# Stochastic Cramer Rao Bounds for Non-Gaussian Signals and Parameters

Weibo Liang, Michael T. Manry, Qiang Yu, Michael S. Dawson, and Adrian K. Fung

Department of Electrical Engineering
University of Texas at Arlington
Arlington, Texas 76019

## ABSTRACT

In minimum mean square estimation, an estimate $\theta'$ of the random parameter vector $\theta$ is obtained from an input vector y. In this paper, we develop bounds on the variances of elements of $\theta'$-$\theta$ for the case where input signal vector y and the parameter vector $\theta$ are non-Gaussian. First, we use linear transformations to obtain a new parameter vector $\phi$ from $\theta$ and a new input vector x from y. These new vectors are approximately Gaussian because of the central limit theorem, so stochastic Cramer-Rao bounds on the variance of $\phi'$ - $\phi$ are tight. Lastly, bounds on variances of elements of $\theta'$-$\theta$ are obtained.

## I. INTRODUCTION

In minimum mean square estimation [1], a random parameter vector $\theta$ is to be estimated from a noisy input vector y. For the additive noise case, elements of y are modelled as

$$y(n) = s'(\theta,n) + e(n) \tag{1}$$

where $s'(\theta,n)$ is the nth element of the N-dimensional signal vector $s'(\theta)$ and e(n) is additive noise with covariance matrix $C_e$. The independent variable $n$ may or may not represent discrete time.

Neural networks can be used to estimate $\theta$ from y. Recently, it has been shown that the training error for such neural network estimators is minimized when the neural network approximates the minimum mean square estimator $E[\theta|y]$ [2]. For the case where both the input signal and parameter vectors are jointly Gaussian, the performance of the estimator is easily characterized by the stochastic Cramer-Rao bound, which is obtained from the stochastic Fisher information matrix (FIM), $J_\theta^{MAP}$[1,2]. Here the superscript *MAP* denotes maximum a posteriori. In neural network applications, the bounds represent target values for the network training error (mean-squared error). When this target is reached, training can be stopped. Failure of the training error to reach the bounds alerts the user to the fact that further or better training is necessary, or that a larger network is required.

When the elements of $\theta$ are not jointly Gaussian, the Fisher information matrix may be impossible to calculate, or the bounds may be too small. Let $\theta'$ denote an estimate of $\theta$. In this paper, we develop bounds on the variances of elements of $\theta'$-$\theta$ for the case where the input signal y and the parameter vector $\theta$ are non-Gaussian.

Let the pseudo stochastic FIM, $\tilde{J}_\theta^{MAP}$, denote a matrix which can be processed into bounds on var($\theta'$-$\theta$) as if it were the stochastic Fisher Information Matrix $J_\theta^{MAP}$. Our goal is to show that in the limit as the dimension of $\theta$ increases, the equation

$$\tilde{J}_\theta^{MAP} = E_\theta[\tilde{J}_\theta^{MLE}] + C_\theta^{-1} \tag{2}$$

is correct for non-Gaussian distributed parameter case, where $C_\theta$ is the covariance matrix of $\theta$ and $J_\theta^{MLE}$ is the regular FIM for nonrandom parameters.

## II. TRANSFORMATIONS OF INPUT AND PARAMETER VECTORS

Assume that the input vector y and parameter vector $\theta$ are non-Gaussian. It is well-known that the stochastic Cramer-Rao bounds are usually not tight for this case. Our goals here are to convert the input vector y and parameter vector $\theta$ to a new input vector x and parameter vector $\phi$ with approximately Gaussian probability density functions (pdfs).



0-7803-2431-5/95 $4.00 © 1995 IEEE

## A. Input Signal Vector

Assume that the input vector **y** is put through a linear transformation, as **x** = **B** • **y**, before it is fed into the estimator. The signal and noise components of **x** are **s**'($\theta$) and **n** respectively. The noise covariance matrix $C_n$ is

$$C_n = B \cdot C_e \cdot B^T$$

The vector **x** is approximately Gaussian because of the central limit theorem [3]. The matrix **B** can be chosen in at least two ways. First, it can be a transformation matrix used for compressing the inputs down to a manageable number, while minimizing the degradation of the estimates [4]. Second, **B** may represent the weights which feed the input vector **y** into net functions [5] of a multilayer perceptron (MLP). The non-stochastic log likelihood function for $\theta$ is

$$\Lambda_\theta^{MLE} = \ln(p_{x|\theta}(x\,|\,\theta)) = C - \frac{1}{2}(x - s(\theta,n))^T C_n^{-1}\ (x - s(\theta,n))$$

where C denotes a constant.

## B. Parameter Vector

Following the same procedure used for the input vectors, we want to transform the parameter vectors as $\phi$ = **A** • $\theta$. The covariance matrix for $\phi$ is easily shown to be

$$C_\phi = A \cdot C_\theta \cdot A^T$$

Here, we want to constrain **A** such that the elements of $\phi$ are approximately statistically independent, and $C_\phi$ is diagonal. We choose the matrix **A** as

$$A = P \cdot S$$

where S denotes a diagonal matrix which normalizes the elements of $\theta$ to unit variance and where the matrix **P** denotes an orthogonal matrix such as the DCT transformation matrix. Clearly then, many matrices **A** exist. As the dimension $M$ of $\theta$ increases, $\phi$ becomes Gaussian via the central limit theorem.

