BAYESIAN CRAMÉR-RAO TYPE BOUND FOR RISK-UNBIASED ESTIMATION WITH DETERMINISTIC NUISANCE PARAMETERS

Shahar Bar and Joseph Tabrikian

Dept. of Electrical and Computer Engineering
Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: shahba@post.bgu.ac.il, joseph@ee.bgu.ac.il

ABSTRACT

In this paper, we derive a Bayesian Cramér-Rao type bound in the presence of unknown nuisance deterministic parameters. The most popular bound for parameter estimation problems which involves both deterministic and random parameters is the hybrid Cramér-Rao bound (HCRB). This bound is very useful especially, when one is interested in both the deterministic and random parameters and in the coupling between their estimation errors. The HCRB imposes locally unbiasedness for the deterministic parameters. However, in many signal processing applications, the unknown deterministic parameters are treated as nuisance, and it is unnecessary to impose unbiasedness on these parameters. In this work, we establish a new Cramér-Rao type bound on the mean square error (MSE) of Bayesian estimators with no unbiasedness condition on the nuisance parameters. Alternatively, we impose unbiasedness in the Lehmann sense for a risk that measures the distance between the estimator and the minimum MSE estimator which assumes perfect knowledge of the nuisance parameters. The proposed bound is compared to the HCRB and MSE of Bayesian estimators with maximum likelihood estimates for the nuisance parameters. Simulations show that the proposed bound provides tighter lower bound for these estimators.

Index Terms— Bayesian Cramér-Rao bound, hybrid Cramér-Rao bound, Lehmann unbiasedness, Risk unbiasedness, nuisance parameters, MSE

1. INTRODUCTION

The Cramér-Rao bound (CRB), introduced in [1, 2], is a milestone for lower bounds on the MSE in non-Bayesian parameter estimation. It is commonly used as a benchmark for asymptotic performance, i.e. for a large number of observations and/or a large signal-to-noise ratio (SNR), where the estimation errors tend to fade. In [3] a Cramér-Rao type bound was established for the Bayesian framework, known as the Bayesian CRB (BCRB). The combination of both scenarios, where the parameter vector is composed of both random and deterministic parameters was first introduced in the context of array processing [4], and then reexamined in a general framework in [5]. The attractivity of this bound, which was named hybrid Cramér-Rao bound (HCRB), comes from its computational efficiency, since it does not involve marginalization of the random parameters.

The problem of non-Bayesian parameter estimation in the presence of random nuisance parameters was investigated in [6–10]. On the other hand, Bayesian estimation based on a family of prior distributions indexed by a set of parameters known as hyperparameters, was thoroughly investigated. This concept is based on three methodologies, which are described in [11, 12]: (1) The approach of hierarchical Bayes assigns a prior distribution to the hyperparameters; (2) The technique of robust Bayes is based on evaluating the performance of an estimator for each member of the prior class, in order to extract an estimator which performs well for the entire class; (3) The approach of empirical Bayes turns to estimation of the hyperparameters from the data.

Only few works in the literature have been focused on the wide context of Bayesian parameter estimation in the presence of deterministic nuisance parameters. This scenario was lately investigated by Yatracos [13], who derived a Cramér-Rao type bound for locally weak-sense unbiased estimators. In [14] the bound was tightened by representing the MSE as a sum of two components: the minimum mean square error (MMSE) and a modified risk which was bounded by Yatracos bound.

Forming a estimator is possible by substituting an estimate of the deterministic nuisance parameters into a Bayesian estimator. This approach is inspired by parametric empirical Bayes [15, 16]. A natural choice would be to substitute the maximum likelihood estimate (MLE) of the nuisance parameter into the MMSE estimator of the Bayesian parameter in order to benefit from the asymptotical properties of the MLE and the optimality of the MMSE estimator. This estimator will be denoted by JMS-ML. Another option is to exploit the joint maximum *a-posteriori* probability-ML (JMAP-ML) estimator. Bar-Shalom [17] was the first to consider this approach in the context of joint state estimation and parameter identification on a discrete-time linear system. This estimator was also considered in [18, 19]. In [20], the relation between the JMAP-ML estimator and the MLE was analyzed for estimation of deterministic vector parameter in additive Gaussian noise.

In order to form the concept of risk-unbiasedness in the context of Bayesian parameter estimation in the presence of deterministic nuisance parameters, two approaches were combined in [21]. The first is Lehmann-unbiasedness criterion [22], which generalizes the conventional mean unbiasedness to arbitrary cost functions, e.g. [23, 24]. The second is the concept of risk-unbiasedness prediction [25], which analyzes a new criterion for unbiasedness of Bayesian estimators when a deterministic nuisance parameter is involved.

