NOISE PARAMETER ESTIMATION IN THE PRESENCE OF RANDOM TIMING ERRORS

Mark R. Morelande

Melbourne Systems Laboratory
Department of Electrical and Electronic Engineering
The University of Melbourne, Australia
m.morelande@ee.unimelb.edu.au

ABSTRACT

Algorithms for estimating the noise parameters of a signal sampled with random timing errors and embedded in additive noise are proposed. Both the timing errors and the additive noise are assumed to be Gaussian and independent and identically distributed. Computationally efficient estimators, derived by maximising an approximation to the likelihood, are proposed. Comparisons with the Cramér-Rao bound demonstrate the performance of the proposed algorithms for a range of parameter values.

1. INTRODUCTION

We consider a continuous-time signal $s(t)$, $t \in \mathbb{R}$ of known functional form which is to be sampled with nominal sampling period T. The sampling procedure is subject to random jitter such that it produces the sequence $\lambda_k(u_k) = s(kT+u_k)$, $k = 1, \ldots, n$ where u_k are iid Gaussian random variables with zero mean and variance τ^2. In the presence of additive noise, the signal model becomes

$$x_k = \lambda_k(u_k) + w_k, \qquad k = 1, \ldots, n, \tag{1}$$

where w_k are iid Gaussian random variables with zero mean and variance σ^2. The goal is to estimate τ^2 and σ^2 from observations $\boldsymbol{x} = (x_1, \ldots, x_n)'$. This problem is of interest, for example, when using high-speed sampling oscilloscopes to test input signals for compliance with relevant standards [1, 4]. In this application an accurate characterisation of the uncertainties in the measuring device are essential.

Asymptotically unbiased and efficient estimates can be obtained by maximising the likelihood $p(\boldsymbol{x}|\tau^2, \sigma^2)$ [5]. The popularity of the maximum likelihood estimator (MLE) is due to the primarily empirical evidence which suggests that these desirable asymptotic properties are often satisfied with sufficient closeness for reasonable sample sizes. Computation of the MLEs for the current problem is complicated by the intractability of the likelihood. One possibility is to use the expectation-maximisation (EM) algorithm [3] to compute the MLEs. However, the obvious formulation in which $\boldsymbol{x}$ is the observed data and $\boldsymbol{u} = (u_1, \ldots, u_n)$ is the unobserved data does not prevent the appearance of intractable integrals. It is possible that an alternative formulation may remove the problem. This topic will not be pursued here.

The search for suitable estimators of the variances is formulated in terms of finding an accurate approximation of the likelihood. This approach is motivated by the intuition that maximising a sufficiently accurate approximation of the likelihood will produce estimators with properties close to those of the MLEs. The first estimator uses a Gaussian mixture likelihood approximation with the approximation becoming exact as the number of components in the mixture tends to infinity. A second class of estimators approximates the likelihood by a single Gaussian with mean and variance dependent on the unknown parameters. Estimators in this second class sacrifice fidelity in the likelihood approximation for reduced computational expense.

A similar approach was taken in [6] where approximate MLEs were derived under a small jitter assumption. In the framework adopted here, these estimators belong to the second class of estimators and are obtained by linearising the signal model (1) about the expected value of the jitter. A more accurate single Gaussian likelihood approximation based on Gaussian quadrature is proposed here. The performances of all estimators are compared with the Cramér-Rao bound (CRB) to assess the effects of the various approximations for a range of actual parameter values.

The paper is organised as follows. Approximate ML estimators are developed in Sections 2 and 3. The CRB is derived in Section 4 and a performance comparison is given in Section 5. Conclusions are given in Section 6.

2. GAUSSIAN MIXTURE APPROXIMATION

The likelihood of the observed data vector $\boldsymbol{x}$ given the parameter $\boldsymbol{\theta} = (\tau^2, \sigma^2)'$ may be written as

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{k=1}^{n} \int N(x_k; \lambda_k(u_k), \sigma^2) N(u_k; 0, \tau^2)\, du_k, \tag{2}$$

where

$$N(z;\lambda,\kappa^2) = \exp\{-(z-\lambda)^2/(2\kappa^2)\}/\sqrt{2\pi\kappa^2}. \quad (3)$$

