Ann. Inst. Statist. Math.
31 (1979), Part A, 487-497

ON SOME SAMPLING SCHEMES FOR ESTIMATING THE PARAMETERS OF A CONTINUOUS TIME SERIES

PHILIP McDUNNOUGH AND DAVID B. WOLFSON

(Received Jan. 9, 1978; revised May 26, 1979)

Summary

In estimating the mean of certain stationary processes it is shown that it is better to sample at fixed equi-spaced time intervals than to sample randomly according to a renewal process. On the other hand it is shown that the estimation of autocorrelation is sometimes better accomplished by random sampling.

1. Introduction

Let $\{X(t),\ t\geqq 0\}$ be a real stationary Gaussian time series with mean μ, variance σ^2, and autocorrelation function $\rho(\tau)$ involving some unknown parameter λ. Estimation of the unknown parameters may be achieved by observing the process $\{X(t)\}$ over some subset T, of the index set $t\geqq 0$. In many cases T and the set $t\geqq 0$ will coincide (asymptotically). This would clearly be the way of extracting, from a sample, the most information about the parameters. On the other hand, T may be confined to be a countable set. This is in fact the most usual situation found in the literature, where T is taken to be the set $\{i\tau,\ i=1, 2, \cdots\}$, $1/\tau$ then representing the sampling rate. The selection of the sampling set T, has received remarkably little attention as has also the case of the relationship between the resulting discrete time process and the underlying continuous time one. Loynes [6] reviews some of the work done on these problems. As he points out, the main contribution in this area seems to be that due to Shapiro and Silverman [9] who deal with the problem of the elimination of "aliasing". They show that if a continuous Gaussian time series is observed over a *random* set T, generated by a renewal process then (subject to conditions on the interarrival distribution which are satisfied for Poisson processes) the autocorrelation function of the underlying process can be completely recovered from that of the sampled one. Recently Gaster and Roberts [5] have considered the effect, on the sampling distribution of the spectral density estimator, of sampling a continuous Gaussian time series using a Poisson process. In addition, they give (see also Daudpota, Dowrich, and Greated [3]) an interesting example involving the measurement of fluid flows where the method used (laser anemometry) yields a randomly sampled time series.

Insofar as parametric estimation is concerned Brillinger [2] has considered the estimation of the mean of a stationary process, using the sample average, for various sampling schemes (see also [3]). Robinson [8] considers the problem of estimating the parameters of a continuous time process, governed by the Langevin equation, from a countable number of unequally spaced (nonrandom) time points. Finally, Taga [10] has shown that, for stationary Markov processes with exponential autocorrelation functions, the sample average is a better estimate of the mean when based on a sample taken at fixed time points, $i\tau$, than at random time points generated by a renewal process with mean interarrival time τ.

This article compares the estimation of the mean and autocorrelation parameter of a stationary Gaussian process under the following two sampling schemes: (i) sampling the continuous time process at fixed equally-spaced times $i\tau$, $i=1, 2, \cdots$ and (ii) sampling at the random times T_i, $i=1, 2, \cdots$ which are independent of $\{X(t)\}$ and form a renewal process with $\Delta T_i \equiv T_i - T_{i-1}$, $T_0 \equiv 0$, having distribution function G. For the most part we shall restrict our attention to the Markov univariate case although a generalization to non-Markov vector processes is given. We also remark that some of the results presented here can be extended in an obvious fashion to include random sampling schemes where $\{\Delta T_i,\ i=1, 2, \cdots\}$ forms a stationary ergodic sequence. Finally the assumption will be made that $X(T_i)$ is a random variable in the sense of being measurable with respect to the product σ-algebra generated by the processes $\{X(t)\}$ and $\{T_i\}$. This in fact need not be true for all processes $\{X(t)\}$ but is certainly the case when $\{X(t)\}$ has right (or left) continuous sample paths (cf. Doob [4]).