In this paper, the problem of Bayesian parameter estimation in the presence of deterministic nuisance parameters is addressed and a new Bayesian lower bound on the MSE is derived using the covariance inequality [26, p. 113] and the concept of risk-unbiasedness. The HCRB imposes mean-unbiasedness of both random and deterministic parameters [27]. The question which arises is: why is it necessary to impose unbiasedness condition on the nuisance parameters. On the contrast, risk-unbiasedness imposes an unbiasedness condition in the Lehmann sense w.r.t. the MSE rather than unbiasedness on the nuisance parameters.

2. RISK-UNBIASEDNESS

Let $(\Omega_{\mathbf{x}} \times \Omega_{\varphi}, \mathcal{F}, P_{\boldsymbol{\theta}})$ denote a probability space where $\Omega_{\mathbf{x}} \subseteq \mathbb{R}^N$ is the observation space, $\Omega_{\varphi} \subseteq \mathbb{R}$ is the parameter space, $\mathcal{F}$ is the σ-algebra on $\Omega_{\mathbf{x}} \times \Omega_{\varphi}$, and $\{P_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta}$ is a family of parameterized probability measures, such that the probability space has a finite second statistical moment w.r.t. $P_{\boldsymbol{\theta}}$. Let $\boldsymbol{\psi} = [\varphi, \boldsymbol{\theta}^T]^T$ be an unknown vector parameter, where the parameter of interest, $\varphi \in \Omega_{\varphi}$, is a random variable and the vector of nuisance parameters, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^M$ is considered as deterministic. We are interested to estimate the parameter of interest φ based on the random observation vector $\mathbf{x} \in \Omega_{\mathbf{x}}$. Let $f(\mathbf{x}, \varphi; \boldsymbol{\theta})$, $f(\varphi|\mathbf{x}; \boldsymbol{\theta})$, $f(\mathbf{x}; \boldsymbol{\theta})$, and $f_{\varphi}(\varphi; \boldsymbol{\theta})$ denote the joint, the posterior, the observation, and the prior probability density functions (pdf's), respectively. The function $\hat{\varphi}(\mathbf{x})$ is an estimator of φ with estimation error $\epsilon = \hat{\varphi}(\mathbf{x}) - \varphi$. $\mathrm{E}_{\boldsymbol{\theta}}[\cdot]$ and $\mathrm{E}_{\boldsymbol{\theta}}[\cdot|\mathbf{x}]$ stand for the expectation operator w.r.t. $f(\mathbf{x}, \varphi; \boldsymbol{\theta})$ and $f(\varphi|\mathbf{x}; \boldsymbol{\theta})$, respectively. The column vector of the gradient operator w.r.t. $\boldsymbol{\theta}$ is denoted by $\frac{\partial}{\partial^T \boldsymbol{\theta}}$, and the Hessian matrix operator w.r.t. $\boldsymbol{\theta}$ is denoted by $\frac{\partial^2}{\partial\boldsymbol{\theta}\partial^T\boldsymbol{\theta}}$.

Under the MSE criterion, the risk is defined as $L(\hat{\varphi}, \boldsymbol{\theta}_t) = \mathrm{E}_{\boldsymbol{\theta}_t}[\epsilon^2]$, where $\boldsymbol{\theta}_t$ denotes the true value of $\boldsymbol{\theta}$. When $\boldsymbol{\theta}_t$ is known, the MMSE estimator is given by the conditional mean, $\hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta}_t) = \mathrm{E}_{\boldsymbol{\theta}_t}[\varphi|\mathbf{x}]$. The estimation error of the MMSE estimator is $\epsilon_{MS}(\mathbf{x}, \boldsymbol{\theta}_t) = \hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta}_t) - \varphi$. If $\boldsymbol{\theta}_t$ is unknown, then $\hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta}_t)$ is not a valid estimator of φ and the conventional Bayesian MSE bounds are not tight.

The JMS-ML estimator is given by $\hat{\varphi}_{JMS-ML}(\mathbf{x}) = \hat{\varphi}_{MS}(\mathbf{x}, \hat{\boldsymbol{\theta}}_{ML})$, where $\hat{\boldsymbol{\theta}}_{ML}$ is the MLE of $\boldsymbol{\theta}$, given by $\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} f(\mathbf{x}; \boldsymbol{\theta})$.

The JMAP-ML estimator [20] is given by $\hat{\varphi}_{JMAP-ML}(\mathbf{x}) = \arg\max_{\varphi} \left[\max_{\boldsymbol{\theta}} \log f(\mathbf{x}, \varphi; \boldsymbol{\theta}) \right]$.

