Optimal Biased Estimation Using Lehmann-Unbiasedness

Eyal Nitzan, Tirza Routtenberg, and Joseph Tabrikian
Department of Electrical and Computer Engineering Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
Email: eyalni@ee.bgu.ac.il, tirzar@bgu.ac.il, joseph@bgu.ac.il

***Abstract*—This paper deals with non-Bayesian parameter estimation under the mean-squared-error (MSE), which is a topic of great interest in various engineering fields. Although the unbiasedness condition is commonly used in non-Bayesian MSE estimation, in many cases biased estimation may result in better performance. However, no method for determining the optimal bias function in general cases is available. We propose a new approach for uniform minimum MSE biased estimation, where the optimal bias is chosen in accordance with Lehmann-unbiasedness definition. The proposed approach is based on modifying the MSE risk by its multiplication with a weighting function of the unknown parameter, g^2. Under this modified risk, Lehmann's definition of unbiasedness provides a condition referred to as g-unbiasedness. By using the g-unbiasedness, we derive a novel Cramér-Rao-type lower bound on the MSE of locally g-unbiased estimators. In addition, we show that if there exists an estimator that achieves the new bound, then it is produced by the penalized maximum likelihood estimator with a penalty function $\log g$. Simulations show that the proposed approach can lead to non-trivial estimators with lower MSE than existing mean-unbiased estimators.**

Keywords*—Non-Bayesian parameter estimation, mean-squared-error (MSE), Lehmann-unbiasedness, penalized maximum likelihood, Cramér-Rao bound

I. Introduction

Various engineering applications require estimation of unknown parameters, for example in signal processing, speech processing, and communications. The performance of estimators is commonly evaluated by the mean-squared-error (MSE) [1], [2]. In the non-Bayesian framework, minimizing the MSE without any restrictions yields the trivial estimator, which has a zero risk. Thus, some restrictions on the estimators are often applied. A commonly-used restriction is mean-unbiasedness [2, C. 2], [3], for which the Cramér-Rao bound (CRB) [4], [5] provides a performance benchmark. However, in several estimation problems the set of mean-unbiased estimators is empty [6], [7]. In addition, in many estimation problems mean-unbiased estimators may be "silly" [8, p. 253] or may not be admissible in the MSE sense [9]. Other approaches for non-Bayesian estimation include shrinkage estimation [9], [10], minimax estimation [2, C. 5] [11], and equivariance [2, C. 3], [12].

It is shown in [9], [13]–[18] that in some scenarios, there may exist biased estimators, which uniformly dominate the MSE of the minimum variance unbiased (MVU) estimator. In [19], [20], biased estimation is considered and uniform Cramér-Rao lower bounds (UCRLBs) on the total variance of any estimator with bounded bias gradient norm are derived. It is shown in [20] that under some conditions, the UCRLB is attained asymptotically by a class of penalized maximum likelihood (PML) estimators. The PML estimator [21] is obtained by maximizing a penalized likelihood function and is widely used in many engineering applications (see e.g. [22]–[24]). A comprehensive tutorial on biased estimation can be found in [25]. In addition, a structured approach for obtaining uniformly best biased estimators is proposed in [26]. However, there is no general procedure for finding the optimal bias function in general biased estimation models.

In this work, we propose a new Cramér-Rao-type bound, where the Lehmann-unbiasedness is imposed on a modified risk. The modified risk is the MSE multiplied by the square of a weighting function, g, where g is a function of the unknown deterministic parameter. For this modified risk, we derive the Lehmann's condition for unbiasedness, denoted by g-unbiasedness. Lehmann's unbiasedness [27] is a generalization of the conventional mean-unbiasedness for arbitrary cost functions that has been used in various models [28]–[38]. The g-unbiasedness condition is defined both in the uniform and the local sense. Then, we derive a novel Cramér-Rao-type bound on the MSE of locally g-unbiased estimators that is lower than or equal to the conventional CRB. The proposed bound is lower due to the mean-bias and not due to lack of tightness. It is shown that when an estimator exists that uniformly attains the bound, named as a g-efficient estimator, it coincides with a PML estimator. We derive an ordinary differential equation (ODE) that the weighting function g should satisfy, in order that g-efficient estimators will be locally g-unbiased. Our approach is examined in two examples in which estimators with uniformly lower MSE than the MVU estimator can be found.