Analytic expressions for the integrals of (2) can be found only if the underlying signal s is an affine function of time, i.e., $s(t) = at + b$ where a and b are real-valued constants. For cases of more general interest the likelihood cannot be evaluated exactly. In such cases a simple scheme suitable for approximating integrals of the form (2) is Gaussian quadrature. The m-point Gaussian quadrature rule leads to the following approximation for the likelihood of the kth observation, $k = 1, \ldots, n$, [2]:

$$p(x_k|\boldsymbol{\theta}) \approx \sum_{i=1}^{m} \rho_i N(x_k; \lambda_k(\upsilon_i), \sigma^2), \quad (4)$$

where $\upsilon_i = \sqrt{2}\tau\xi_i$, $i = 1, \ldots, m$ with ξ_i the ith root of the mth order Hermite polynomial H_m and

$$\rho_i = 2^{m-1} m! / \{m H_{m-1}(\xi_i)\}^2. \quad (5)$$

Hermite polynomials can be generated using, for $m = 1, 2, \ldots,$

$$H_m(u) = 2uH_{m-1}(u) - 2(m-1)H_{m-2}(u), \quad (6)$$

with $H_{-1}(u) = 0$ and $H_0(u) = 1$. Approximate MLEs are obtained as

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{k=1}^{n} \log\left[\sum_{i=1}^{m} \rho_i N(x_k; \lambda_k(\upsilon_i), \sigma^2)\right]. \quad (7)$$

Newton's method initialised with a rough initial estimate can be used to solve (7) in a computationally efficient manner. The details are omitted for the sake of brevity.

For integrals of the form

$$\int f(z) N(z; \mu, \kappa^2)\, dz, \quad (8)$$

m-point Gaussian quadrature is exact if f is a polynomial of order less than $2m$. For a given order m the accuracy of (4) then depends on how well $N(x; \lambda_k(\sqrt{2}\tau u), \sigma^2)$ is approximated by a polynomial in u of order $2m-1$. In particular, the error in the approximation is bounded by [2]

$$\max_{-\infty<\xi<\infty} \frac{m!}{2^m(2m)!} \left.\frac{d^{2m}\, N(x; \lambda_k(\sqrt{2}\tau u), \sigma^2)}{du^{2m}}\right|_{u=\xi}, \quad (9)$$

provided that λ_k is $2m$-times differentiable. It can be shown that, for fixed m and a given λ_k, the upper bound (9) will tend to increase as σ^2 decreases. Convergence of the approximation (4) as $m \to \infty$ follows from [2, Eq. (3.7.5)].

3. GAUSSIAN LIKELIHOOD APPROXIMATIONS

In this section the likelihood of the kth sample, $k = 1, \ldots, n$, is approximated by a single Gaussian of the form

$$p(x_k|\boldsymbol{\theta}) \approx N(x_k; \hat{x}_k(\boldsymbol{\theta}), \nu_k(\boldsymbol{\theta})). \quad (10)$$

Approximate ML estimates of the variances τ^2 and σ^2 are obtained as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{n} \left[\log\{\nu_k(\boldsymbol{\theta})\} + \frac{\{x_k - \hat{x}_k(\boldsymbol{\theta})\}^2}{\nu_k(\boldsymbol{\theta})}\right]. \quad (11)$$

Newton's method initialised with a rough initial estimate is used solve (11).

3.1. Linearisation

Under the assumption that the jitter variance τ^2 is small, the sampled signal can be approximated by an affine function of the jitter noise. In particular, truncating the Taylor series after the first-order term yields

$$\lambda_k(u_k) \approx s_k + u_k g_k, \quad (12)$$

where $s_k = s(kT)$ and $g_k = ds(t)/dt|_{t=kT}$. The signal s_k obtained from jitter-free sampling is subtracted from the observations to obtain the residuals $\epsilon_k = x_k - s_k$, $k = 1, \ldots, n$. The likelihood of the kth observation can then be written as

$$p(x_k|\boldsymbol{\theta}) \approx \int N(\epsilon_k; u_k g_k, \sigma^2) N(u_k; 0, \tau^2)\, du_k. \quad (13)$$

It can be shown that

$$N(\epsilon_k; u_k g_k, \sigma^2) N(u_k; 0, \tau^2) = N(\epsilon_k; 0, \nu_k(\boldsymbol{\theta})) N\left(u_k; \frac{\tau^2 g_k \epsilon_k}{\nu_k(\boldsymbol{\theta})}, \frac{\sigma^2\tau^2}{\nu_k(\boldsymbol{\theta})}\right), \quad (14)$$

where

$$\nu_k(\boldsymbol{\theta}) = \sigma^2 + \tau^2 g_k^2. \quad (15)$$

Note that the first term of the RHS of (14) is independent of the jitter noise u_k. Therefore substituting (14) into (13) leads to a likelihood approximation of the form (10) with $\hat{x}_k(\boldsymbol{\theta}) = s_k$ and $\nu_k(\boldsymbol{\theta})$ as given in (15). Although formulated differently here this is the approach taken in [6].

