On the Choice of Sampling Rates in Parametric Identification of Time Series

Åström, Karl Johan

1968

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Åström, K. J. (1968). *On the Choice of Sampling Rates in Parametric Identification of Time Series*. (Technical Reports TFRT-7003). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

ON THE CHOICE OF SAMPLING RATES IN PARAMETRIC IDENTIFICATION OF TIME SERIES

K. J. ÅSTRÖM

REPORT 6807 NOVEMBER 1968
LUND INSTITUTE OF TECHNOLOGY
DIVISION OF AUTOMATIC CONTROL

ON THE CHOICE OF SAMPLING RATES IN PARAMETRIC IDENTIFICATION OF TIME SERIES †

K.J. Åström

ABSTRACT

Aliasing gives a lower bound for the sampling rate in ordinary spectral analysis of a time series. In parametric it appears at first sight that no such limitations are present. In this note we will get insight into this paradox by analysing a simple gauss-markov process. We assume that a time series analysis is performed based on N samples of the series at equal spacing h. The result shows that there is an optimal choice of h and that the variance increases rapidly when h increases from the optimal value. The results obtained when a time series of fixed length T is analysed with a different number of samples are also analysed.

† This work was partially supported by the Swedish Board for Technical Development under Contract 68-336-f

1. STATEMENT OF THE PROBLEM

Consider the stochastic differential equation

$$dx = - \alpha x dt + dw \tag{1}$$

where $\{w(t)\}$ is a Wiener process with variance parameter r i.e. $E\, w^2(t) = rt$

Assume that the values of x are observed at equidistant sampling points with spacing h and that we analyse the stochastic process $\{x(t)\}$ using these values. We will then investigate how the result depends on the choice of sampling interval h.

According to the sampling theorem we find that using a sampling interval h we get no useful information of the frequency content above the Nyquist frequency

$$f_c = \frac{1}{2h} \, [\mathrm{Hz}] \tag{3}$$

Using ordinary spectral analysis we also have the effect of aliasing which means that the spectral density in the interval $(-f_c, f_c)$ can be distorted if the signal contains frequencies outside this interval. Using common rules of thumb [Bendat-Piersol p. 288] we get the following rule for the choice of h

$$h < \frac{1}{2f_c}$$

When the stochastic process $\{x(t)\}$ is analysed using parametric models, we formulate the analysis problem as to estimate the parameters α and r which completely describe the process. After the parameters have been obtained we can then compute the spectral density using wellknown formulas. With this approach there are no apparent limitations on the sampling interval due to aliasing. The accuracy of the parameter estimates will, however, depend on the sampling interval. This dependence is analysed in the next section.

2. ANALYSIS

The equation (1) gives

$$x(t+h) = e^{-\alpha h} x(t) + \int_t^{t+h} e^{-\alpha(t+h-s)} dw(s) \tag{4}$$

Hence

$$x(t+1) = a\ x(t) + \sigma\ e(t) \tag{5}$$

where h is chosen as the time unit

$$a = e^{-\alpha h} \tag{6}$$

$$\sigma^2 = E \iint_0^h e^{-2\alpha h}\ e^{\alpha(s+s')}\ dw(s)\ dw(s')$$

$$= \frac{r}{2\alpha}\ \left[1 - e^{-2\alpha h}\right] \tag{7}$$

and $\{e(t)\}$ is a sequence of independent normal (0,1) random variables. In parametric analysis we thus estimate the parameters a and σ from the sampled series $\{x(t),\ t = h,\ 2h,\ \ldots\}$. The estimates of the parameters α and r are then computed from the equations (6) and (7). It has been shown by Mann and Wold that the maximum likelihood estimate is consistent and asymptotic efficient in this case. The likelihood function is given by

$$-\log L = \frac{1}{2\sigma^2} \sum_{t=1}^{N} \varepsilon^2(t) + N \log \sigma + \text{const} \tag{8}$$

where

$$\varepsilon(t+1) = x(t+1) - a\ x(t) \tag{9}$$

To compute the Cramér-Rao lower bound of the variance [3] of a and σ we form

$$-\frac{\partial^2}{\partial \alpha^2} \log L = \frac{1}{\sigma^2} \sum_{t=1}^{N} x^2(t+1)$$

$$-\frac{\partial^2}{\partial \alpha \partial \sigma} \log L = -\frac{2}{\sigma^3} \sum_{t=1}^{N} \varepsilon(t)\ x(t-1)$$

$$-\frac{\partial^2}{\partial \sigma^2} \log L = \frac{3}{\sigma^4} \sum_{t=1}^{N} \varepsilon^2(t) - \frac{N}{\sigma^2}$$