2. Estimating the mean

Consider a stationary Gaussian process $\{X(t)\}$ with mean μ, variance σ^2, and autocorrelation function $\rho(\tau)=e^{-\lambda\tau}$. If this process is observed at the fixed equidistant times $\tau, 2\tau, \cdots, n\tau$ then it is well known that the maximum likelihood estimator of μ is simply the sample mean $n^{-1}\sum_{i=1}^{n} X(i\tau)$ which is asymptotically normal with asymptotic mean μ and variance $I_F(\hat{\mu})^{-1}/n$ where $I_F(\hat{\mu})$ may be interpreted as the (asymptotic) information per observation extracted by the sample mean and is given by

$$I_F(\hat{\mu})=(1-e^{-\lambda\tau})/\sigma^2(1+e^{-\lambda\tau}) .$$

On the other hand if $\{X(t)\}$ is observed at the random renewal times T_i, then the sampled process $\{X(T_i),\ i=1,2,\cdots\}$ is stationary with mean μ, variance σ^2, and autocorrelation sequence r_k given by

$$r_k=\mathrm{E}\,[\rho(T_{i+k}-T_i)]=[\mathrm{E}\,(\exp(-\lambda\Delta T_i))]^k ,$$

the expectation being with respect to the distribution, of ΔT_i, G. It is then readily established that the sample average $n^{-1}\sum_{i=1}^{n} X(T_i)$ is asymptotically normal with mean μ and variance $I_{RA}(\hat{\mu})^{-1}/n$ where

$$I_{RA}(\hat{\mu})=[1-\mathrm{E}\,(\exp(-\lambda\Delta T_1))]/\sigma^2[1+\mathrm{E}\,(\exp(-\lambda\Delta T_1))] .$$

In order to show this we set $r_1=\mathrm{E}\,(\rho(\Delta T_1))$ and $b=\mathrm{E}\,(\rho(2\Delta T_1))$. Let

$$Y_n=\frac{\sigma^2}{n}\sum_{i,j}\rho(T_j-T_i) .$$

Now

$$\mathrm{E}\,(Y_n)=\frac{\sigma^2}{n}\sum_{i,j}\mathrm{E}\,(\rho(T_j-T_i))=\frac{\sigma^2}{n}\sum_{i,j} r_1^{|j-i|} ,$$

so that as $n\to\infty$

$$\mathrm{E}\,(Y_n)\to\sigma^2(1+r_1)/(1-r_1)=[I_{RA}(\hat{\mu})]^{-1} .$$

Using the fact that

$$T_{i+k}-T_i=\sum_{j=i+1}^{i+k}\Delta T_j$$

and the exponential nature of $\rho(t)$ we can rewrite Y_n as

$$Y_n=\sigma^2+\frac{2\sigma^2}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\rho(\Delta T_{i+1})\cdots\rho(\Delta T_j) .$$

We shall first establish that $\mathrm{Var}\,(Y_n)\to 0$ as $n\to\infty$. In the calculation of $\mathrm{Var}\,(Y_n)$ there are several cases to consider when evaluating

$$\mathrm{Cov}\,(\rho(\Delta T_{i+1})\cdots\rho(\Delta T_j),\ \rho(\Delta T_{k+1})\cdots\rho(\Delta T_l)) .$$

Case A: $i+1\leqq j<k+1\leqq l$.

Here the covariance is clearly zero.

Case B: $i+1\leqq k+1\leqq j\leqq l$.

It is obviously sufficient to consider only

$$\mathrm{E}\{\rho(\Delta T_{i+1})\cdots\rho(\Delta T_j)\rho(\Delta T_{k+1})\cdots\rho(\Delta T_l)\}$$

(since the convergence of $\mathrm{Var}(Y_n)$ to zero is all that is important). Now this expected value is equal to

$$\begin{aligned}&\mathrm{E}\{\rho(\Delta T_{i+1})\cdots\rho(\Delta T_k)\rho^2(\Delta T_{k+1})\cdots\rho^2(\Delta T_j)\rho(\Delta T_{j+1})\cdots\rho(\Delta T_l)\}\\&=r_1^{k-i}b^{j-k}r_1^{l-j}\,.\end{aligned}$$