3.2. Numerical approximation

The Gaussian approximation to the likelihood (10) is completely specified by the quantities $\hat{x}_k(\boldsymbol{\theta})$ and $\nu_k(\boldsymbol{\theta})$. Intuition suggests that these quantities should be as close as possible to the mean and variance of the observations conditioned on the parameter $\boldsymbol{\theta}$. A more accurate approximation to these quantities than that given by linearisation can be obtained by numerical integration. Specifically, the expected value of the kth sample conditional on $\boldsymbol{\theta}$ is

$$\mathsf{E}\,(x_k|\boldsymbol{\theta}) = \int \lambda_k(u_k) N(u_k; 0, \tau_k)\, du_k. \tag{16}$$

The integral (16) can be approximated efficiently and accurately using Gaussian quadrature. The m-point Gaussian quadrature approximation to the mean is

$$\hat{x}_k(\boldsymbol{\theta}) = \sum_{i=1}^{m} \rho_i \lambda_k(\sqrt{2}\tau\xi_i), \tag{17}$$

where the weights $\rho_1, \ldots, \rho_m$ are given by (5) and the sample point ξ_i, $i = 1, \ldots, m$ is the ith root of the mth order Hermite polynomial H_m. The conditional variance is approximated by

$$\nu_k(\boldsymbol{\theta}) = \sigma^2 + \sum_{i=1}^{m} \rho_i \{\lambda_k(\sqrt{2}\tau\xi_i) - \hat{x}_k(\boldsymbol{\theta})\}^2. \tag{18}$$

Identifying (16) with (8), it can be seen that the accuracy of (17) depends on how well $\lambda_k(u) = s(kT + u)$ is approximated by a polynomial of order $2m - 1$. For signals s which are reasonably smooth this will permit the use of a small value of m thus resulting in a computationally efficient scheme. Convergence of (17) as $m \to \infty$ follows from [2, Eq. (3.7.5)].

4. DERIVATION OF THE CRAMÉR-RAO BOUND

The Cramér-Rao bound (CRB) provides a lower bound on the variance of any unbiased estimator. It is given by the diagonal elements of the inverse of the Fisher information matrix (FIM), $\boldsymbol{J} = -\mathsf{E}\,\nabla_{\boldsymbol{\theta}}\nabla'_{\boldsymbol{\theta}} \log\{p(\boldsymbol{x}|\boldsymbol{\theta})\}$ where $\nabla_{\boldsymbol{\theta}} = (\partial\tau^2, \partial\sigma^2)'$. The FIM can be evaluated as follows.

The likelihood is written as

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{k=1}^{n} \int e(x_k, u_k; \boldsymbol{\theta})\, du_k, \tag{19}$$

where $e(x_k, u_k; \boldsymbol{\theta}) = N(x_k; \lambda_k(u_k), \sigma^2) N(u_k; 0, \tau^2)$. Denoting the ith element of $\boldsymbol{\theta}$ as θ_i, it is straightforward to show that

$$\frac{\partial^2 \log\{p(\boldsymbol{x}|\boldsymbol{\theta})\}}{\theta_i\theta_j} = \sum_{k=1}^{n} \left[\frac{\frac{\partial^2 p(x_k|\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}}{p(x_k|\boldsymbol{\theta})} - \frac{\frac{\partial p(x_k|\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial p(x_k|\boldsymbol{\theta})}{\partial\theta_j}}{p(x_k|\boldsymbol{\theta})^2} \right].$$

