CRAMÉR-RAO BOUNDS FOR CONTINUOUS-TIME AR PARAMETER ESTIMATION WITH IRREGULAR SAMPLING

Erik K. Larsson and Erik G. Larsson

Dept. of Systems and Control, Uppsala University,
P.O. Box 27, SE-751 03 Uppsala, Sweden.
Email: {egl,ekl}@syscon.uu.se

ABSTRACT

Consider the problem of estimating the parameters in a continuous-time autoregressive (AR) model given measurements taken at arbitrary time instants. In this paper the Cramér-Rao bound for this problem is derived by using a technique based on the Slepian-Bang's formula and residue calculus. Furthermore, we investigate by means of numerical experiments how different sampling schemes can affect accuracy. Interestingly enough, however, for the examples studied, the estimation accuracy is relatively insensitive to the choice of sampling strategy.

1. INTRODUCTION

The problem of estimating the parameters in a continuous-time AR model given samples taken at irregularly spaced time instants has been the subject of research for a fairly long time [1, 2]. In this paper the Cramér-Rao bound (CRB) for the problem under study is considered. The motivation for this is twofold: to provide bounds for judging the performance of various estimators, for instance those in [1]; and to illustrate by numerical examples how the sampling scheme can affect the estimation accuracy. We note here that the corresponding CRB in the case of regular sampling has been derived in [3, 4]. However, the results and methods therein appear not to be extendable to irregular sampling in a straightforward manner, so in this paper a different approach is considered. As a further remark, notice that the CRBs for the related problem of estimating the parameters of polynomial-phase signals from irregularly sampled data have been presented in [5]. Apart from the mentioned papers, not much literature on the topic appears to exist.

2. CRAMÉR-RAO BOUND

Consider the following real-valued scalar continuous-time AR process:

$$A\Big(\frac{d}{dt}\Big)y(t) = e(t) \tag{1}$$

where $e(t)$ is continuous-time white noise of variance σ^2, $\frac{d}{dt}$ is the differentiation operator and $A(s)$ is a polynomial of degree $2n$, $n \geq 1$, with all its zeros strictly in the left half plane. For a rigorous treatment of continuous-time stochastic processes see, e.g., [6, 7]. Assume further that all zeros of $A(s)$ are *distinct* and that $A(s)$ has no zeros on the real axis. Under these assumptions, we can write, since the process is real-valued,

$$A(s) = \prod_{j=1}^{n}(s-p_j)(s-p_j^*) \tag{2}$$

where $\{p_j, p_j^*\}_{j=1}^n$ are the zeros of $A(s)$ and $(\cdot)^*$ denotes the complex conjugate. The model is fully parameterized by the *real* parameter vector

$$\boldsymbol{\theta} = \begin{bmatrix}\bar{p}_1 & \tilde{p}_1 & \cdots & \bar{p}_n & \tilde{p}_n & \sigma^2\end{bmatrix}^T \tag{3}$$

where $\bar{(\cdot)}$ and $\tilde{(\cdot)}$ denote the real and imaginary parts, respectively.

The assumptions that all roots $\{p_j, p_j^*\}_{j=1}^n$ are *distinct*, that $\tilde{p}_j > 0$ for $j = 1, \ldots, n$, and that $A(s)$ is of even degree are clearly restrictive, but not seriously so, as we believe for the following reasons: Firstly, in many applications, including spectral line analysis, the poles of $A(s)$ are distinct and non-real. Secondly, our choice of model and its parameterization are motivated by the clarity of presentation. However, it should be stressed that the results herein can easily be extended to the case where $A(s)$ has distinct zeros on the real axis as well.

Assume now that the process is sampled at N time instants $t_1 < \cdots < t_N$. From (1) it follows [8] that the covariance matrix

$$\boldsymbol{R} \triangleq E\left\{\begin{bmatrix}y(t_1)\\ \vdots \\ y(t_N)\end{bmatrix}\begin{bmatrix}y(t_1) & \cdots & y(t_N)\end{bmatrix}\right\} \tag{4}$$

can be expressed as

$$\boldsymbol{R} = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\Phi(s)\boldsymbol{\Gamma}(s)ds \tag{5}$$

where $\int_{-i\infty}^{i\infty}$ denotes integration along the imaginary axis, element k, l of $\boldsymbol{\Gamma}(s)$ equals

$$\boldsymbol{\Gamma}_{k,l}(s) = e^{|t_k - t_l|s} \tag{6}$$

This work was partly supported by the Swedish Foundation for Strategic Research (SSF) and the Swedish Research Council for Engineering Sciences under contract 95-800.

and

$$\Phi(s) \triangleq \frac{\sigma^2}{A(s)A(-s)} \tag{7}$$

is the spectrum of the process $y(t)$.