## III. STEPS IN THE DERIVATION

In this section, our goal is to find the FIM for $\phi$ and use it to find an approximate FIM for $\theta$. First, relevant conditional pdfs are found as

$$p_{x|\theta}(x|\theta) = p_n(x - s(\theta))$$
$$p_{x|\phi}(x|\phi) = p_{x|\theta}(x|A^{-1}\phi)$$

The non-stochastic log likelihood function for $\phi$ is now found as

$$\Lambda_\phi^{MLE} = \ln(p_{x|\phi}(x|\phi))$$
$$= C - \frac{1}{2}(x - s(A^{-1}\phi))^T C_n^{-1}(x - s(A^{-1}\phi))$$

where C is a constant. Similarly, the a priori log likelihood function for $\phi$ is

$$\Lambda_\phi^{AP} = \ln(p_\phi)$$

which can be approximated via the central limit theorem as

$$\Lambda_\phi^{AP} \approx -\frac{1}{2}(\phi - m_\phi)^T C_\phi^{-1}(\phi - m_\phi)$$
$$= -\frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M}(\phi_m - m_\phi(m))\ b_\phi(m,n)\ (\phi_n - m_\phi(n)) \tag{3}$$

where $b_\phi(m,n)$ denotes an element of matrix $C_\phi^{-1}$ and $m_\phi(m)$ denotes an element of the mean vector $\mathbf{m}_\phi$ .

The next step in the derivation is to find the stochastic FIM $J_\phi^{MAP}$ for parameter vector $\phi$. Elements of $J_\phi^{MAP}$ are found as

$$J_\phi(i,j) = E_\phi[E_x[\frac{\partial\Lambda_\phi^{MLE}}{\partial\phi_i}\frac{\partial\Lambda_\phi^{MLE}}{\partial\phi_j}]\ ] + E_\phi[\frac{\partial\Lambda_\phi^{AP}}{\partial\phi_i}\frac{\partial\Lambda_\phi^{AP}}{\partial\phi_j}]$$

Using,

$$\frac{\partial\Lambda_\phi^{MLE}}{\partial\phi_i} = (\frac{\partial s(A^{-1}\phi)}{\partial\phi_i})^T C_n^{-1}(x - s(A^{-1}\phi))/$$

$$\frac{\partial\Lambda_\phi^{MLE}}{\partial\phi_j} = (x - s(A^{-1}\phi)^T C_n^{-1}(\frac{\partial s(A^{-1}\phi)}{\partial\phi_j})$$

the first part of the FIM is found as

$$E_\phi[E_x(\frac{\partial\Lambda_\phi^{MLE}}{\partial\phi_i})(\frac{\partial\Lambda_\phi^{MLE}}{\partial\phi_j})]$$
$$= E_\phi[(\frac{\partial s(A^{-1}\phi)}{\partial\phi_i})^T C_n^{-1} E_x[(x - s(A^{-1}\phi))(x - s(A^{-1}\phi))^T] C_n^{-1}(\frac{\partial s(A^{-1}\phi}{\partial\phi_j}$$
$$= E_\phi[(\frac{\partial s(A^{-1}\phi)}{\partial\phi_i})^T C_n^{-1}(\frac{\partial s(A^{-1}\phi)}{\partial\phi_j})]$$
$$= E_\phi[(\sum_{k=1}^{n}(\frac{\partial s}{\partial\theta_k})^T d_{ki})C_n^{-1}(\sum_{m=1}^{n}(\frac{\partial s}{\partial\theta_m})d_{mj})]$$
$$= \sum_{k=1}^{n}\sum_{m=1}^{n} d_{ki}d_{mj}E_\theta[(\frac{\partial s(\theta)}{\partial\theta_k})^T C_n^{-1}(\frac{\partial s(\theta)}{\partial\theta_m})]$$

where $d_{mj}$ denotes an element of $A^{-1}$.

Using the Gaussian approximation of the pdf of $\phi$,

$$\frac{\partial\Lambda_\phi^{AP}}{\partial\phi_j} \approx -\sum_{m=1}^{M} b_\phi(j,m)\ (\phi_m - m_\phi(m))$$

$$\frac{\partial \Lambda_{\phi}^{AP}}{\partial \phi_i} \approx -\frac{1}{2}\sum_{n=1}^{M} b_{\phi}(i,n)(\phi_n - m_{\phi}(n)) - \frac{1}{2}\sum_{n=1}^{M} (\phi_n - m_{\phi}(n)) b_{\phi}(i,n)$$
$$= -\sum_{n=1}^{M} b_{\phi}(i,n)(\phi_n - m_{\phi}(n))$$

Elements of the a priori part of the stochastic FIM are found as

$$E_{\phi}[\frac{\partial \Lambda_{\phi}^{AP}}{\partial \phi_i}\frac{\partial \Lambda_{\phi}^{AP}}{\partial \phi_j}] \approx \sum_{m=1}^{M}\sum_{n=1}^{M} b_{\phi}(i,m)\ b_{\phi}(j,n)\ c_{\phi}(n,m) = b_{\phi}(i,j)$$

where $c_{\phi}(m,n)$ denotes an element of matrix $\mathbf{C}_{\phi}$.