II. Problem formulation and definitions

We consider the following estimation problem: Let $(\Omega_{\mathbf{x}}, \mathcal{F}, P_\theta)$ denote a probability space, where $\Omega_{\mathbf{x}}$ is the observation space, $\mathcal{F}$ is the σ-algebra on $\Omega_{\mathbf{x}}$, $\{P_\theta\}$ is a family of probability measures parameterized by the deterministic unknown parameter $\theta \in \Omega_\theta$, and Ω_θ is the parameter space, which is an open subset of $\mathbb{R}$. In this work, we will focus on scalar unknown parameter, θ. Extension to vector parameter is beyond the focus of this paper. The probability measure, P_θ, is assumed to have an associated probability density function, $f_{\mathbf{x}}(\cdot;\theta)$. Expectation with respect to (w.r.t.) P_θ, $\theta \in \Omega_\theta$, is denoted by $\mathrm{E}[\cdot;\theta]$. For simplicity of notations, we omit θ from the notation of expectation and denote it by $\mathrm{E}[\cdot]$, whenever the value of θ is clear from the context. In addition, we denote by $\hat{\theta}: \Omega_{\mathbf{x}} \to \Omega_\theta$ an arbitrary estimator of θ based on a random observation vector $\mathbf{x} \in \Omega_{\mathbf{x}}$.

In the following, we denote vectors by boldface lowercase letters. The kth element of the vector $\mathbf{b}$ is denoted by b_k. The derivative of a function $g(\theta)$ at point θ_0 is denoted by $\frac{\mathrm{d}}{\mathrm{d}\theta}g(\theta)\Big|_{\theta_0}$ or $g'(\theta_0)$.

We define the following g-squared-error (g-SE) cost function

$$g\text{-SE}(\hat{\theta},\theta) \triangleq \left(g(\theta)(\hat{\theta}-\theta)\right)^2, \tag{1}$$

where $g : \Omega_\theta \to \mathbb{R}_+$ is a positive weighting function of differentiability class C^2. The g-MSE of $\hat{\theta}$ at θ is defined as $g\text{-MSE}_{\hat{\theta}}(\theta) \triangleq g^2(\theta)\text{MSE}_{\hat{\theta}}(\theta)$, where $\text{MSE}_{\hat{\theta}}(\theta) \triangleq \mathrm{E}\left[(\hat{\theta}-\theta)^2\right]$ is the MSE of $\hat{\theta}$ evaluated at θ. In the following, the weighting function g is used to define g-unbiasedness.

According to Lehmann's concept of unbiasedness [27], [28], an estimator is unbiased w.r.t. some cost function if on the average it is "closer" to the true parameter rather than to any other value in the parameter space. The closeness between the estimator and the parameter is measured by the chosen cost function, which leads to different unbiasedness definitions for different cost functions [28]–[38].

The generalized Lehmann-unbiasedness definition with an arbitrary cost function can be found, e.g. in Definition 1 in [28]. Based on this generalized definition, an estimator $\hat{\theta}$ of θ is unbiased w.r.t. the g-SE cost function if

$$\mathrm{E}\left[g^2(\theta)(\hat{\theta}-\theta)^2;\theta\right] \leq \mathrm{E}\left[g^2(\eta)(\hat{\theta}-\eta)^2;\theta\right],\ \forall\eta,\theta\in\Omega_\theta. \tag{2}$$

The condition in (2) is equivalent to requiring that the global minimum of $\mathrm{E}\left[g^2(\eta)(\hat{\theta}-\eta)^2;\theta\right]$ w.r.t. η is achieved at $\eta=\theta,\ \forall\theta\in\Omega_\theta$. In the following theorem, we present a necessary condition for an estimator to be unbiased in the Lehmann sense w.r.t. the g-SE cost function. We denote this condition on $\hat{\theta}$ as g-unbiasedness.

Theorem 1. *Let $\hat{\theta}$ be an estimator of θ. Then, a necessary condition for $\hat{\theta}$ to be unbiased in the Lehmann sense w.r.t. the g-SE cost function is*

$$g(\theta)\mathrm{E}[\hat{\theta}-\theta] = g'(\theta)\mathrm{E}\left[(\hat{\theta}-\theta)^2\right],\ \forall\theta\in\Omega_\theta. \tag{3}$$

Proof 1. *Assuming that $\mathrm{E}\left[g^2(\eta)(\hat{\theta}-\eta)^2;\theta\right]$ is a differentiable function of η. Then, a necessary condition for $\hat{\theta}$ to be unbiased in the Lehmann sense w.r.t. the g-SE cost function is that*

$$\frac{\mathrm{d}}{\mathrm{d}\eta}\mathrm{E}\left[g^2(\eta)(\hat{\theta}-\eta)^2;\theta\right]\Big|_{\eta=\theta} = 0. \tag{4}$$