The Cramér-Rao inequality states that for any unbiased estimate $\hat{\boldsymbol{\theta}}$ of the parameter-vector $\boldsymbol{\theta}$ we must have [8]

$$E\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\} \geq \boldsymbol{J}^{-1} \tag{8}$$

where $\boldsymbol{J}$ is the *Fisher Information Matrix* (FIM) and the matrix inequality $\boldsymbol{A} \geq \boldsymbol{B}$ means that $\boldsymbol{A}-\boldsymbol{B}$ is nonnegative definite. Under the additional assumption that $e(t)$ is Gaussian, a convenient expression for $\boldsymbol{J}$ is given by Slepian-Bang's formula (see, e.g., [9]):

$$\left[\boldsymbol{J}_{k,l}|\{t_i\}_{i=1}^N\right] = \frac{1}{2}\,\mathrm{Tr}\left\{\boldsymbol{R}^{-1}\boldsymbol{R}_k'\boldsymbol{R}^{-1}\boldsymbol{R}_l'\right\} \tag{9}$$

Here $\mathrm{Tr}\{\cdot\}$ denotes the trace, $\boldsymbol{R}_k' \triangleq \frac{d}{d\theta_k}\boldsymbol{R}$ and θ_k denotes the kth element of the parameter-vector $\boldsymbol{\theta}$ which corresponds to any of the parameters $\bar{p}_i$, $\tilde{p}_i$ or σ^2. We write $\left[\boldsymbol{J}_{k,l}|\{t_i\}_{i=1}^N\right]$ to stress that $\boldsymbol{J}$ is conditioned on the sampling instants.

The matrix $\boldsymbol{R}$ in (5) and its derivatives can be evaluated straightforwardly by residue calculus [10]. A detailed derivation can be found in [11], whereas in this paper we merely summarize the results. For notational convenience, let us introduce the following functions:

$$\begin{aligned}\beta_l(s) &\triangleq \frac{1}{(s-p_l)(s-p_l^*)(-s-p_l)(-s-p_l^*)}\\ \beta_{l,\theta_k}(s) &\triangleq \frac{d}{d\,\theta_k}\beta_l(s) = \begin{cases}\frac{d}{d\theta_k}\beta_k(s), & l=k\\ 0, & l\neq k\end{cases}\end{aligned} \tag{10}$$

In [11] it is shown that $\boldsymbol{R}$ and $\boldsymbol{R}_k'$ can be written as

$$\boldsymbol{R} = 2\sum_{j=1}^{n}\bar{\boldsymbol{\Psi}}_{p_j} \tag{11}$$

$$\boldsymbol{R}_k' = 2\sum_{j=1}^{n}\bar{\boldsymbol{\Psi}}_{p_j,\theta_k} \tag{12}$$

where $\boldsymbol{\Psi}_{p_j}$ and $\boldsymbol{\Psi}_{p_j,\theta_k}$ are $N\times N$ residue matrices. The matrix $\boldsymbol{\Psi}_{p_j}$ is given by [11]

$$\boldsymbol{\Psi}_{p_j} = -i\frac{\sigma^2}{8\bar{p}_j\tilde{p}_j p_j}\Big[\prod_{\substack{l=1\\ l\neq j}}^{n}\beta_l(p_j)\Big]\boldsymbol{\Gamma}(p_j) \tag{13}$$