By utilizing the orthogonality principle of the MMSE estimator, the MSE risk can be expressed as:

$$L(\hat{\varphi}, \boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}[(\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta}))^2] + \mathrm{E}_{\boldsymbol{\theta}}[\epsilon_{MS}^2(\mathbf{x}, \boldsymbol{\theta})]. \quad (1)$$

The term $\mathrm{E}_{\boldsymbol{\theta}}[\epsilon_{MS}^2(\mathbf{x}, \boldsymbol{\theta})]$ in the right hand side (r.h.s.) of (1) is independent of $\hat{\varphi}(\mathbf{x})$. Accordingly, as in [21], we focus on a modified risk, defined by the first term in the r.h.s. of (1), which measures the "closeness" between the estimator $\hat{\varphi}(\mathbf{x})$ to the optimal estimation procedure when $\boldsymbol{\theta}$ is known:

$$R(\hat{\varphi}, \boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}[(\hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta}))^2]. \quad (2)$$

The modified estimation error is now defined as $z_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta}) \triangleq \hat{\varphi}(\mathbf{x}) - \hat{\varphi}_{MS}(\mathbf{x}, \boldsymbol{\theta})$ and accordingly, the modified cost function is $r_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta}) = z_{\hat{\varphi}}^2(\mathbf{x}, \boldsymbol{\theta})$.

In order to provide an appropriate unbiasedness criterion for the modified cost function, we utilize Lehmann's concept of unbiasedness, which was first introduced in the context of arbitrary cost functions in the non-Bayesian framework [22].

Consider the following criterion for optimal parameter estimation $\hat{\boldsymbol{\theta}}_{opt} = \arg\min_{\hat{\boldsymbol{\theta}}\in S} \mathrm{E}_{\boldsymbol{\theta}}[C(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\theta})]$ where S denotes a given subspace of estimators.

Definition 1. *The estimator $\hat{\boldsymbol{\theta}}(\mathbf{x})$ is said to be point-wise unbiased at $\boldsymbol{\theta}$ in the Lehmann sense w.r.t. the cost function $C(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\theta})$ if*

$$\mathrm{E}_{\boldsymbol{\theta}}[C(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\theta})] \leq \mathrm{E}_{\boldsymbol{\theta}}[C(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\eta})], \ \forall \boldsymbol{\eta} \in \boldsymbol{\Theta}. \quad (3)$$

If (3) *is valid for all values of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then $\hat{\boldsymbol{\theta}}$ is said to be uniformly unbiased.*

The Lehmann-unbiasedness definition implies that an estimator is unbiased if on the average it is "closest" to the true parameter, $\boldsymbol{\theta}$, rather than to any other value in the parameter space, $\boldsymbol{\eta} \in \boldsymbol{\Theta}$. The measure of "closeness" between the estimator and the parameter is the cost function $C(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\theta})$. It is shown in [22] that under the quadratic cost function, $C(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\|^2$, the Lehmann-unbiasedness in (3) is reduced to the conventional mean-unbiasedness, $\mathrm{E}_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}] = \mathbf{0}_M$, where $\mathbf{0}_M$ is a column vector of length M, whose entries are equal to 0. Applying Lehmann-unbiasedness condition to the risk in (2), leads to the following definition.

Definition 2. *The estimator $\hat{\varphi}(\mathbf{x})$ is said to be point-wise risk-unbiased at $\boldsymbol{\theta}$ if*

$$\mathrm{E}_{\boldsymbol{\theta}}[r_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\theta})] \leq \mathrm{E}_{\boldsymbol{\theta}}[r_{\hat{\varphi}}(\mathbf{x}, \boldsymbol{\eta})], \ \forall \boldsymbol{\eta} \in \boldsymbol{\Theta}. \quad (4)$$

If (4) *is valid for all values of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, then $\hat{\varphi}(\mathbf{x})$ is said to be uniformly risk-unbiased.*

The next theorem states that under some mild regularity conditions, the above definition can be expressed in a simpler manner.

Theorem 1. If $\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\eta})$ is once differentiable w.r.t. $\boldsymbol{\eta}$ for a.e. $\mathbf{x}\in\Omega_{\mathbf{x}}$, a necessary condition for the estimator $\hat{\varphi}(\mathbf{x})$ to be point-wise risk-unbiased at $\boldsymbol{\theta}_t$, is given by:

$$\mathrm{E}_{\boldsymbol{\theta}_t}\left[z_{\hat{\varphi}}(\mathbf{x},\boldsymbol{\theta}_t)\mathbf{d}(\mathbf{x},\boldsymbol{\theta}_t)\right]=\mathbf{0}_M \tag{5}$$

where $\mathbf{d}(\mathbf{x},\boldsymbol{\theta}_t)\triangleq\left.\frac{\partial\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\theta})}{\partial^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_t}$. If, in addition, $\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\eta})$ is twice differentiable w.r.t. $\boldsymbol{\eta}$ and $r(\hat{\varphi},\boldsymbol{\eta})$ is convex in $\boldsymbol{\eta}$ for a.e. $\mathbf{x}\in\Omega_{\mathbf{x}}$, then the condition in (5) is also sufficient.