Therefore consider

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=i+1}^{j-1}\sum_{l=j}^{n}r_1^{k-i}b^{j-k}r_1^{l-j}\\&\leqq\frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=i+1}^{\infty}r_1^{k-i}b^{j-k}r_1^{l-j}\\&=\frac{1}{1-r_1}\frac{1}{n^2}\sum_{i=1}^{n-1}r_1^{-i}\sum_{j=i+1}^{n}\left(\frac{b}{r_1}\right)^j\sum_{k=i+1}^{j-1}\left(\frac{r_1}{b}\right)^k r_1^j\\&=\frac{1}{1-r_1}\frac{1}{(1-r_1/b)}\frac{1}{n^2}\sum_{i=1}^{n-1}r_1^{-i}\sum_{j=i+1}^{n}b^j\left(\frac{r_1}{b}\right)^{i+1}\left(1-\left(\frac{r_1}{b}\right)^{j-i}\right)\\&=\frac{r_1}{b}\frac{1}{1-r_1}\frac{1}{(1-r_1/b)}\frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}b^{j-i}\left(1-\left(\frac{r_1}{b}\right)^{j-i}\right)\\&=\frac{r_1}{b}\frac{1}{1-r_1}\frac{1}{(1-r_1/b)}\frac{1}{n^2}\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}b^{j-i}-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}r_1^{j-i}\right].\end{aligned}$$

This last expression converges to zero as $n\to\infty$.

Case C: $i+1\leqq k+1\leqq l\leqq j$.

Here it is sufficient to consider

$$\begin{aligned}&\mathrm{E}\{\rho(\Delta T_{i+1})\cdots\rho(\Delta T_k)\rho^2(\Delta T_{k+1})\cdots\rho^2(\Delta T_l)\rho(\Delta T_{l+1})\cdots\rho(\Delta T_j)\}\\&=r_1^{k-i}b^{l-k}r_1^{j-l}\,,\end{aligned}$$

which leads to an examination of

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=i+1}^{j-1}\sum_{l=k+1}^{j}r_1^{k-i}b^{l-k}r_1^{j-l}\\&\qquad\leqq\frac{1}{(1-b/r_1)}\frac{1}{n^2}\sum_{i=1}^{n-1}r_1^{-i}\sum_{j=i+1}^{n}r_1^j\sum_{k=i+1}^{j-1}\left(\frac{r_1}{b}\right)^k\left(\frac{b}{r_1}\right)^{k+1}\end{aligned}$$

(since $b/r_1<1$)

$$=\frac{1}{(1-b/r_1)}\frac{b}{r_1}\frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j-i=1}^{n}r_1^{j-i}(j-i-1)$$

which, again, tends to zero as $n\to\infty$.

The remaining cases need not be worked out separately, because of symmetry. In summary, we thus have, using self-explanatory notation,

$$\operatorname{Var}(Y_n)=\frac{4\sigma^4}{n^2}\sum_{i,j,k,l}\operatorname{Cov}(i,j,k,l)$$
$$=\frac{4\sigma^4}{n^2}\{\sum_A \operatorname{Cov}(i,j,k,l)+\sum_B \operatorname{Cov}(i,j,k,l)+\sum_C \operatorname{Cov}(i,j,k,l)\}$$

where the three sums are over the i, j, k and l of the three cases considered and their symmetric counterparts.

The first sum is identically zero while

$$0\leqq\sum_B \operatorname{Cov}(i,j,k,l)+\sum_C \operatorname{Cov}(i,j,k,l)$$
$$\leqq\sum_B \mathrm{E}(i,j,k,l)+\sum_C \mathrm{E}(i,j,k,l) \qquad \text{where } \mathrm{E}(i,j,k,l)$$

denotes expectation of the appropriate product.

Consequently we have that $\operatorname{Var}(Y_n)\to 0$ as $n\to\infty$. We therefore conclude that Y_n converges to $I_{RA}(\hat{\mu})^{-1}$ in mean square and thus in probability. Consider the standardized sum

$$Z_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X(T_i)-\mu)\ .$$

Conditional on $T_1,\cdots,T_n$, Z_n is normal with mean 0 and variance Y_n so that its characteristic function is given by

$$\mathrm{E}(\exp(itZ_n))=\mathrm{E}[\mathrm{E}(\exp(itZ_n)\,|\,T_1,\cdots,T_n)]=\mathrm{E}(\exp(-t^2Y_n/2))\ .$$

As $Y_n\geqq 0$, $\exp(-t^2Y_n/2)$ is bounded and, moreover, tends to $\exp(-t^2\cdot I_{RA}(\hat{\mu})^{-1}/2)$ in probability. Therefore by the dominated convergence theorem,

$$\mathrm{E}(\exp(-t^2Y_n/2))\to\exp(-t^2I_{RA}(\hat{\mu})^{-1}/2)$$

which shows that $n^{-1}\sum_{i=1}^{n}X(T_i)$ is asymptotically normal with mean μ and asymptotic variance $I_{RA}(\hat{\mu})^{-1}/n$.