The procedure of evaluating $\boldsymbol{\Psi}_{p_j,\theta_k}$ is divided into three cases, namely, θ_k corresponding to $\bar{p}_k$, $\tilde{p}_k$ or σ^2. The residue matrix $\boldsymbol{\Psi}_{p_j,\sigma^2}$ is

$$\boldsymbol{\Psi}_{p_j,\sigma^2} = \frac{\boldsymbol{\Psi}_{p_j}}{\sigma^2} \tag{14}$$

where $\boldsymbol{\Psi}_{p_j}$ is given by (13). To evaluate $\boldsymbol{\Psi}_{p_j,\theta_k}$ for θ_k corresponding to $\bar{p}_k$ and $\tilde{p}_k$ the cases $k\neq j$ and $k=j$ need to be treated separately. For $k\neq j$ it holds that

$$\boldsymbol{\Psi}_{p_j,\theta_k} = -i\frac{\sigma^2}{8\bar{p}_j\tilde{p}_j p_j}\Big[\prod_{\substack{l=1\\ l\neq j,k}}^{n}\beta_l(p_j)\Big]\beta_{k,\theta_k}(p_j)\boldsymbol{\Gamma}(p_j) \tag{15}$$

where

$$\beta_{k,\theta_k}(s) = \begin{cases}-4\bar{p}_k\big(|p_k|^2-s^2\big)\beta_k^2(s), & \theta_k\sim\bar{p}_k\\ -4\tilde{p}_k\big(|p_k|^2+s^2\big)\beta_k^2(s), & \theta_k\sim\tilde{p}_k\end{cases} \tag{16}$$

For the case $k=j$ it can be shown that

$$\begin{aligned}\boldsymbol{\Psi}_{p_j,\theta_j} = \sigma^2\Bigg\{&h_{j,\theta_j}(p_j)\sum_{\substack{m=1\\ m\neq j}}^{n}\Big[\frac{d}{d\,s}\beta_m(s)\Big|_{s=p_j}\prod_{\substack{t=1\\ t\neq j,m}}^{n}\beta_t(p_j)\Big]\boldsymbol{\Gamma}(p_j)\\ &+\Big[\prod_{\substack{l=1\\ l\neq j}}^{n}\beta_l(p_j)\Big]\cdot\Big[\frac{d}{d\,s}h_{j,\theta_j}(s)\Big|_{s=p_j}\boldsymbol{\Gamma}(p_j)\\ &+h_{j,\theta_j}(p_j)\frac{d}{d\,s}\boldsymbol{\Gamma}(s)\Big|_{s=p_j}\Big]\Bigg\}\end{aligned} \tag{17}$$

where

$$\begin{aligned}h_{j,\theta_j}(p_j) &= \begin{cases}-\frac{i}{8\bar{p}_j\tilde{p}_j p_j}, & \theta_j\sim\bar{p}_j\\ \frac{1}{8\bar{p}_j\tilde{p}_j p_j}, & \theta_j\sim\tilde{p}_j\end{cases}\\ \frac{d}{d\,s}\beta_m(s)\Big|_{s=p_j} &= 4p_j\big(\bar{p}_m^2-\tilde{p}_m^2-p_j^2\big)\beta_m^2(p_j)\\ \frac{d}{d\,s}h_{j,\theta_j}(s)\Big|_{s=p_j} &= \begin{cases}i\frac{p_j+\bar{p}_j}{8\bar{p}_j^2\tilde{p}_j p_j^2}, & \theta_j\sim\bar{p}_j\\ i\frac{p_j+i\tilde{p}_j}{8\bar{p}_j\tilde{p}_j^2 p_j^2}, & \theta_j\sim\tilde{p}_j\end{cases}\\ \frac{d}{d\,s}\boldsymbol{\Gamma}(s)\Big|_{s=p_j} &= \boldsymbol{\Omega}\odot\boldsymbol{\Gamma}(p_j)\end{aligned} \tag{18}$$

and $\boldsymbol{\Omega}$ is an $N\times N$ matrix whose element k,l equals $\boldsymbol{\Omega}_{k,l} = |t_k-t_l|$. In (18) $\odot$ denotes element-wise multiplication.

We now summarize the steps for computing the CRB:

I) Compute the matrix $\boldsymbol{\Psi}_{p_j}$ for $j=1,\ldots,n$ according to (13) and (10).