Under the assumption that the expectation and derivative can be interchanged, we can rewrite (4) *as*

$$2g^2(\theta)\mathrm{E}[\hat{\theta}-\theta;\theta] = 2g(\theta)g'(\theta)\mathrm{E}\left[(\hat{\theta}-\theta)^2;\theta\right]. \tag{5}$$

By dividing both sides of (5) *by the positive function* $2g(\theta)$, *one obtains the g-unbiasedness condition from* (3). □

It can be verified that by substituting $g(\theta)=\text{const}$ in (3), one obtains the conventional mean-unbiasedness, $\mathrm{E}[\hat{\theta}-\theta]=0,\ \forall\theta\in\Omega_\theta$. In general, the choice of g imposes a relation between the estimator's bias and MSE.

Theorem 1 provides a uniform unbiasedness condition, i.e. unbiasedness for any $\theta\in\Omega_\theta$. In the non-Bayesian framework, local unbiasedness in which the estimator is assumed to be unbiased only in the vicinity of the true parameter, can be useful as well, e.g. for derivation of the CRB [39]. In the following, we define local g-unbiasedness by using Theorem 1 in the vicinity of an interior point $\theta_0\in\Omega_\theta$.

Definition 1. *An estimator $\hat{\theta}$ of θ is said to be locally g-unbiased at $\theta_0\in\Omega_\theta$ if it satisfies*

$$g(\theta)\mathrm{E}[\hat{\theta}-\theta] = g'(\theta)\mathrm{E}\left[(\hat{\theta}-\theta)^2\right] \tag{6}$$

at $\theta=\theta_0+\delta,\ \forall|\delta|<\varepsilon,\ \varepsilon\to 0$.

Under continuity assumptions, the condition

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(g(\theta)\mathrm{E}[\hat{\theta}-\theta]\right)\Big|_{\theta=\theta_0} = \frac{\mathrm{d}}{\mathrm{d}\theta}\left(g'(\theta)\mathrm{E}\left[(\hat{\theta}-\theta)^2\right]\right)\Big|_{\theta=\theta_0} \tag{7}$$

is a necessary condition for an estimator to satisfy (6) at $\theta=\theta_0+\delta,\ \forall|\delta|<\varepsilon,\ \varepsilon\to 0$. For the sake of simplicity, in the following we refer to an estimator $\hat{\theta}$ as a locally g-unbiased estimator at $\theta_0\in\Omega_\theta$ if it satisfies (7).

III. The g-CRB

In this section, we derive a Cramér-Rao-type lower bound, denoted by g-CRB, on the MSE of locally g-unbiased estimators. In order to obtain a simple estimator-independent bound, we select the weighting functions g such that the left hand side of (7) equals zero. That is,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(g(\theta)\mathrm{E}[\hat{\theta}-\theta;\theta]\right)\Big|_{\theta=\theta_0} = 0, \tag{8}$$

where $\hat{\theta}$ is a locally g-unbiased estimator of θ at $\theta_0\in\Omega_\theta$. In the following theorem we derive the proposed g-CRB.

Theorem 2. *Let $\hat{\theta}$ be a locally g-unbiased estimator of θ at $\theta_0\in\Omega_\theta$ and assume that* (8) *and the following regularity conditions are satisfied:*

C.1) $0<I(\theta)<\infty,\ \forall\theta\in\Omega_\theta$, *where*

$$I(\theta) \triangleq \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f_{\mathbf{x}}(\mathbf{x};\theta)\right)^2\right] \tag{9}$$

is the Fisher information for estimating θ.

C.2) *For any function $h:\Omega_{\mathbf{x}}\times\Omega_\theta\to\mathbb{R}$, which is differentiable w.r.t. θ:* $\int_{\Omega_{\mathbf{x}}}\frac{\partial}{\partial\theta}\left\{h(\mathbf{x},\theta)f_{\mathbf{x}}(\mathbf{x};\theta)\right\}\mathrm{d}\mathbf{x} = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{E}\left[h(x,\theta)\right]$.