Proof 1. *Let $\hat{\varphi}(\mathbf{x})$ be point-wise risk-unbiased at $\boldsymbol{\theta}_t$. Then* (4) *implies that $\boldsymbol{\theta}_t=\arg\min_{\boldsymbol{\eta}}\mathrm{E}_{\boldsymbol{\theta}_t}[r(\hat{\varphi},\boldsymbol{\eta})]$. Since $\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\eta})$ is once differentiable w.r.t. $\boldsymbol{\eta}$ and $\mathrm{E}_{\boldsymbol{\theta}_t}[r(\hat{\varphi},\boldsymbol{\eta})]$ depends on $\boldsymbol{\eta}$ only through $\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\eta})$, then $\mathrm{E}_{\boldsymbol{\theta}_t}[r(\hat{\varphi},\boldsymbol{\eta})]$ is also once differentiable. Taking its derivative w.r.t. $\boldsymbol{\eta}$ yields the necessary condition for the estimator $\hat{\varphi}(\mathbf{x})$ to be point-wise risk-unbiased at $\boldsymbol{\theta}_t$:*

$$\begin{aligned}\left.\frac{\partial\mathrm{E}_{\boldsymbol{\theta}_t}\left[r(\hat{\varphi},\boldsymbol{\eta})\right]}{\partial^T\boldsymbol{\eta}}\right|_{\boldsymbol{\theta}_t}&=\left.\frac{\partial\mathrm{E}_{\boldsymbol{\theta}_t}\left[(\hat{\varphi}(\mathbf{x})-\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\eta}))^2\right]}{\partial^T\boldsymbol{\eta}}\right|_{\boldsymbol{\theta}_t}\\&=\mathrm{E}_{\boldsymbol{\theta}_t}\left[z(\hat{\varphi},\mathbf{x},\boldsymbol{\theta}_t)\mathbf{d}(\mathbf{x},\boldsymbol{\theta}_t)\right]=\mathbf{0}_M.\end{aligned} \tag{6}$$

Suppose that (5) holds and that $r(\hat{\varphi},\boldsymbol{\eta})$ is convex on $\boldsymbol{\Theta}$ for a.e. $\mathbf{x}\in\Omega_{\mathbf{x}}$. Then, using Taylor's theorem:

$$\begin{aligned}&\forall\boldsymbol{\eta}\in\boldsymbol{\Theta},\exists c(\boldsymbol{\eta})\in[0,1]:\mathrm{E}_{\boldsymbol{\theta}_t}[r(\hat{\varphi},\boldsymbol{\eta})]-\mathrm{E}_{\boldsymbol{\theta}_t}[r(\hat{\varphi},\boldsymbol{\theta}_t)]=\\&=(\boldsymbol{\eta}-\boldsymbol{\theta}_t)^T\mathrm{E}_{\boldsymbol{\theta}_t}\left.\left[\frac{\partial^2 r(\hat{\varphi},\boldsymbol{\zeta})}{\partial\boldsymbol{\zeta}\partial^T\boldsymbol{\zeta}}\right]\right|_{\zeta=c\boldsymbol{\theta}_t+(1-c)\boldsymbol{\eta}}(\boldsymbol{\eta}-\boldsymbol{\theta}_t)\geq 0\end{aligned} \tag{7}$$

where the last inequality stems from the property of convexity.

■

Thus, equation (5) can be utilized as an alternative definition for point-wise risk-unbiasedness.

If $\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\theta})$ is twice differentiable w.r.t. $\boldsymbol{\theta}$, then the left hand side (l.h.s.) of (5) is once differentiable. Taking the derivative of both sides of (5) in $\boldsymbol{\theta}_t$ yields the following definition.