Since the derivative can be taken inside the integral, the required first-order partial derivatives are

$$\frac{\partial e(x_k, u_k; \boldsymbol{\theta})}{\partial\sigma^2} = \frac{e(x_k, u_k; \boldsymbol{\theta})}{2\sigma^2} \left(\frac{\epsilon_k(u_k)^2}{\sigma^2} - 1 \right), \tag{20}$$

$$\frac{\partial e(x_k, u_k; \boldsymbol{\theta})}{\partial\tau^2} = \frac{e(x_k, u_k; \boldsymbol{\theta})}{2\tau^2} \left(\frac{u_k^2}{\tau^2} - 1 \right), \tag{21}$$

where $\epsilon_k(u_k) = x_k - \lambda_k(u_k)$. The required second-order partial derivatives can be found as

$$\frac{\partial^2 e(x_k, u_k; \boldsymbol{\theta})}{\partial\sigma^2\partial\sigma^2} = \frac{e(x_k, u_k; \boldsymbol{\theta})}{4\sigma^4} \left(\frac{\epsilon_k(u_k)^4}{\sigma^4} - 6\frac{\epsilon_k(u_k)^2}{\sigma^2} + 3 \right),$$

$$\frac{\partial^2 e(x_k, u_k; \boldsymbol{\theta})}{\partial\tau^2\partial\tau^2} = \frac{e(x_k, u_k; \boldsymbol{\theta})}{4\tau^4} \left(\frac{u_k^4}{\tau^4} - 6\frac{u_k^2}{\tau^2} + 3 \right),$$

$$\frac{\partial^2 e(x_k, u_k; \boldsymbol{\theta})}{\partial\sigma^2\partial\tau^2} = \frac{e(x_k, u_k; \boldsymbol{\theta})}{4\tau^2\sigma^2} \left(\frac{u_k^2}{\tau^2} - 1 \right) \left(\frac{\epsilon_k(u_k)^2}{\sigma^2} - 1 \right).$$

The expected value of the second-order partial derivative with respect to σ^2 can be found as

$$\mathsf{E}\, \frac{\frac{\partial^2 p(x_k|\boldsymbol{\theta})}{\partial\sigma^2\partial\sigma^2}}{p(x_k|\boldsymbol{\theta})} = \frac{1}{4\sigma^4} \int N(u_k; 0, \tau^2) \left(3\frac{\sigma^4}{\sigma^4} - 6\frac{\sigma^2}{\sigma^2} + 3 \right) du_k = 0.$$

In a similar manner it can be shown that

$$\mathsf{E}\, \frac{\partial^2 p(x_k|\boldsymbol{\theta})/\partial\sigma^2\partial\tau^2}{p(x_k|\boldsymbol{\theta})} = 0, \quad \mathsf{E}\, \frac{\partial^2 p(x_k|\boldsymbol{\theta})/\partial\tau^2\partial\tau^2}{p(x_k|\boldsymbol{\theta})} = 0.$$

The remaining expectations, involving products of partial derivatives of the likelihood, are of the form (8) and can be approximated with arbitrary accuracy using a Gaussian quadrature rule with a sufficient number of points. The details are omitted for the sake of brevity.

5. SIMULATION RESULTS

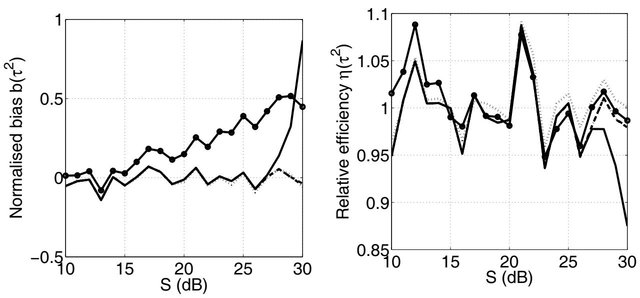
In this section the performances of the algorithms of Sections 2 and 3 are analysed using Monte Carlo simulations. The acronyms GM-MLE, G-MLE(L) and G-MLE(Q) will be used for the estimators of Sections 2, 3.1 and 3.2, respectively.

The continuous-time signal of interest is a sinusoid $s(t) = a\cos(\omega t + \psi)$ which is to be sampled with nominal sampling period $T = 1$s. The sinusoidal parameters are $a = 1$, $\omega = 1$ and $\psi = 0.5$. The sample size is $n = 4096$. The noise parameters τ^2 and σ^2 will be varied between simulation runs with 500 realisations used for each parameter set. Jitter noise variances of $\tau^2 = 0.01, 0.05$ are considered with additive noise variances such that the ratio $S = a^2/\sigma^2$ varies between 10dB and 30dB. Two measures of performance, defined as follows for a scalar parameter θ, are used:

- the normalised bias, $b(\theta) = (\mathsf{E}(\theta) - \theta)/\sqrt{\mathsf{var}(\theta)}$,
- the relative efficiency, $\eta(\theta) = \sqrt{\mathsf{var}(\theta)/\mathsf{CRB}(\theta)}$.