If we now assume that $\mathrm{E}(\varDelta T_1)=\tau$, so that both sampling schemes yield the same average distance between observations, it is easily shown (as noted by Taga [10]) by using the convexity of $e^{-\lambda\tau}$ and Jensen's inequality, that $I_F(\hat{\mu})\geqq I_{RA}(\hat{\mu})$. That is, for estimating μ by the sample average, random is never better than fixed sampling. It should be remembered however, as mentioned in the introduction, that one is sometimes forced to sample at random.

For sampling according to a Poisson process an exact comparison can be made between $I_F(\hat{\mu})$ and $I_{RA}(\hat{\mu})$, for here $\mathrm{E}(\exp(-\lambda\varDelta T_1))=1/(1+$

$\lambda\tau$) so that $I_{RA}(\hat{\mu})$ reduces to $\lambda\tau/\sigma^2(2+\lambda\tau)$ which, as is $I_F(\hat{\mu})$, is an increasing function of τ with a maximum (supremum) of $1/\sigma^2$ at $\tau=\infty$. Thus the maximum information about μ, for varying sampling rates, that can be extracted by the sample average, is the *same* for both random and fixed sampling schemes.

It should be mentioned that the process $\{X(T_i)\}$ is not Gaussian. Furthermore, for random sampling the sample average is no longer the maximum likelihood estimator of μ. To derive this latter estimator we note, setting $Y_i=X(T_i)$, that the joint likelihood of $Y_1,\cdots,Y_n$; $T_1,\cdots,T_n$ (neglecting the marginal of Y_1 which has no asymptotic effect) is

$$\left\{\prod_{i=1}^{n}[2\pi\sigma^2(1-\exp(-2\lambda\Delta T_i))]^{-1/2}\right. \cdot\exp\left(\frac{-[Y_i-\mu-(\exp(-\lambda\Delta T_i))(Y_{i-1}-\mu)]^2}{2\sigma^2(1-\exp(-2\lambda\Delta T_i))}\right)\Bigg\}h(T_1,\cdots,T_n) \tag{2.1}$$

where h represents the joint density of $T_1,\cdots,T_n$. If the T_i's are constants then we set $h=1$. From (2.1) the maximum likelihood estimator of μ is obtained as

$$\hat{\mu}_R=\frac{n^{-1}\sum_{i=1}^{n}(Y_i-Y_{i-1}\exp(-\lambda\Delta T_i))/(1+\exp(-\lambda\Delta T_i))}{n^{-1}\sum_{i=1}^{n}(1-\exp(-\lambda\Delta T_i))/(1+\exp(-\lambda\Delta T_i))}. \tag{2.2}$$

Now, using the strong law of large numbers we have, almost surely (with respect to the product measure obtained from the processes $\{X(t)\}$ and $\{T_i\}$.)

$$\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}(1-\exp(-\lambda\Delta T_i))/(1+\exp(-\lambda\Delta T_i)) =\mathrm{E}[(1-\exp(-\lambda\Delta T_1))/(1+\exp(-\lambda\Delta T_1))] \tag{2.3}$$

and

$$\begin{aligned}\lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}&[(Y_i-(\exp(-\lambda\Delta T_i))Y_{i-1})/(1+\exp(-\lambda\Delta T_i))]\\ &=\mathrm{E}[(Y_i-(\exp(-\lambda\Delta T_1))Y_{i-1})/(1+\exp(-\lambda\Delta T_1))]\\ &=\mu\,\mathrm{E}[1-\exp(-\lambda\Delta T_1))/(1+\exp(-\lambda\Delta T_1))]\end{aligned}$$