Hence

$$\lim_{N\to\infty} -\frac{1}{N}\frac{\partial^2}{\partial\alpha^2}\log L = \lim_{N\to\infty}\frac{1}{\sigma^2 N}\sum_{t=1}^{N} x^2(t-1) = \frac{1}{1-a^2}$$

$$\lim_{N\to\infty} -\frac{1}{N}\frac{\partial^2}{\partial\alpha\partial\sigma}\log L = \lim_{N\to\infty} -\frac{2}{N\sigma^3}\sum_{t=1}^{N}\varepsilon(t)\,x(t-1) = 0$$

$$\lim_{N\to\infty} -\frac{1}{N}\frac{\partial^2}{\partial\sigma^2}\log L = \lim_{N\to\infty}\left(\frac{3}{N\sigma^4}\sum_{t=1}^{N}\varepsilon^2(t) - \frac{1}{\sigma^2}\right) = \frac{2}{\sigma^2}$$

where convergence is with probability one. For large N we thus have the following estimates of the variances of the estimates

$$\mathrm{Var}\ \hat{a} \geq \frac{1}{N}(1-a^2) \tag{10}$$

$$\mathrm{Var}\ \hat{\sigma} \geq \frac{\sigma^2}{2N} \tag{11}$$

It has been shown by Mann and Wold that the maximum likelihood estimates achieve the lower bounds for large N. The estimate of α is given by

$$\hat{\alpha} = -\frac{1}{h}\log\hat{a} \tag{12}$$

which is the maximum likelihood estimate of α. Using Carpenters formula we get after elementary calculations

$$\mathrm{Var}\ \hat{\alpha} = \frac{\alpha^2}{N} f(\alpha h) \tag{13}$$

$$\mathrm{Var}\ \hat{r} = \frac{2r^2}{N} \tag{14}$$

where

$$f(x) = \frac{e^{2x}-1}{x^2} \tag{15}$$

The graphs of the function f is shown in Fig. 1. We thus find that there is a choice of the sampling interval

$$\alpha h_o = 0.797$$

which gives the smallest variance to the α estimate

$$\min\ \mathrm{Var}\ \hat{\alpha} = 6.177\ \frac{\alpha^2}{N}$$

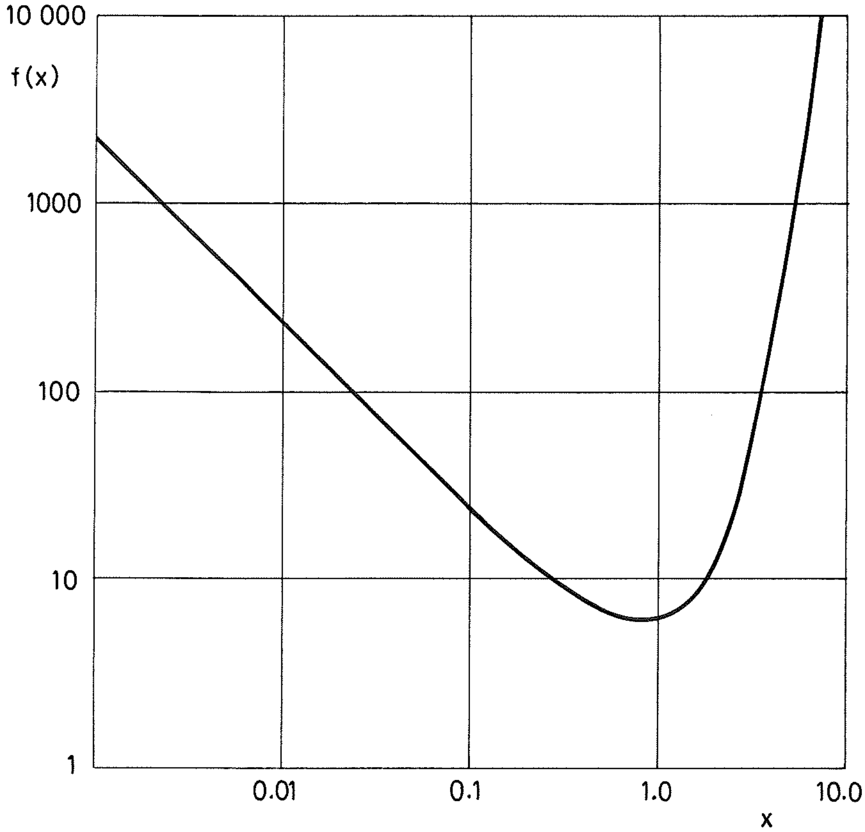

Fig. 1
Graph of the function $f(x) = (e^{2x}-1)x^{-2}$. The variance of the estimate of α using N values with the spacing h is $\alpha^2 f(\alpha h)/N$.

Notice that the variance of $\hat{\alpha}$ increases very rapidly with increasing $h \geq h_o$, but that it increases more moderately with decreasing h. We have e.g. $f(10x_o) = 213401 f(x_o)$ but $f(0.1x_o) =$ $f(0.1x_o) = 4.009 f(x_o)$. The variance of $\hat{r}$ decreases monotonically with increasing sampling interval.