Next, $\hat{\mathbf{J}}_{\theta}^{MAP}$ is found in terms of $\mathbf{J}_{\theta}^{MLE}$ (the FIM for nonrandom parameters) and $\mathbf{C}_{\theta}$ (the covariance matrix of $\boldsymbol{\theta}$.) as

$$J_{\phi}^{MAP} = (A^T)^{-1} E_{\theta}[J_{\theta}^{MLE}] A^{-1} + (A^T)^{-1} C_{\theta}^{-1} A^{-1}$$
$$= (A^T)^{-1} \hat{J}_{\theta}^{MAP} A^{-1}$$

and

$$\hat{J}_{\theta}^{MAP} = A^T J_{\phi}^{MAP} A = E_{\theta}[J_{\theta}^{MLE}] + C_{\theta}^{-1} \tag{4}$$

Lastly, we need to establish that $\hat{\mathbf{J}}_{\theta}^{MAP}$ has a useful relationship to our estimation error variances. Let $\mathbf{C}_{\theta'-\theta}$ and $\mathbf{C}_{\phi'-\phi}$ denote covariance matrice for $\boldsymbol{\theta}'-\boldsymbol{\theta}$ and $\boldsymbol{\phi}'-\boldsymbol{\phi}$ respectively. We can get

$$C_{\phi'-\phi} = A \cdot C_{\theta'-\theta} \cdot A^T$$

It is obvious that off-diagonal elements of $\mathbf{C}_{\theta'-\theta}$ and $\mathbf{C}_{\phi'-\phi}$ are zero. The eestimation error variances can be written as

$$Var(\phi'_i - \phi_i) = [\ C_{\phi'-\phi}\ ]_{ii}$$
$$Var(\theta'_i - \theta_i) = [\ C_{\theta'-\theta}\ ]_{ii}$$

where $[\mathbf{W}]_{ii}$ denotes the ith diagonal element of the matrix $\mathbf{W}$.

Cramer-Rao bounds for parameter vector $\boldsymbol{\phi}$ are diagonal elements of $(\mathbf{J}_{\phi}^{MAP})^{-1}$ so

$$Var(\phi'_i - \phi) \geq [\ (J_{\phi}^{MAP})^{-1}\ ]_{ii}$$

$$[\ C_{\phi'-\phi}\ ]_{ii} \geq [\ (A^T)^{-1}\ (\hat{J}_{\theta}^{MAP})^{-1}\ A^{-1}\ ]_{ii}$$

$$[\ A\ C_{\theta'-\theta}\ A^T\ ]_{ii} \geq [A\ (\hat{J}_{\theta}^{MAP})^{-1}\ A^T\ ]_{ii}$$

$$[\ C_{\theta'-\theta}\ ]_{ii} \geq [\ (\hat{J}_{\theta}^{MAP}\ )^{-1}]_{ii}$$

The expression for this pseudo stochastic FIM $\hat{\mathbf{J}}_{\theta}^{MAP}$ is the same one used when the parameter vector $\boldsymbol{\theta}$ is Gaussian. However, we have now shown that it can be used for non-Gaussian parameter vectors when M is sufficiently large.


## Acknowledgement

This work was funded by NASA under Grant NAGW-3091, by the NSF under grant IRI-9216545, by EPRI under grant RP 8030-09, and by a grant from the state of Texas.


## III. REFERENCES


[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory - Part I*, New York, NY: John Wiley and Sons, 1968.

[2] Q. Yu, S.J. Apollo, and M.T. Manry, "MAP Estimation and the Multilayer Perceptron," *Proceedings of the 1993 IEEE Workshop on Neural Networks for Signal Processing*, Linthicum Heights, Maryland, Sept. 6-9, 1993, pp. 30-39.

[3] Berry, A.C. "The accuracy of the Gaussian Approximation to the sums of Independent Variables", *Trans. Amer. Math. Soc.* vol. 49, 1941, pp.122-136.

[4] S.J. Apollo, M.T. Manry, L.S. Allen, and W.D. Lyle, "Optimality of transforms for parameter estimation," *Conference Record of the Twenty-Sixth Annual Asilomar Conference on Signals, Systems, and Computers*, Oct. 1992, vol. 1, pp. 294-298.

[5] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning internal representations by error propagation," in D.E. Rumelhart and J.L. McClelland (Eds.), **Parallel Distributed Processing**, Vol. I, Cambridge, Massachusetts: The MIT Press, 1986.