so that $\hat{\mu}_R$ is a (strongly) consistent estimator of μ. Now let $t_1,t_2,\cdots$ be a realization of $T_1,T_2,\cdots$ and let $\hat{\mu}_t$ denote the estimate of μ obtained by replacing ΔT_i in (2.2) by Δt_i. Note that $\hat{\mu}_t$ is the maximum likelihood estimator of μ based on a sample $X(t_1),\cdots,X(t_n)$ taken at the nonrandom unequally spaced, times $t_1,\cdots,t_n$. Clearly $\mathrm{E}(\hat{\mu}_t)=\mu$. Furthermore, it is readily shown that the $X(t_i)-X(t_{i-1})\exp(-\lambda\Delta t_i)$ are uncorrelated, and hence independent, for different i. Therefore

$$\text{Var}(\hat{\mu}_t)=(nA_n)^{-2}\sum_{i=1}^{n}\text{Var}\,[(X(t_i)-X(t_{i-1})\exp(-\lambda\Delta t_i))/(1+\exp(-\lambda\Delta t_i))]$$

where we have put

$$A_n=n^{-1}\sum_{i=1}^{n}(1-\exp(-\lambda\Delta t_i))/(1+\exp(-\lambda\Delta t_i))\ .$$

Since $\text{Var}(X(t_i)-X(t_{i-1})\exp(-\lambda\Delta t_i))=\sigma^2(1-\exp(-\lambda\Delta t_i))^2$, we get

$$\text{Var}(\hat{\mu}_t)=(nA_n)^{-2}\sum_{i=1}^{n}(1-\exp(-\lambda\Delta t_i))/(1+\exp(-\lambda\Delta t_i))=(nA_n)^{-1}\ .$$

From (2.3) we have $A_n\to\text{E}\,[(1-\exp(-\lambda\Delta T_1))/(1+\exp(-\lambda\Delta T_1))]$, this limit holding for almost every $\boldsymbol{t}$. Here the "almost every" need only refer to the probability measure generated by the renewal process. Since the cumulants, of order greater than 2, of $\hat{\mu}_t$ are zero we obtain that for almost every $\boldsymbol{t}$, $\hat{\mu}_t$ is asymptotically normal with mean μ and variance $I_R(\hat{\mu})^{-1}/n$ where

$$I_R(\hat{\mu})=\sigma^{-2}\,\text{E}\,[(1-\exp(-\lambda\Delta T_1))/(1+\exp(-\lambda\Delta T_1))]\ .$$

This (asymptotic distribution), not depending on $\boldsymbol{t}$, is thus the asymptotic distribution of $\hat{\mu}_R$. It is interesting to note that $I_F(\hat{\mu})\geqq I_R(\hat{\mu})\geqq I_{RA}(\hat{\mu})$. It should also be pointed out that $\text{E}(\hat{\mu}_R)=\text{E}\,[\text{E}(\hat{\mu}_R|\boldsymbol{T}=\boldsymbol{t})]=\mu$ so that the maximum likelihood estimator of μ based on $X(T_1),\cdots,X(T_n)$ is unbiased and is therefore the minimum variance unbiased estimator of μ.

Finally we note that the asymptotic variance of $\hat{\mu}_R$ is given by I_n^{-1} where $I_n\equiv nI_R(\hat{\mu})$ is easily calculated, using (2.1), to be the Fisher information about μ contained in $X(T_1),\cdots,X(T_n)$. Thus (roughly speaking) $\hat{\mu}_R$ asymptotically achieves the Cramér-Rao lower bound for variances.

3. Estimation of the mean and autocorrelation

As we have seen, the estimation of the mean is better accomplished by using a fixed sampling scheme as opposed to a random one. This, surprisingly, is not always the case when estimating other parameters. For instance, consider the estimation of λ in the autocorrelation of the Markov process considered in section two. It is well known that the maximum likelihood estimator of λ, calculated on the basis of the fixed sampling times $\{i\tau\}$, is asymptotically normal with mean λ and variance $I_F(\hat{\lambda})^{-1}/n$ where

$$I_F(\hat{\lambda})=\tau^2e^{-2\lambda\tau}(1+e^{-2\lambda\tau})/(1-e^{-2\lambda\tau})^2\ .$$

On the other hand, random sampling can be shown to yield a consistent root of the likelihood equation which is asymptotically normal with mean λ_0 (the true value of λ) and variance $I_R(\hat{\lambda})^{-1}/n$ where

$$I_R(\hat{\lambda})=\mathrm{E}\,[(\varDelta T_1)^2 \exp(-2\lambda_0\varDelta T_1)(1+\exp(-2\lambda_0\varDelta T_1))/(1-\exp(-2\lambda_0\varDelta T))^2]\ .$$