Then,

$$MSE_{\hat{\theta}}(\theta_0) \geq B_{g-CRB}(\theta_0), \tag{10}$$

where

$$B_{g-CRB}(\theta_0) \triangleq \frac{1}{I(\theta_0)+\left(\frac{g'(\theta_0)}{g(\theta_0)}\right)^2}. \tag{11}$$

Equality in (10) *is obtained* iff

$$\frac{\partial}{\partial\theta}\log\left(f_{\mathbf{x}}(\mathbf{x};\theta)g(\theta)\right)\Big|_{\theta=\theta_0} = \left(I(\theta_0)+\left(\frac{g'(\theta_0)}{g(\theta_0)}\right)^2\right)(\hat{\theta}-\theta_0). \tag{12}$$

Proof 2. *The proof is based on using the Cauchy-Schwarz inequality (see e.g. [40])*

$$\mathrm{E}\left[\epsilon^2(\mathbf{x},\theta_0)\right] \geq \frac{\mathrm{E}^2\left[\epsilon(\mathbf{x},\theta_0)\eta_g(\mathbf{x},\theta_0)\right]}{\mathrm{E}\left[\eta_g^2(\mathbf{x},\theta_0)\right]}, \tag{13}$$

where $\eta_g(\mathbf{x},\theta) \stackrel{\triangle}{=} \frac{\partial}{\partial\theta}\log\left(f_{\mathbf{x}}(\mathbf{x};\theta)g(\theta)\right)$ *and* $\epsilon(\mathbf{x},\theta) \stackrel{\triangle}{=} \hat{\theta}-\theta$. *We substitute the condition in* (8) *in the right hand side of* (13) *in order to obtain an estimator-independent bound. Due to space limitations, the full proof is omitted.* □

The proposed g-CRB is a valid lower bound on the MSE of any estimator with function g that satisfies (8). It can be seen from (11) that for any θ_0, the g-CRB is lower than or equal to the conventional CRB, $B_{\mathrm{CRB}}(\theta_0) \stackrel{\triangle}{=} \frac{1}{I(\theta_0)}$. Therefore, the g-CRB is a valid bound on the MSE of mean-unbiased estimators. For $g(\theta) = const$, the g-CRB and CRB coincide, $\forall\theta \in \Omega_\theta$.

Let us denote the estimator that satisfies (12) as $\hat{\theta}_{g,\theta_0}$. This estimator attains the g-CRB from (11) at θ_0. The general estimator that attains the g-CRB, $\forall\theta \in \Omega_\theta$, is denoted by $\hat{\theta}_{g,\theta}$. In general, $\hat{\theta}_{g,\theta}$ is a function of θ and therefore, is not a practical estimator. However, in some cases $\hat{\theta}_{g,\theta}$ is not a function of θ and achieves the g-CRB, $\forall\theta \in \Omega_\theta$. In this case, we replace $\hat{\theta}_{g,\theta}$ with $\hat{\theta}_g^{\mathrm{eff}}$ and refer to it as a g-efficient estimator. In the following proposition, we show that in case a g-efficient estimator exists, it is given by the PML estimator [21]:

$$\hat{\theta}_g^{\mathrm{PML}} = \arg\max_{\theta\in\Omega_\theta}\left\{\log f_{\mathbf{x}}(\mathbf{x};\theta) + \log g(\theta)\right\}, \tag{14}$$

where it is assumed that

$$\left.\frac{\partial}{\partial\theta}\log\left(f_{\mathbf{x}}(\mathbf{x};\theta)g(\theta)\right)\right|_{\theta=\hat{\theta}_g^{\mathrm{PML}}} = 0. \tag{15}$$

Proposition 3. *Assume that the regularity conditions C.1-C.2 and* (15) *are satisfied and that* $\hat{\theta}_g^{\mathit{eff}}$ *is a g-efficient estimator. Then,*

$$\hat{\theta}_g^{\mathit{eff}} = \hat{\theta}_g^{\mathit{PML}},\ \forall\mathbf{x}\in\Omega_{\mathbf{x}}. \tag{16}$$

Proof 3. *By using* (12)*, it can be seen that if there exists a g-efficient estimator,* $\hat{\theta}_g^{\mathit{eff}}$*, it satisfies*

$$\frac{\partial}{\partial\theta}\log\left(f_{\mathbf{x}}(\mathbf{x};\theta)g(\theta)\right) = \left(I(\theta)+\left(\frac{g'(\theta)}{g(\theta)}\right)^2\right)(\hat{\theta}_g^{\mathit{eff}}-\theta),\ \forall\theta\in\Omega_\theta. \tag{17}$$

Under Conditions C.1-C.2, by substituting $\theta = \hat{\theta}_g^{\mathit{PML}}$ *in* (17) *and using* (15)*, one obtains*

$$\hat{\theta}_g^{\mathit{eff}} - \hat{\theta}_g^{\mathit{PML}} = 0,\ \forall\mathbf{x}\in\Omega_{\mathbf{x}}, \tag{18}$$

which completes the proof. □

The relation between the g-efficient estimator and the PML estimator in Proposition 3 is similar to the relation between the efficient estimator and the ML estimator [41, p. 68]. In particular, for $g(\theta) = \mathrm{const}$, (18) is reduced to the relation between the conventional efficient estimator and the ML estimator [41, p. 68]. In the following section we choose the weighting function such that a g-efficient estimator will be locally g-unbiased.