Definition 3. *The estimator $\hat{\varphi}(\mathbf{x})$ is said to be locally risk-unbiased around $\boldsymbol{\theta}_t$ if* (5) *holds and*

$$\mathrm{E}_{\boldsymbol{\theta}_t}[z(\hat{\varphi},\mathbf{x},\boldsymbol{\theta}_t)\mathbf{H}(\mathbf{x},\boldsymbol{\theta}_t)]=\mathbf{A}(\boldsymbol{\theta}_t) \tag{8}$$

where $\mathbf{A}(\boldsymbol{\theta}_t)\triangleq\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{d}(\mathbf{x},\boldsymbol{\theta}_t)\mathbf{d}^T(\mathbf{x},\boldsymbol{\theta}_t)]$ and
$\mathbf{H}(\mathbf{x},\boldsymbol{\theta}_t)\triangleq\left.\frac{\partial^2\hat{\varphi}_{MS}(\mathbf{x},\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial^T\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_t}+\mathbf{d}(\mathbf{x},\boldsymbol{\theta}_t)\left.\frac{\partial\log f(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_t}$.

3. MSE LOWER BOUND FOR RISK-UNBIASED ESTIMATORS

In this section, we derive a Bayesian lower bound on the MSE of risk-unbiased estimators in the presence of deterministic nuisance parameters. For the simplicity of notations, in the sequel we will use the symbol $\boldsymbol{\theta}$ instead of the true parameter $\boldsymbol{\theta}_t$.

3.1. The Proposed Bound

Given $u:\Omega_{\mathbf{x}}\times\boldsymbol{\Theta}\to\mathbb{R}$ and $v:\Omega_{\mathbf{x}}\times\boldsymbol{\Theta}\to\mathbb{R}^K$ for some $K\in\mathbb{N}$, The covariance inequality [26, p. 113] is given by:

$$\mathrm{E}_{\boldsymbol{\theta}}[u^2]\geq\mathrm{E}_{\boldsymbol{\theta}}[u\mathbf{v}^T]\mathrm{E}_{\boldsymbol{\theta}}^{-1}[\mathbf{v}\mathbf{v}^T]\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{v}u]. \tag{9}$$

By setting $u(\mathbf{x},\boldsymbol{\theta})=z_{\hat{\varphi}}(\mathbf{x},\boldsymbol{\theta})$, the l.h.s. of (9) turns into the modified risk in (2) while the r.h.s. constitutes a lower bound. Denote $\mathbf{h}(\mathbf{x},\boldsymbol{\theta})=\mathrm{vec}(\mathbf{H}(\mathbf{x},\boldsymbol{\theta}))$ where $\mathrm{vec}(\cdot)$ stands for the vectorization operation. Letting $\mathbf{v}(\mathbf{x},\boldsymbol{\theta})=\left[\mathbf{d}^T(\mathbf{x},\boldsymbol{\theta}),\mathbf{h}^T(\mathbf{x},\boldsymbol{\theta})\right]^T$, where and then applying the constraints in (5) and (8) yields the following terms for locally risk-unbiased estimators:

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\theta}}[u(\mathbf{x},\boldsymbol{\theta})\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})]&=\mathrm{E}_{\boldsymbol{\theta}}\left[z_{\hat{\varphi}}(\mathbf{x},\boldsymbol{\theta})\left[\mathbf{d}^T(\mathbf{x},\boldsymbol{\theta}),\mathbf{h}^T(\mathbf{x},\boldsymbol{\theta})\right]\right]\\&=\left[\mathbf{0}_M^T,\mathrm{vec}^T(\mathbf{A}(\boldsymbol{\theta}))\right]\end{aligned} \tag{10}$$

$$\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{v}(\mathbf{x},\boldsymbol{\theta})\mathbf{v}^T(\mathbf{x},\boldsymbol{\theta})]=\begin{bmatrix}\mathbf{A}(\boldsymbol{\theta}) & \mathbf{B}^T(\boldsymbol{\theta})\\ \mathbf{B}(\boldsymbol{\theta}) & \mathbf{C}(\boldsymbol{\theta})\end{bmatrix} \tag{11}$$

where $\mathbf{B}(\boldsymbol{\theta})\triangleq\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{h}(\mathbf{x},\boldsymbol{\theta})\mathbf{d}^T(\mathbf{x},\boldsymbol{\theta})]$ and
$\mathbf{C}(\boldsymbol{\theta})\triangleq\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{h}(\mathbf{x},\boldsymbol{\theta})\mathbf{h}^T(\mathbf{x},\boldsymbol{\theta})]$. Substituting (10) and (11) into (9) forms a lower bound on the modified risk for locally risk-unbiased estimators:

$$\mathrm{E}_{\boldsymbol{\theta}}[z_{\hat{\varphi}}^2(\mathbf{x},\boldsymbol{\theta})]\geq\mathrm{vec}^T(\mathbf{A})(\mathbf{C}-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)^{-1}\mathrm{vec}(\mathbf{A}) \tag{12}$$