II) Compute the matrix $\boldsymbol{\Psi}_{p_j,\theta_k}$ for $j=1,\ldots,n$ and $k=1,\ldots,2n+1$. To do so, use the fact that $\boldsymbol{\Psi}_{p_j,\sigma^2}$ is given by (14), and that $\boldsymbol{\Psi}_{p_j,\bar{p}_k}$ and $\boldsymbol{\Psi}_{p_j,\tilde{p}_k}$ for $k=1,\ldots,n$ are given by (15)-(16) for $k\neq j$ and by (17)-(18) for $k=j$.

III) Evaluate $\boldsymbol{R}$ according to (11) and $\boldsymbol{R}_k'$ for $k=1,\ldots,2n+1$ according to (12).

IV) Compute the FIM according to (9).

3. NUMERICAL EXAMPLES

The CRB derived in Section 2 is conditioned on the sampling instants $\{t_i\}_{i=1}^N$. However, if the sampling process is random and independent of $e(t)$ we rather want to compute, cf. [5],

$$CRB = \Big\{E_{\{t_i\}}\left[\boldsymbol{J}|\{t_i\}\right]\Big\}^{-1} \tag{19}$$

where $[\boldsymbol{J}|\{t_i\}]$ is the conditional Fisher information matrix in (9). Finding an analytic expression for (19) given a probabilistic description of the sampling process appears to be far from trivial. We state this as an open research problem and resort to numerical evaluation of (19) by means of Monte-Carlo simulations. In the examples below we have used 200 Monte-Carlo runs.

We consider the following sampling strategies:

a) *(Deterministic) uniform*: $t_i = iT$, $i = 1, \ldots, N$.

b) *Uniformly distributed*: $t_i = \sum_{k=1}^{i} \Delta t_k$, $i = 1, \ldots, N$, where

$$\Delta t_k = T + \delta_k, \qquad \delta_k \sim \text{Uniform}(-\delta_0, \delta_0) \tag{20}$$

and δ_k are independent of $e(t)$ for all t, and independent of δ_j for all $j \neq k$. Here $T = E\{\Delta t_k\}$ is the mean sample interval. In Example 1 and Example 2 below, we chose $\delta_0 = T/2$.

c) *Two-point distributed*: as b) above but

$$\Delta t_k = \begin{cases} \Delta_0 & \text{with probability } p_0 \\ \Delta_1 & \text{with probability } 1 - p_0. \end{cases} \tag{21}$$

Also here, Δ_{t_k} are independent of each other and of $e(t)$. In this case the mean sample interval is $T = E\{\Delta t_k\} = p_0\Delta_0 + (1 - p_0)\Delta_1$. In Example 3.1 and Example 3.2 below, we chose $p_0 = 1/2$, $\Delta_0 = T/2$ and $\Delta_1 = 3T/2$.

d) *Gapped-data sampling*:

$$\begin{aligned} t_i = {} & 0, \tilde{\Delta}, \ldots, (\tilde{N}-1)\tilde{\Delta}, \Delta, \Delta + \tilde{\Delta}, \ldots, \Delta + (\tilde{N}-1)\tilde{\Delta}, \\ & \ldots, (N_\# - 1)\Delta, (N_\# - 1)\Delta + \tilde{\Delta}, \ldots, \\ & (N_\# - 1)\Delta + (\tilde{N}-1)\tilde{\Delta} \end{aligned} \tag{22}$$

In this case the mean sample interval is

$$T = \frac{(N_\# - 1)\Delta + (\tilde{N}-1)\tilde{\Delta}}{N_\#\tilde{N} - 1} \tag{23}$$

In the examples below, the values of the parameters are chosen as $\tilde{N} = 5$, $\tilde{\Delta} = T/10$, $N_\# = N/\tilde{N}$ and Δ so that (23) is fulfilled.

Note that for all considered sampling strategies, a)-d) the mean sample interval is by definition $T \triangleq E\{t_i - t_{i-1}\}$.

Example 1 In the first example we consider an AR process with poles at $\bar{p}_1 \pm i\tilde{p}_1 = -0.05 \pm i$. Two experiments are performed: Firstly we fix the number of samples $N = 100$ and vary the mean sample interval T between 0.01 and 0.5. Secondly N is varied between 10 and 200, for a fixed T. Figure 1 shows the result of the first experiment. The CRB for estimating $\tilde{p}_1$ is shown for different values of T. Note that $\tilde{p}_1$ can be interpreted as the location of a "spectral line", so estimating $\tilde{p}_1$ is of particular practical relevance. We note two things from Figure 1:

1. CRB decreases for increasing T (this behavior was also seen in [1]).
2. There is, maybe somewhat surprisingly no big difference between the CRB for different sampling schemes. This behavior was also seen in [5] for a different but related problem. Note that the length of the interval during which the process is observed is $\approx NT$, which is independent of the sampling scheme used.