Conclusion

If the parameters α and r of the process (1) are estimated using N values of the process which equal spacing h, then there is an optimal choice $h_o = 0.797/\alpha$ which gives the smallest variance of $\hat{\alpha}$. The variance increases rapidly for sampling intervals greater than h_o. The variance of $\hat{r}$ decreases monotonically with increasing sampling interval.

3. COMPARISON WITH CONTINUOUS TIME ESTIMATION

It is of interest to compare the estimates discussed in the previous section with the estimates based on a continuous record of the process. Let us therefore assume that a realization of the process is known over the whole interval (0,T) and that we estimate the parameters α and r from this record. Such estimation problems have been considered by Arato [1]. His results are based on Striebels [6] explicit formulas for the conditional measure of continuous processes. There is one drastic difference in comparison with the discrete time case namely that the parameter r can be estimated without error. The likelihood function is given by

$$- \log L = \frac{1}{2r} \quad \alpha(x^2(T) - rT) + \alpha^2 \int_0^T x^2(s)\, ds$$
$$- \frac{1}{2} \log \alpha T + \frac{1}{2} \log \pi \qquad (16)$$

Routine calculations now give the following expression for the variance of $\hat{\alpha}$

$$\text{Var } \hat{\alpha} \sim \frac{2\alpha}{T} \qquad (17)$$

The expression holds for large T. To be able to compare with the results for discrete time system, we consider a process of length T which is analysed using discrete data with equal spacing h. From (13) and (15) we get the following asymptotic expression for the variance in the discrete time case

$$\text{Var } \hat{\alpha}_d = \frac{2\alpha}{T} g(\alpha h)$$

where

$$g(x) = \frac{e^{2x} - 1}{2x}$$

A graph of the function g is shown in Fig. 2. As can be seen from this figure the loss of accuracy due to sampling is moderate as long as we have sampling intervals $\alpha h < 1$. For $\alpha h = 1$ we have an increase in variance by a factor 3.2. The variance increases rapidly with increasing sampling interval.

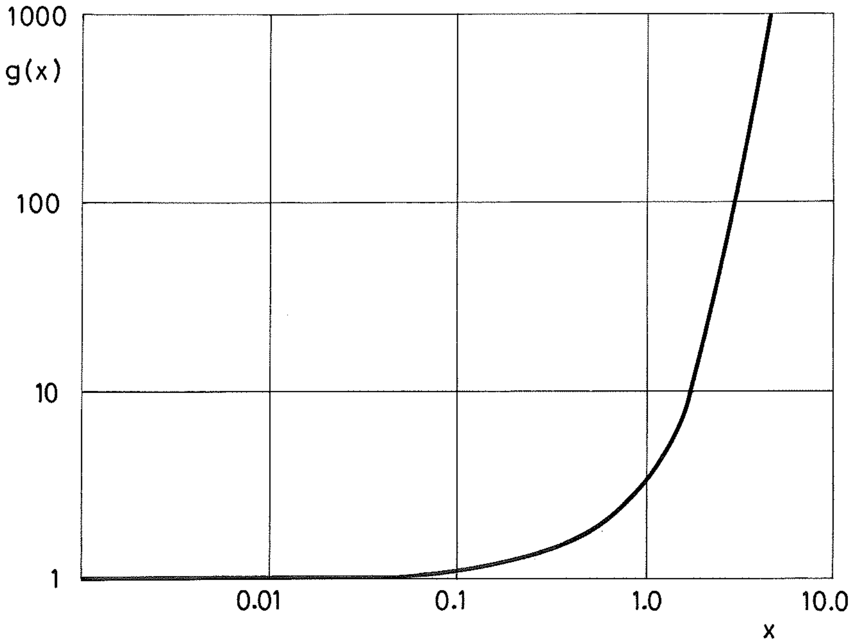

Fig. 2

Graph of the function g. The variance of the estimate of α based on equidistant sampling with spacing h of a record of length T is $2\alpha g(\alpha h)/T$.

References

[1] Arato, M. "Estimation of the Parameters of a Stationary Gaussion Markov Process", Dokl. Akad. Nauk. SSSR 145 (1962) 13 pp.

[2] Åström, K.J. and Bohlin, T. "Numerical Identification of Linear Dynamic Systems from Normal Operating Records", Proc IFAC Symposium on Self-Adaptive Control Systems, Teddington 1965.

[3] Åström, K.J. "On the achievable accuracy in identification problems", Proc. IFAC Symposium on the Problem of Identification in Automatic Control Systems, Prague 1967.

[4] Bendat, J.S. and Piersol, A.G. "Measurement and Analysis of Random Data", Wiley New York 1966.

[5] Mann, H.B. and Wald, A. "On the Statistical Treatment of Linear Stochastic Difference Equations", Econometrica 11 (1943) 173-220.

[6] Striebel, C.T. "Densities for Stochastic Processes", Ann. Math. Stat. 30 (1959) 559-567.