where for the simplicity of notations we omitted the dependency of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ on $\boldsymbol{\theta}$. Substituting (12) into (1) results in a lower bound on the MSE of $\hat{\varphi}(\mathbf{x})$:

$$\begin{aligned}L(\hat{\varphi},\boldsymbol{\theta})\geq\;&\mathrm{E}_{\boldsymbol{\theta}}[\epsilon_{MS}^2(\mathbf{x},\boldsymbol{\theta})]\\&+\mathrm{vec}^T(\mathbf{A})(\mathbf{C}-\mathbf{B}\mathbf{A}^{-1}\mathbf{B}^T)^{-1}\mathrm{vec}(\mathbf{A}).\end{aligned} \tag{13}$$

3.2. The Proposed Bound for Scalar Nuisance Parameter

For the case of a scalar nuisance parameter, θ, the bound in (13) can be simplified to

$$L(\hat{\varphi},\theta)\geq\mathrm{E}_{\theta}\left[\epsilon_{MS}^2(\mathbf{x},\theta)\right]+\frac{a^3(\theta)}{a(\theta)c(\theta)-b^2(\theta)} \tag{14}$$

where $a(\theta)\triangleq\mathrm{E}_{\theta}\left[\left(\frac{\partial\hat{\varphi}_{MS}(\mathbf{x},\theta)}{\partial\theta}\right)^2\right]$,
$b(\theta)\triangleq\mathrm{E}_{\theta}\left[\frac{\partial\hat{\varphi}_{MS}(\mathbf{x},\theta)}{\partial\theta}\left(\frac{\partial^2\hat{\varphi}_{MS}(\mathbf{x},\theta)}{\partial\theta^2}+\frac{\partial\hat{\varphi}_{MS}(\mathbf{x},\theta)}{\partial\theta}\frac{\partial\log f(\mathbf{x};\theta)}{\partial\theta}\right)\right]$
and $c(\theta)\triangleq\mathrm{E}_{\theta}\left[\left(\frac{\partial^2\hat{\varphi}_{MS}(\mathbf{x},\theta)}{\partial\theta^2}+\frac{\partial\hat{\varphi}_{MS}(\mathbf{x},\theta)}{\partial\theta}\frac{\partial\log f(\mathbf{x};\theta)}{\partial\theta}\right)^2\right]$.

The form of the bound for scalar nuisance parameter in (14) sheds light over the full form of the bound in (13) and the concept of risk-unbiasedness. The proposed bound is based on the sensitivity of the MMSE estimator to perturbations around $\boldsymbol{\theta}$. Incorporation of the first and second order derivatives with the derivative of the marginal likelihood function, provides a small error bound for estimators which exploit the form of dependency of the MMSE estimator on the deterministic nuisance parameters, as the next section states.

4. EXAMPLE - SIGNAL ESTIMATION

In this example, we examine the problem of random source signal estimation using an array of sensors. Consider the following observation model:

$$\mathbf{x}_n = c\mathbf{a}(\phi)s_n + \mathbf{w}_n,\ n = 1, \ldots, N \tag{15}$$

where $\mathbf{s} = [s_1, \ldots, s_N] \in \mathbb{C}^N$ is a sequence of zero-mean complex proper jointly Gaussian random variables with a known covariance $\sigma_s^2 \mathbf{I}_N$, $c > 0$ is an unknown amplitude, $\mathbf{a}(\phi) \in \mathbb{C}^K$ is a normalized steering vector of a uniform linear array of half wavelength inter element spacing with K elements, and $\{\mathbf{w}_n\}_{n=1}^N$ is a white zero-mean complex proper Gaussian random noise vector sequence with a known covariance matrix $\sigma_w^2 \mathbf{I}_K$. The MMSE estimator of $\varphi_n = s_n$ from $\mathbf{x} = \left[\mathbf{x}_1^T, \ldots, \mathbf{x}_N^T\right]^T$ for an unknown value of $\boldsymbol{\theta} = [\phi, c]^T$ is given by:

$$\hat{s}_{n_{MS}}(\mathbf{x}, \boldsymbol{\theta}) = \frac{c\sigma_s^2}{\sigma_w^2 + c^2\sigma_s^2} \mathbf{a}^H(\phi)\mathbf{x}_n. \tag{16}$$