Figure 2 shows the result of the second experiment. Here $T = 0.5$ as N is varied. Owing to obvious reasons the CRB decreases as N increases and for sufficiently large NT it appears that the CRB decreases linearly with N (for a fixed T). We will return to this observation later. Furthermore, we see also in this example that the CRB is almost independent of the sampling scheme.

Example 2 In this example the behavior of the CRB for estimating the location of two closely spaced spectral lines is considered. The first spectral line is fixed, as $\bar{p}_1 = -0.1$ and $\tilde{p}_1 = 1$, while the second spectral line is varying according to $\bar{p}_2 = \bar{p}_1/(1 + \delta)$ and $\tilde{p}_2 = \tilde{p}_1 + \delta$. For $\bar{p}_i << \tilde{p}_i$ the normalization of $\bar{p}_2$ will make the peaks of the spectral lines to be of approximately the same height. The number of samples is $N = 100$ and the mean sample interval is $T = 0.5$. In Figure 3 the CRB for estimating $\tilde{p}_1$, as a function of δ for different sampling strategies is shown. It is observed that the accuracy increases rapidly with increasing δ, until a certain threshold is reached, which seems to constitute a "fundamental" bound on the achievable accuracy. Moreover it is seen, as in the previous example, that the achievable accuracy, for a fixed mean sample interval, is very insensitive to the choice of sampling scheme.

4. ASYMPTOTIC BEHAVIOR

Direct evaluation of the expression in (9) for the CRB becomes impractical as N increases. However, the CRB can be *extrapolated* for large N. The following result aim at providing some justification for doing that.

Proposition 1 Assume that the following conditions hold:

$\mathcal{C}1$: All sampling intervals $t_k - t_{k-1}$ are multiples of a minimum value $\varepsilon > 0$. In other words, $t_k - t_{k-1} = M_k\varepsilon$ and $t_1 = M_1\varepsilon$ for some sequence of strictly positive integers $\{M_k\}_{k=1}^{N}$.

$\mathcal{C}2$: The sequence of sampling intervals $\{t_k - t_{k-1}\}$ (or, equivalently, M_k) is a realization of a white stationary random process. Furthermore, there is a constant $c < \infty$ such that $\sup_{1 \leq k \leq \infty} M_k \leq c$.

Then it follows that

$$\frac{E_{\{t_i\}_{i=1}^{N}}\left[\boldsymbol{J}_{k,k}|\{t_i\}_{i=1}^{N}\right]}{E_{\{t_i\}_{i=1}^{2N}}\left[\boldsymbol{J}_{k,k}|\{t_i\}_{i=1}^{2N}\right]} \to \frac{1}{2} \tag{24}$$

as $N \to \infty$. In other words, the diagonal elements of the average Fisher Information matrix are proportional to N if N is sufficiently large.

Proof 1 *See [11].*

5. CONCLUDING REMARKS

The CRB for estimating the parameters in a continuous-time AR model given irregularly sampled data was considered. A conceptually simple method for computing the CRB by using Slepian-Bang's formula in combination with residue calculus was developed. Furthermore, the influence on the parameter estimation of different sampling patterns was demonstrated by means of some numerical examples. One of the conclusions from these examples is that the CRB is relatively insensitive to the sampling scheme as long as the mean inter-sampling distance is maintained.

6. ACKNOWLEDGEMENT

The authors are grateful to Prof. Torsten Söderström for his comments on an early version of this paper.

7. REFERENCES

[1] E. K. Larsson and T. Söderström, "Approaches for identifying continuous-time AR processes from unevenly sampled data," in *Proc. of IFAC SYSID 2000 (Symposium on System Identification)*, (Santa Barbara, California), June 21 – 23 2000.