Indeed, let $l_n(\lambda)$ denote the log likelihood function of $(X(T_1), T_1),\cdots,(X(T_n), T_n)$. From (2.1) we have, omitting the marginal for $X(T_1)$,

$$\begin{aligned} l_n(\lambda)=-n\log(2\pi\sigma)-\frac{1}{2}\sum_{i=1}^{n}\Big\{&\log(1-\exp(-2\lambda\varDelta T_i)) \\ &+\frac{[Y_i-\mu-(\exp(-\lambda\varDelta T_i))(Y_i-\mu)]^2}{\sigma^2(1-\exp(-2\lambda\varDelta T_i))}\Big\} \\ &+\log h(T_1, T_2,\cdots, T_n)\ . \end{aligned}$$

In order to show that the likelihood equation, $(\partial/\partial\lambda)\log\{l_n(\lambda)\}=0$, has a consistent root $\hat{\lambda}$ it suffices to show that (cf. Rao [7], p. 300)

$$\frac{1}{n}l_n(\lambda)\xrightarrow{p}\mathrm{E}\,[l_n(\lambda)/n]\ , \tag{3.1}$$

which is, in fact independent of n.

Clearly, since the $\varDelta T_i$ are independent and have the same distribution,

$$\frac{1}{n}\sum_{i=1}^{n}\log(1-\exp(-2\lambda\varDelta T_i))\xrightarrow{p}\mathrm{E}\,[\log(1-\exp(-2\lambda\varDelta T_1))]\ .$$

We are implicitly assuming here that $\mathrm{E}\,[|\log(1-\exp(-2\lambda\varDelta T_1))|]<+\infty$. This in fact holds for any bounded continuous density function (of $\varDelta T_1$). We also have

$$h(T_1,\cdots, T_n)=\prod_{i=1}^{n} g(\varDelta T_i)\ ,$$

where g is the p.d.f. of $\varDelta T_1$.

$$\frac{1}{n}\log h(T_1,\cdots, T_n)\xrightarrow{p}\mathrm{E}\,[\log g(\varDelta T_1)]\ .$$

Finally consider

$$\mathrm{E}\,\{[Y_i-\mu-(\exp(-\lambda\varDelta T_i))(Y_{i-1}-\mu)]^2/(1-\exp(-2\lambda\varDelta T_i))\}\ .$$

By conditionning first on $\varDelta T_i$ this reduces to

$$\mathrm{E}\,\{(1+\exp(-2\lambda\varDelta T_i)-2\exp(-(\lambda+\lambda_0)\varDelta T_i))/(1-\exp(-2\lambda\varDelta T_i))\}$$

which clearly is finite. Furthermore, a straightforward calculation yields

$$\mathrm{Var}\,\frac{1}{n}\sum_{i=1}^{n}[Y_i-\mu-(\exp(-\lambda\Delta T_i))(Y_{i-1}-\mu)]^2/(1-\exp(-2\lambda\Delta T_i))\to 0$$

so that

$$\frac{1}{n}\sum_{i=1}^{n}[Y_i-\mu-(\exp(-\lambda\Delta T_i))(Y_{i-1}-\mu)]^2/(1-\exp(-2\lambda\Delta T_i))$$

converges to its expectation in probability. These results establish (3.1). Now expand the likelihood function about λ_0 in a Taylor series. Usual arguments yield

$$\sqrt{n}(\hat{\lambda}-\lambda_0)=\frac{(1/\sqrt{n})(dl_n(\lambda)/d\lambda_0)}{d^2l_n(\lambda)/d\lambda_0^2}+\varepsilon_n \tag{3.2}$$

where $\varepsilon_n \xrightarrow{p} 0$. A direct calculation (as done previously) yields

$$\frac{1}{n}\frac{d^2l_n(\lambda)}{d\lambda_0^2}\xrightarrow{p}\mathrm{E}\left(\frac{1}{n}\frac{d^2l_n(\lambda)}{d\lambda_0^2}\right)=-I_R(\hat{\lambda})\,. \tag{3.3}$$