IV. Choosing the Weighting Function

It can be seen that the g-CRB from (11) depends on the weighting function, g. In order for the g-CRB to be practical, we need to choose g such that the bound will be achievable or closely approximated by a valid estimator, $\forall\theta\in\Omega_\theta$. In general, we would like to consider a family of weighting functions for which g-efficient estimators exist and then choose the efficient estimator that results in the uniformly lowest g-CRB.

We use the local g-unbiasedness condition from (7) under the assumption that a g-efficient estimator exists. The MSE of the g-efficient estimator is equal to the g-CRB, $\forall\theta\in\Omega_\theta$. By substituting the g-CRB from (11) in the right hand side of (7) and using (8), one obtains

$$\left.\frac{\mathrm{d}}{\mathrm{d}\theta}\left(g'(\theta)\frac{1}{I(\theta)+\left(\frac{g'(\theta)}{g(\theta)}\right)^2}\right)\right|_{\theta=\theta_0} = 0. \tag{19}$$

We require that (19) is satisfied, $\forall\theta_0\in\Omega_\theta$ and obtain an ODE that g must satisfy in order to be a member of the family of weighting functions to be considered:

$$g'(\theta) - c\left(I(\theta)+\left(\frac{g'(\theta)}{g(\theta)}\right)^2\right) = 0,\ c\in\mathbb{R}. \tag{20}$$

The ODE in (20) generates a number of valid weighting functions. Given a valid weighting function g, we can attempt to find a g-efficient estimator by substituting g in (12). In case a g-efficient estimator cannot be found, we can consider specific solutions of (20) and derive their corresponding g-unbiasedness. If a g-unbiased estimator with weighting function g that satisfies (8), $\forall\theta_0\in\Omega_\theta$, can be found, then we can examine this estimator's proximity to the corresponding g-CRB. It can be seen that $g(\theta) = \mathrm{const}$ solves (20) for $c = 0$ and is always a valid weighting function. As stated in previous sections, for $g(\theta) = \mathrm{const}$ we return to the conventional mean-unbiasedness and CRB. In the following section we will show two examples in which the MSE can be reduced by using non-constant weighting function and shrinkage estimators that are shown to be g-unbiased estimators.

V. Examples

A. Example 1 -Variance estimation of a Gaussian process

We consider a set of N observations, x_n, $n = 1,\dots,N$, that are independent identically distributed Gaussian random variables with zero-mean and variance θ, where the variance, $\theta\in\mathbb{R}_+$, is the parameter to be estimated. In this case, for $g(\theta) = g_1(\theta) = \mathrm{const}$ we can obtain an efficient/MVU estimator given by $\hat{\theta}_{g_1}^{\mathrm{eff}} = \frac{1}{N}\sum_{n=1}^N x_n^2$. It can be shown that $g(\theta) = g_2(\theta) = \frac{1}{\theta}$ is also a valid solution of the ODE in (20) and that we can find a g_2-efficient estimator given by $\hat{\theta}_{g_2}^{\mathrm{eff}} = \frac{1}{N+2}\sum_{n=1}^N x_n^2$, which coincides with the PML estimator from (14) with $g = g_2$. This estimator also coincides with the shrinkage estimator for this case and has the lowest MSE among estimators with linear bias (see [16], [17]).

In Fig. 1, the MSEs of $\hat{\theta}_{g_1}^{\mathrm{eff}}$ and $\hat{\theta}_{g_2}^{\mathrm{eff}}$ are evaluated by using 10^6 Monte-Carlo simulations and presented versus the number of observations, N, for $\theta = 4$. The estimator $\hat{\theta}_{g_1}^{\mathrm{eff}}$ is compared to the g_1-CRB, which coincides with the conventional CRB in this case, while the estimator $\hat{\theta}_{g_2}^{\text{eff}}$ is compared to the g_2-CRB. The MSEs of the estimators and the corresponding bounds are normalized by the CRB. We denote the normalized MSE by NMSE. It can be seen that both estimators achieve their associated bounds and that $\hat{\theta}_{g_2}^{\text{eff}}$ has a lower NMSE for any N. Finally, the NMSE of $\hat{\theta}_{g_2}^{\text{eff}}$ approaches the NMSE of $\hat{\theta}_{g_1}^{\text{eff}}$ as N grows.