[2] T. Söderström and J. Ängeby, "On the prediction error method for estimating continuous-time models from unevenly sampled data," Tech. Rep. UPTEC 90084R, Department of Technology, Uppsala University, Uppsala, Sweden, 1990.

[3] T. Söderström, "On the Cramér-Rao lower bound for estimating continuous-time autoregressive parameters," in *Proc. 14th World Congress of IFAC*, vol. H, (Beijing, P.R. China), pp. 175–180, July 5–9 1999.

[4] T. Söderström, "On computing the asymptotic accuracy of estimated linear state space models," Tech. Rep. UPTEC 97031R, Department of Technology, Uppsala University, Uppsala, Sweden, 1997.

[5] J. A. Legg and D. A. Gray, "Performance bounds for polynomial phase parameter estimation with nonuniform and random sampling schemes," *IEEE Transactions on Signal Processing*, vol. 48, pp. 331–337, February 2000.

[6] K. J. Åström, *Introduction to Stochastic Control Theory*. New York, NY: Academic Press, 1970.

[7] T. Söderström, *Discrete-Time Stochastic Systems*. Hemel Hempstead, United Kingdom: Prentice Hall International, 1994.

[8] T. Söderström and P. Stoica, *System Identification*. United Kingdom: Prentice Hall International, 1989.

[9] P. Stoica and R. Moses, *Introduction To Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, 1997.

[10] R. V. Churchill and J. W. Brown, *Complex Variables and Applications*. United States of America: McGraw-Hill International Edition, 1990.

[11] E. G. Larsson and E. K. Larsson, "On the Cramér-Rao bound for continuous-time AR parameter estimation with irregular sampling," *Submited to IEEE Transactions on Signal Processing*, August 2000.

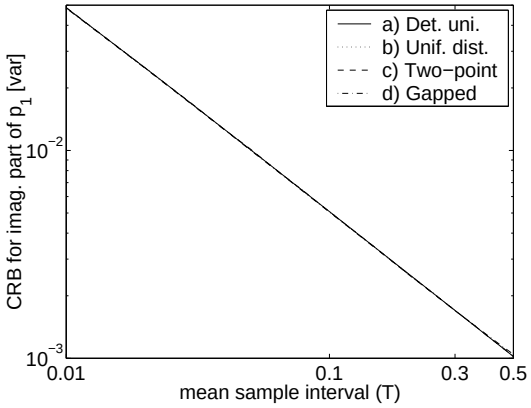

Figure 1: CRB for $\tilde{p}_1$ as a function of T for the process described in Example 1. Here $N = 100$.

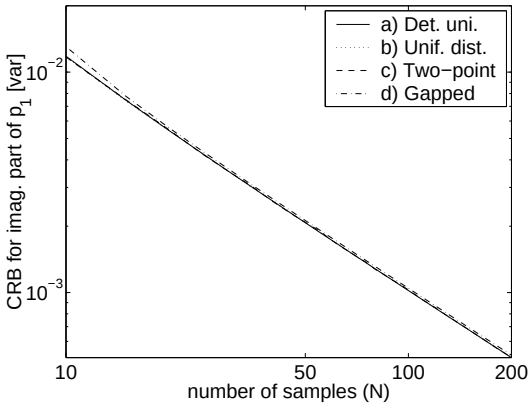

Figure 2: CRB for $\tilde{p}_1$ as a function of N for the process described in Example 1. Here $T = 0.5$.

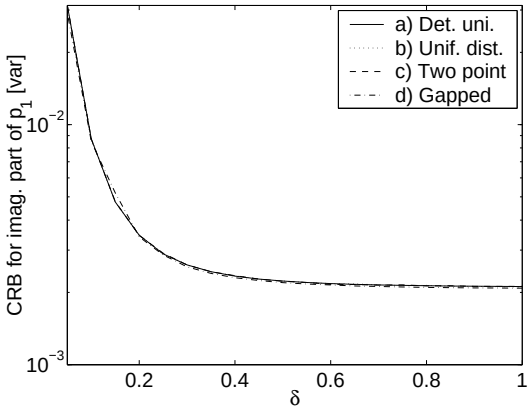

Figure 3: CRB for $\tilde{p}_1$ as a function of δ for the process in Example 2. Here $N = 100$ and $T = 0.5$.