_{j+1} \geqq \gamma_\alpha$$

with probability at least $1 - \alpha/2$. Since we assumed the a_j's decreasing, we have

$$\textstyle\sum_{j=0}^{\infty} a_j^2 - \sum_{j=0}^{\infty} a_{j+1}^2 \geqq \gamma_\alpha,$$

i.e.,

(3.4) $a_0^2 \geqq \gamma_\alpha$ with probability at least $1 - \alpha/2$.

From (3.3), (3.4) and our $1 - \alpha/4$ probability bound for σ^2 we have with probability at least $1 - 3\alpha/4$,

(3.5) $\operatorname{Var}(k^{-1}\sum_{n=1}^{k} Y_n) \leqq \sigma_{\alpha/4}^2 k^{-1}(1 + 2N(\epsilon,\gamma_\alpha)) + \sigma_{\alpha/4}^2(k-1)k^{-1}\epsilon^{\frac{1}{2}}.$

We now choose ϵ so that $\sigma_{\alpha/4}^2(k-1)k^{-1}\epsilon^{\frac{1}{2}}$ is small, and then choose k so that $\sigma_{\alpha/4}^2 k^{-1}(1 + 2N(\epsilon,\gamma_\alpha))$ is small (the total small enough for the variance of a single normal random variable permitting an l-length $1 - \alpha/4$ confidence interval for its mean). Combining these, we obtain the desired interval. Q.E.D.

REFERENCES

[1] ANDERSON, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. Wiley, New York.
[2] ABBOTT, J. H. and ROSENBLATT, JUDAH. (1963). Two stage estimation with one observation on the first stage. *Ann. Inst. Statist. Math.* **15** 3.
[3] DANTZIG, G. B. (1940). On the nonexistence of tests of "Student's" hypothesis having power functions independent of σ. *Ann. Math. Statist.* **11** 186–192.
[4] BLUM, J. R. and ROSENBLATT, JUDAH. (1963). On multistage estimation. *Ann. Math. Statist.* **34** 1452–1458.