Consider (setting $\rho=e^{-\lambda_0}$, $\mu=0$, and $Z_i=Y_i-\rho^{\Delta T_i}Y_{i-1}$ for convenience),

$$-\frac{1}{\sqrt{n}}\frac{dl_n(\lambda)}{d\lambda_0}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\Delta T_i\rho^{\Delta T_i}}{1-\rho^{2\Delta T_i}}\left[\rho^{\Delta T_i}+\frac{\rho^{\Delta T_i}Z_i^2}{1-\rho^{2\Delta T_i}}+Z_iY_{i-1}\right]\,. \tag{3.4}$$

We consider the behaviour of (3.4) for fixed $\Delta T_i=\Delta t_i$. We have, since $dl_n(\lambda)/d\lambda_0$ represents the score function of $X(t_1),\cdots,X(t_n)$,

$$\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{dl_n(\lambda)}{d\lambda_0}\right)=0\,.$$

Also, noting that $Z_1,\cdots,Z_n$ are independent, as well as Z_i and Y_j for $j\leqq i$, it is straightforward (though lengthy) to show that

$$\mathrm{E}\left(\frac{1}{\sqrt{n}}\frac{dl_n(\lambda)}{d\lambda_0}\right)^2\to I_R(\hat{\lambda})$$

while the higher-order cumulants tend to zero. Consequently, for almost every $\boldsymbol{t}$, $(1/\sqrt{n})(dl_n(\lambda)/d\lambda_0)$ is asymptotically normal with mean zero and variance $I_R(\hat{\lambda})^{-1}/n$. This also holds unconditionally and also for $\hat{\lambda}$ by (3.2), (3.3) and (3.4).

Now it is straightforward (though tedious) to show that there exists an x_0 such that the function $x^2e^{-2x}(1+e^{-2x})/(1-e^{-2x})^2$ is concave for $0<x<x_0$ and convex for $x>x_0$. Hence (using Jensen's inequality) there exists a τ_0 such that for $\Delta T_1<\tau_0$ almost surely

$$I_R(\hat{\lambda})\geqq(\mathrm{E}(\Delta T_1))^2\exp(-2\lambda\,\mathrm{E}(\Delta T_1))(1+\exp(-2\lambda\,\mathrm{E}(\Delta T_1)))/$$
$$(1-\exp(-2\,\mathrm{E}(\Delta T_1)))^2$$

$$=\tau^2 e^{-2\lambda\tau}(1+e^{-2\lambda\tau})/(1-e^{-2\lambda\tau})^2=I_F(\hat{\lambda}) .$$

Similarly for $\Delta T_1>\tau_0$ almost surely, $I_R(\hat{\lambda})\geqq I_F(\hat{\lambda})$. Thus, roughly speaking, if one samples rapidly, then fixed is better than random while for slow sampling random is better than fixed. It should also be noted that $I_F(\hat{\lambda})$ is a decreasing function of τ with a maximum of $1/2\lambda^2$ at $\tau=0+$. On the other hand for random sampling $\tau=\mathrm{E}(\Delta T_1)$ converging to 0 implies that ΔT_1 converges to 0 in probability and hence $(\Delta T_1)^2\cdot(\exp(-2\lambda\Delta T_1))(1+\exp(-2\lambda\Delta T_1))/(1-\exp(-2\Delta T_1))^2$ converges in probability to $1/2\lambda^2$. Thus as $\mathrm{E}(\Delta T_1)\to 0$, $I_R(\hat{\lambda})\to 1/2\lambda^2$. From these results and those of Section 2 we see that both random and fixed sampling yield the same maximal precision for estimating either the mean or the autocorrelation of a stationary normal Markov process. That this is not necessarily the case may be seen by considering the joint estimation of μ and λ. We recall (cf. Rao [7], p. 271) that the Fisher information matrix is a measure of the information contained in a sample about the unknown parameters. A direct calculation for both fixed equispaced and random sampling schemes yields information matrices I_F and I_R, respectively, which are both diagonal. The diagonal elements of I_F are $nI_F(\hat{\mu})$ and $nI_F(\hat{\lambda})$ while those of I_R are $nI_R(\hat{\mu})$ and $nI_R(\hat{\lambda})$. As is well known (see, for example, T. W. Anderson [1], Chapter 7) the determinant of a covariance matrix may be considered as a generalized variance. Hence one possible comparison of the information about μ and λ extracted by fixed and random sampling schemes may be obtained by comparing