By applying the Bayes law, the logarithm of the joint pdf of $\mathbf{x}$ and $\mathbf{s}$ is obtained:

$$\begin{aligned}\log f(\mathbf{x}, \mathbf{s}; \boldsymbol{\theta}) &= \log f(\mathbf{x}|\mathbf{s}; \boldsymbol{\theta}) + \log f(\mathbf{s}; \boldsymbol{\theta}) \\ &= -NM \log(\pi\sigma_w^2) - N\log(\pi\sigma_s^2) \\ &\quad - \sum_{n=1}^{N} \frac{\|\mathbf{x}_n - c\mathbf{a}(\phi)s_n\|^2}{\sigma_w^2} - \frac{|s_n|^2}{\sigma_s^2}.\end{aligned} \tag{17}$$

Since the measurements and the parameter of interest are jointly Gaussian, the BCRB equals the MMSE [3]. In addition, the hybrid Fisher information matrix [27] is block diagonal, such that the estimation errors of the elements of $\mathbf{s}$ are uncoupled to each other, and to those of the nuisance parameters. Thus, the BCRB and the HCRB for estimation of the elements of $\mathbf{s}$ are also equal:

$$BCRB(s_n) = HCRB(s_n) = \frac{\sigma_s^2\sigma_w^2}{\sigma_w^2 + c^2\sigma_s^2},\ n = 1, \ldots, N. \tag{18}$$

Using (17), the logarithm of the marginal pdf of the observations can be verified to take the form:

$$\begin{aligned}\log f(\mathbf{x}; \boldsymbol{\theta}) &= -NM\log(\pi\sigma_w^2) - N\log\left(1 + \frac{\sigma_s^2}{\sigma_w^2}\right) \\ &\quad - \sum_{n=1}^{N} \frac{1}{\sigma_w^2}\mathbf{x}_n^H\left(\mathbf{I}_N - \frac{c^2\sigma_s^2\mathbf{a}(\phi)\mathbf{a}^H(\phi)}{\sigma_w^2 + c^2\sigma_s^2}\right)\mathbf{x}_n.\end{aligned} \tag{19}$$

The proposed bound can be computed by using (14), (16), (18), and (19).

For each experiment, the MLE of $\boldsymbol{\theta}$ is obtained by maximizing $\log f(\mathbf{x}; \boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$, such that:

$$\begin{aligned}\hat{\phi}_{ML} &= \arg\max_{\phi} \sum_{n=1}^{N} |\mathbf{a}^H(\phi)\mathbf{x}_n|^2 \\ \hat{c}_{ML}^2 &= \max\left(\frac{\frac{1}{N}\sum_{n=1}^{N}|\mathbf{a}^H(\hat{\phi}_{ML})\mathbf{x}_n|^2 - \sigma_w^2}{\sigma_s^2}, 0\right).\end{aligned} \tag{20}$$

The JMAP-ML estimator can be verified to be identical to the JMS-ML estimator.

The MSE of the JMS-ML, the HCRB, and the proposed bound versus SNR are presented in Fig. 1, where $SNR \triangleq \frac{c^2\sigma_s^2}{\sigma_w^2}$. The MSE was evaluated using 10,000 Monte-Carlo simulations with $\sigma_s^2 = 1$, $N = 20$, $M = 2$, $c = 1$ and $\phi = \frac{\pi}{3}$. $\hat{\phi}_{ML}$ was evaluated using a grid search with a resolution of 10^{-3}. The proposed bound provides a tighter lower bound for the JMS-ML estimator, for all SNR values.

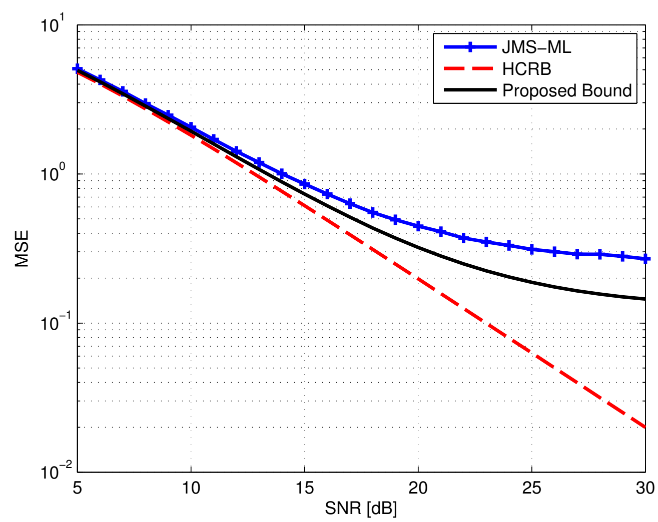

Fig. 1. The HCRB, proposed bound, and MSE of JMS-ML.