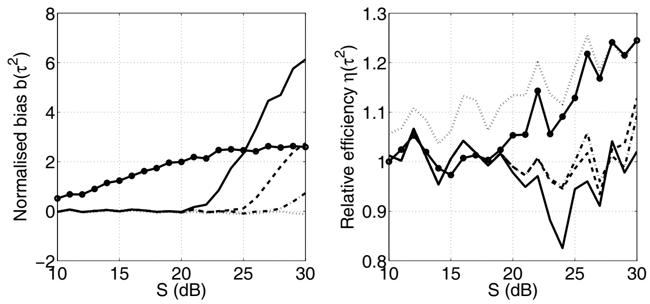
Ideally, the normalised bias is zero and the relative efficiency is one. The performance analysis will assess the degree to which the various estimators depart from these ideal outcomes. Although the relative efficiency is lower-bounded by one for unbiased estimators this is not necessarily the case for biased estimators.

The normalised bias and relative efficiency for estimators of τ^2 are plotted against the ratio S in Figures 1 and 2 for $\tau^2 = 0.01$ and $\tau^2 = 0.05$, respectively. The G-MLE(Q) is implemented with $m = 3$ while the GM-MLE is implemented with $m = 15, 30$ and 60. Initialisation is done via a coarse grid search. For the given scenario, the G-MLE(Q) performs best. The normalised bias is close to zero and relative efficiency is close to one for all parameter values considered. The GM-MLE is the next best estimator although large values of m are required for accurate estimation for small σ^2. This becomes increasingly true as τ^2 increases. The superiority of the G-MLE(Q) over the GM-MLE is surprising given the potentially greater accuracy of the Gaussian mixture approximation compared to the Gaussian approximation. A likely explanation is that the likelihood (2) is more difficult to approximate than the conditional mean (16). The worst estimator is G-MLE(L) as it has significant bias even for the smaller value of τ^2 when σ^2 is small. Similar results are obtained for the estimators of σ^2.



(a) (b)

Figure 1: (a) Normalised bias and (b) relative efficiency of $\hat{\tau}^2$ plotted against $S = a^2/\sigma^2$ (in dB) for the GM-MLE with $m = 15$ (solid), $m = 30$ (dashed) and $m = 60$ (dash-dot), the G-MLE(Q) (dotted) and the G-MLE(L) (–o–). The jitter noise variance $\tau^2 = 0.01$.



(a) (b)

Figure 2: (a) Normalised bias and (b) relative efficiency of $\hat{\tau}^2$ plotted against $S = a^2/\sigma^2$ (in dB) for the GM-MLE with $m = 15$ (solid), $m = 30$ (dashed) and $m = 60$ (dash-dot), the G-MLE(Q) (dotted) and the G-MLE(L) (–o–). The jitter noise variance $\tau^2 = 0.05$.

6. CONCLUSIONS

The problem of estimating the noise parameters of a signal sampled with random timing errors and embedded in additive noise was considered. Since the likelihood is intractable, estimators based on maximising an approximation to the likelihood were proposed. Three likelihood approximations were considered; (1) a Gaussian mixture approximation, (2) a Gaussian approximation with moments calculated by linearisation and (3) a Gaussian approximation with moments calculated by numerical integration. Estimator 3 demonstrated the best performance in a series of Monte Carlo simulations. This performance is obtained with a computational expense about twice that of estimator 2 but about half that of estimator 1. Although Gaussian distributed timing errors were considered here the proposed estimators are not limited to this case. In particular estimators 1 and 3 used Gaussian quadrature rules which can be easily adapted to a number of timing error distributions.

7. REFERENCES

[1] M. Bertocco, L. Garbin, and C. Narduzzi, "Correction of systematic effects in digitizing oscilloscopes," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 3, pp. 871–877, 2003.

[2] P. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed. Academic Press, 1984.

[3] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society* **B**, vol. 39, no. 1, pp. 1–38, 1977.

[4] Y.-C. Jenq, "Digital spectra of nonuniformly sampled signals: Fundamentals and high-speed waveform digitizers," *IEEE Transactions on Instrumentation and Measurement*, vol. 37, no. 2, pp. 245–251, 1988.

[5] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, New Jersey: Prentice Hall, 1993.

[6] G. Vandersteen and R. Pintelon, "Maximum likelihood estimator for jitter noise models," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 6, pp. 1282–1283, December 2000.