$$\det(I_F/n)=I_F(\hat{\mu})I_F(\hat{\lambda})$$

with

$$\det(I_R/n)=I_R(\hat{\mu})I_R(\hat{\lambda}) .$$

Both these quantities depend, of course, on the sampling rate. Taking λ and σ^2 to both be 1 we find numerically that the maximum of $\det(I_F/n)$ is approximately .0950 this being attained at a sampling interval of $\tau=.9889$. On the other hand for Poisson random sampling $\det(I_R)$ has a maximum of around .1037 this being achieved at a slower average sampling rate corresponding to $\mathrm{E}(\Delta T_i)=1.3982$. Assuming then, that the determinant of the Fisher information matrix provides a reasonable measure of the information contained in a sample, about μ and λ, we conclude that random is "better" than fixed sampling for extracting information (about the mean and autocorrelation parameter) from the underlying process.

4. A concluding note

While the above analysis has been confined to real stationary Markov Gaussian processes some extensions are readily obtained. Consider for example the estimation of the mean vector $\boldsymbol{\mu}$ of a stationary vector process $\{\boldsymbol{x}(t),\ t\geqq 0\}$ having autocovariance matrix $C(\tau)\equiv \mathrm{E}\,[(\boldsymbol{x}(t)-\boldsymbol{\mu})\cdot(\boldsymbol{x}(t+\tau)-\boldsymbol{\mu})']$. Assume that the matrix $P(\tau)=(C(\tau)+C(\tau)')/2$ is convex in the sense that every quadratic form $\boldsymbol{s}'P(\tau)\boldsymbol{s}$ is a convex function of τ. An example of such a case is a stationary vector Markov Gaussian process with $C(\tau)=C(\tau)'=\exp(-\Lambda\tau)$ for some positive (semi) definite matrix Λ. If we estimate $\boldsymbol{\mu}$ by the sample mean then

$$\begin{aligned}\Sigma_R &\equiv \lim_{n\to\infty} n\,\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{X}(T_i)\right)\\ &=C(0)+2\sum_{k=1}^{\infty}R_k\,, \qquad \text{where } R_k\equiv \mathrm{E}\,[P(T_{i+k}-T_i)]\\ &\geqq C(0)+2\sum_{k=1}^{\infty}P(k\tau)\\ &=\Sigma_F\equiv \lim_{n\to\infty} n\,\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}\boldsymbol{X}(i\tau)\right).\end{aligned}$$

Note that $\Sigma_R\geqq\Sigma_F$ means that $\Sigma_R-\Sigma_F$ is positive semi-definite.

This research was supported by the National Research Council of Canada Personal Operating Grants A9261 and A8347.

UNIVERSITY OF TORONTO
MCGILL UNIVERSITY

REFERENCES

[1] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons Inc.
[2] Brillinger, D. R. (1972). Estimation of the mean of a stationary time series by sampling, *J. Appl. Prob.*, **10**, 419-431.
[3] Daudpota, Q. I., Dowrich, G. and Greated, C. A. (1977). Estimation of moments of a Poisson-sampled random process, *J. Phys.* A: Math., Gen., **10**, 471-483.
[4] Doob, J. L. (1953). *Stochastic Processes*, John Wiley, New York, 518-526.
[5] Gaster, M. and Roberts, J. B. (1977). The spectral analysis of randomly sampled records by a direct transform, *Proc. R. Soc.*, A, **354**, 27-58.
[6] Loynes, R. M. (1969). Aliasing and related questions in stationary processes, *Proc. Twelfth Biennial Sem. Canad. Math. Cong. on Time Series and Stochastic Processes* (Vancouver, B.C.), 89-98.
[7] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, 2nd edition, John Wiley & Sons Inc.
[8] Robinson, P. M. (1977). Estimation of a time series model from unequally spaced data, *Stoch. Processes Appl.*, **6**, 9-24.
[9] Shapiro, Harold S. and Silverman, Richard A. (1960). Alias-free sampling of random noise, *J. Soc. Indust. Appl. Math.*, **8**, 225-248.
[10] Taga, Y. (1965). The optimal sampling procedure for estimating the mean of a stationary Markov process, *Ann. Inst. Statist. Math.*, **17**, 105-112.