5. CONCLUSION

In this paper, the concept of Bayesian parameter estimation in the presence of deterministic nuisance parameters is introduced and a Cramér-Rao type bound for the MSE was developed. Unlike the HCRB, the proposed bound does not assume unbiasedness for the nuisance parameter. The proposed bound assumes risk-unbiasedness which is more appropriate for the case of nuisance parameters. The asymptotic properties of the proposed bound are investigated and its relation to conventional estimators was explored. It was shown that for the problem of Gaussian source signals estimation using an array of sensors, the proposed bound provides a tighter bound than the HCRB on the performance of the JMS-ML estimator.

6. REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, NJ, 1946.

[2] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.

[3] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, vol. 1, Wiley, New York, 1968.

[4] Y. Rockah and P. M. Schultheiss, "Array shape calibration using sources in unknown locations–part I: Far-field sources," *IEEE Audio, Speech, Language Process.*, vol. 35, no. 3, pp. 286–299, 1987.

[5] I. Reuven and H. Messer, "A Barankin-type lower bound on the estimation error of a hybrid parameter vector," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 1084–1093, 1997.

[6] R. Miller and C. Chang, "A modified Cramér-Rao bound and its applications," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 398–400, 1978.

[7] H. Messer and Y. Adar, "Lower bounds on bearing estimation errors of any number of narrowband far-field sources," in *Proc. ICASSP 1989*, 1989, vol. 4, pp. 2113–2116.

[8] F. Gini, R. Reggiannini, and U. Mengali, "The modified Cramér-Rao bound in vector parameter estimation," *IEEE Trans. Commun.*, vol. 46, no. 1, pp. 52–60, 1998.

[9] M. Moeneclaey, "On the true and the modified Cramér-Rao bounds for the estimation of a scalar parameter in the presence of nuisance parameters," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1536–1544, 1998.

[10] Y. Noam and H. Messer, "Notes on the tightness of the hybrid Cramér-Rao lower bound," *IEEE Trans. Signal Processing*, vol. 57, no. 6, pp. 2074–2084, 2009.

[11] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer, 1985.

[12] C. P. Robert, "The Bayesian choice: From decision-theoretic foundations to computational implementation (springer texts in statistics) by," 2001.

[13] Y. G. Yatracos, "On prediction and mean squared error," *Canadian Journal of Statistics*, vol. 20, no. 2, pp. 187–200, 1992.

[14] T. K. Nayak, "On best unbiased prediction and its relationships to unbiased estimation," *Journal of Statistical Planning and Inference*, vol. 84, no. 1, pp. 171–189, 2000.

[15] J. J. Deely and D. V. Lindley, "Bayes empirical Bayes," *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 833–841, 1981.

[16] C. N. Morris, "Parametric empirical Bayes inference: theory and applications," *Journal of the American Statistical Association*, vol. 78, no. 381, pp. 47–55, 1983.

[17] Y. Bar-Shalom, "Optimal simultaneous state estimation and parameter identification in linear discrete-time systems," *IEEE Trans. Autom. Control*, vol. 17, no. 3, pp. 308–319, 1972.

[18] J. Lim and A. Oppenheim, "All-pole modeling of degraded speech," *IEEE Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 197–210, 1978.

[19] B. R. Musicus, "Iterative algorithms for optimal signal reconstruction and parameter identification given noisy and incomplete data," Tech. Rep., DTIC Document, 1982.

[20] A. Yeredor, "The joint MAP-ML criterion and its relation to ML and to extended least-squares," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3484–3492, 2000.

[21] T. K. Nayak and M. Qin, "The concept of risk unbiasedness in statistical prediction," *Journal of Statistical Planning and Inference*, vol. 140, no. 7, pp. 1923–1938, 2010.

[22] E. L. Lehmann, "A general concept of unbiasedness," *The Annals of Mathematical Statistics*, vol. 22, no. 4, pp. 587–592, 1951.

[23] T. Routtenberg and J. Tabrikian, "Periodic CRB for non-Bayesian parameter estimation," in *Proc. ICASSP 2011*, 2011, pp. 2448–2451.

[24] T. Routtenberg and J. Tabrikian, "Non-Bayesian periodic Cramér-Rao bound," *IEEE Trans. Signal Processing*, vol. 61, no. 4, pp. 1019–1032, 2013.

[25] Y. Xiao, "Linex unbiasedness in a prediction problem," *Annals of the Institute of Statistical Mathematics*, vol. 52, no. 4, pp. 712–721, 2000.

[26] E. L. Lehmann and G. Casella, *Theory of point estimation*, vol. 31, Springer, 1998.

[27] H. Messer, "The hybrid Cramér-Rao lower bound-from practice to theory," in *Proc. 2006 Fourth IEEE Workshop on Sensor Array and Multichannel Processing*, 2006, pp. 304–307.