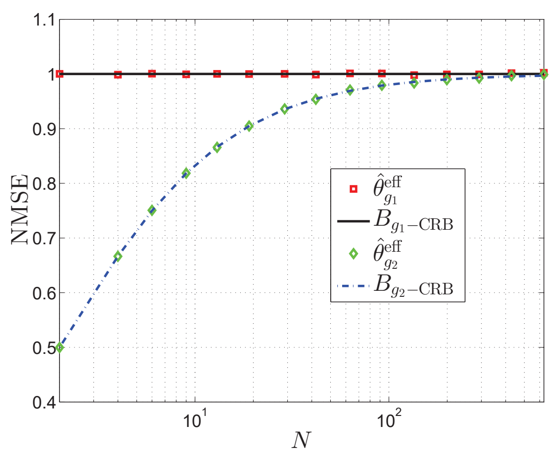

Fig. 1. Variance estimation of a Gaussian process: The NMSEs of the g_i-efficient estimators compared to the normalized g_i-CRB, $i = 1, 2$, where the normalization is by the CRB.

B. Example 2 - Parameter estimation with nonlinear model and additive Gaussian noise

We consider parameter estimation with additive white Gaussian noise, according to the following observation model:

$$x_n = a(\theta) + u_n, \quad n = 1, \dots, N, \tag{21}$$

where the sequence $u_1, \dots, u_N$ is white Gaussian noise with known variance, σ_{u}^2, $a(\theta) = \log\theta$, and $\theta \in \mathbb{R}_+$ is the parameter to be estimated. This model is used for distance estimation from received signal strength [42], [43]. It can be shown that $g(\theta) = \frac{1}{\theta}$ is a valid solution of the ODE in (20). For this choice of weighting function there is no g-efficient estimator. However, we show in the following that we can find a g-unbiased estimator, for which $g(\theta) = \frac{1}{\theta}$ satisfies (8). This estimator is also uniformly optimal in terms of MSE among a set of estimators, which are given by a scale of the MVU estimator.

It is shown in [42] that the MVU estimator for this case exists and is given by $\hat{\theta}_{\text{MVU}} = e^{\frac{1}{N}\sum_{n=1}^{N} x_n - \frac{1}{2}\frac{\sigma_{\text{u}}^2}{N}}$. We will consider estimators of the form $c\hat{\theta}_{\text{MVU}}$ that have a linear bias, $(c-1)\theta$. It can be shown that the weighting function $g(\theta) = \frac{1}{\theta}$ satisfies (8) for these estimators. Thus, the corresponding g-CRB, $B_{g\text{-CRB}}(\theta) \triangleq \frac{\theta^2\sigma_{\text{u}}^2}{N+\sigma_{\text{u}}^2}$ is a valid bound for their MSE. It is shown in [42] that the ML estimator is equal to $\hat{\theta}_{\text{ML}} = e^{\frac{1}{2}\frac{\sigma_{\text{u}}^2}{N}}\hat{\theta}_{\text{MVU}}$ and that the minimum MSE estimator of the form $c\hat{\theta}_{\text{MVU}}$, is given by $\hat{\theta}_{c_{\text{opt}}} = e^{-\frac{\sigma_{\text{u}}^2}{N}}\hat{\theta}_{MVU}$. For $g(\theta) = \frac{1}{\theta}$, the g-unbiasedness is reduced to the following condition [2, p. 171]

$$\text{E}[\hat{\theta}^2 - \theta\hat{\theta}] = 0, \ \forall\theta \in \Omega_\theta. \tag{22}$$

By substituting the general estimator $c\hat{\theta}_{\text{MVU}}$ in (22), it can be verified that $\hat{\theta}_{c_{\text{opt}}}$ is the unique g-unbiased estimator among estimators of the form $c\hat{\theta}_{\text{MVU}}$. It is worth noting that unlike the CRB that grows linearly in σ_{u}^2 [42], the g-CRB with $g(\theta) = \frac{1}{\theta}$ is bounded by θ^2, similar to the MSE of $\hat{\theta}_{c_{\text{opt}}}$ [42].

In Fig. 2, the NMSEs of $\hat{\theta}_{\text{ML}}$, $\hat{\theta}_{\text{MVU}}$, and $\hat{\theta}_{c_{\text{opt}}}$, are evaluated using 10^6 Monte-Carlo simulations and presented versus the number of observations, N, for $\theta = e^2$, $\sigma_{\text{u}}^2 = 3.5$. These estimators are compared to the normalized CRB and the normalized g-CRB with $g(\theta) = \frac{1}{\theta}$, where the normalization is by the CRB. It can be seen that $\hat{\theta}_{c_{\text{opt}}}$ achieves lower NMSE than the normalized CRB and that the normalized g-CRB is valid for all of the considered estimators.

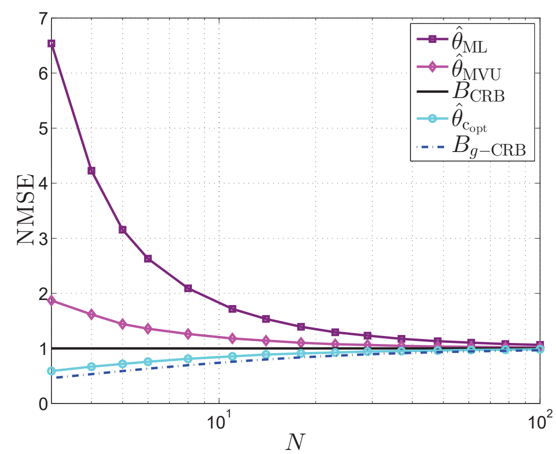

Fig. 2. Parameter estimation with nonlinear model and additive Gaussian noise: The NMSE's of $\hat{\theta}_{\text{ML}}$, $\hat{\theta}_{\text{MVU}}$, and $\hat{\theta}_{c_{\text{opt}}}$ compared to the normalized CRB and the normalized g-CRB, where $g(\theta) = \frac{1}{\theta}$ and the normalization is by the CRB.

VI. Conclusion

In this paper, we propose a new approach for estimation under the MSE risk. In this approach, by multiplying the MSE risk by a weighting function, we obtain a new unbiasedness definition, called g-unbiasedness, that stems from Lehmann's definition of unbiasedness. The g-unbiasedness is reduced to the conventional mean-unbiasedness for a constant g. By restricting the weighting function, we derive a new Cramér-Rao-type bound on the MSE of g-unbiased estimators. This Cramér-Rao-type bound is equal to or lower than the conventional CRB. By using the local g-unbiasedness condition we obtain a family of weighting functions to be considered, which are solutions to an ODE. It is shown that our method can lead to estimators that uniformly outperform the conventional MVU estimator. In particular, in some cases the PML estimator with an appropriate choice of penalizing function achieves the proposed bound. A topic for future research is developing a method to choose the ODE solution, among the family of solutions, which is optimal in terms of MSE. Another topic is to extend our general approach to vector parameters and to general Barankin-type lower bounds [44], [45].

Acknowledgment

This research was partially supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 1160/15).

REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[2] E. L. Lehmann and G. Casella, *Theory of Point Estimation (Springer Texts in Statistics), 2nd ed*, New York: Springer, 1998.

[3] E. W. Barankin, "Locally best unbiased estimates," *Ann. Math. Stat.*, vol. 20, pp. 477–501, 1946.

[4] C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.

[5] H. Cramér, "A contribution to the theory of statistical estimation," *Skand. Akt. Tidskr.*, vol. 29, pp. 85–94, 1946.

[6] K. Peters and S. Kay, "Unbiased estimation of the phase of a sinusoid," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2004, vol. 2, pp. 493–496.

[7] K. Todros, R. Winik, and J. Tabrikian, "On the limitations of Barankin type bounds for MLE threshold prediction," *Signal Processing*, vol. 108, pp. 622–627, 2015.

[8] D. R. Cox and D. V. Hinkley, *Theoretical statistics*, New York: Chapman & Hall, 1974.

[9] S. Kay and Y. C. Eldar, "Rethinking biased estimation [lecture notes]," *IEEE Signal Process. Mag.*, vol. 25, no. 3, pp. 133–136, May 2008.

[10] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, Sep. 2011.

[11] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[12] T. Kariya, "Equivariant estimation in a model with an ancillary statistic," *The Annals of Statistics*, vol. 17, no. 2, pp. 920–928, 1989.

[13] W. James and C. Stein, "Estimation with quadratic loss," in *Proc. of the fourth Berkeley symposium on mathematical statistics and probability*, 1961, vol. 1, pp. 361–379.

[14] B. J. N. Blight, "Some general results on reduced mean square error estimation," *The American Statistician*, vol. 25, no. 3, pp. 24–25, 1971.

[15] M. D. Perlman, "Reduced mean square error estimation for several parameters," *Sankhyā: The Indian Journal of Statistics, Series B*, pp. 89–92, 1972.

[16] P. Stoica and R. L. Moses, "On biased estimators and the unbiased Cramér-Rao lower bound," *Signal Processing*, vol. 21, no. 4, pp. 349–350, 1990.

[17] Y. C. Eldar, "Uniformly improving the Cramér-Rao bound and maximum-likelihood estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2943–2956, Aug. 2006.

[18] Y. C. Eldar, "MSE bounds with affine bias dominating the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3824–3836, Aug. 2008.

[19] A. O. Hero, J. A. Fessler, and M. Usman, "Exploring estimator bias-variance tradeoffs using the uniform CR bound," *IEEE Trans. Signal Process.*, vol. 44, no. 8, pp. 2026–2041, Aug. 1996.

[20] Y. C. Eldar, "Minimum variance in biased estimation: bounds and asymptotically optimal estimators," *IEEE Trans. on Signal Process.*, vol. 52, no. 7, pp. 1915–1930, July 2004.

[21] I. J. Good and R. A. Gaskins, "Nonparametric roughness penalties for probability densities," *Biometrika*, vol. 58, no. 2, pp. 255–277, 1971.

[22] J. A. Fessler and A. O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized em algorithms," *IEEE Trans. Image Process.*, vol. 4, no. 10, pp. 1417–1429, Oct. 1995.

[23] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 493–506, Mar. 1996.

[24] P. Jancovic and M. Kokuer, "Acoustic recognition of multiple bird species based on penalized maximum likelihood," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1585–1589, Oct. 2015.

[25] Y. C. Eldar, "Rethinking biased estimation: Improving maximum likelihood and the Cramér-Rao bound," *Foundations and Trends in Signal Processing*, vol. 1, no. 4, pp. 305–449, 2008.

[26] K. Todros and J. Tabrikian, "Uniformly best biased estimators in non-Bayesian parameter estimation," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7635–7647, Nov. 2011.

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

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

[29] T. Routtenberg and J. Tabrikian, "Performance bounds for constrained parameter estimation," in *Proc. of the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2012, pp. 513–516.

[30] S. Bar and J. Tabrikian, "Bayesian estimation in the presence of deterministic nuisance parameters–part I: Performance bounds," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6632–6646, Dec. 2015.

[31] T. Routtenberg and J. Tabrikian, "Cyclic Barankin-type bounds for non-Bayesian periodic parameter estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 13, pp. 3321–3336, July 2014.

[32] T. Routtenberg and L. Tong, "Estimation after parameter selection: Performance analysis and estimation methods," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5268–5281, Oct. 2016.

[33] T. Routtenberg, "Two-stage estimation after parameter selection," in *Proc. of the IEEE Statistical Signal Processing Workshop (SSP)*, June 2016.

[34] E. Nitzan, T. Routtenberg, and J. Tabrikian, "Mean-cyclic-error lower bounds via integral transform of likelihood-ratio function," in *Proc. of the 9th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2016.

[35] T. Routtenberg and J. Tabrikian, "Cyclic Cramér-Rao-type bounds for periodic parameter estimation," in *Proc. of the 19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 1797–1804.

[36] S. Bar and J. Tabrikian, "A risk-unbiased approach to a new Cramér-Rao bound," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 2921–2925.

[37] S. Bar and J. Tabrikian, "A risk-unbiased bound for information fusion with nuisance parameters," in *Proc. of the 19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 504–511.

[38] S. Bar and J. Tabrikian, "New observations on efficiency of variance estimation of white Gaussian signal with unknown mean," in *Proc. of the 9th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2016.

[39] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, "New results on deterministic Cramér-Rao bounds for real and complex parameters," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1032–1049, Mar. 2012.

[40] J. S. Abel, "A bound on mean-square-estimate error," *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1675–1680, Sep. 1993.

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

[42] S. D. Chitte, S. Dasgupta, and Z. Ding, "Distance estimation from received signal strength under log-normal shadowing: Bias and variance," *IEEE Signal Processing Letters*, vol. 16, no. 3, pp. 216–218, Mar. 2009.

[43] S. Wu, D. Xu, and S. Liu, "Weighted linear least square localization algorithms for received signal strength," *Wireless Personal Communications*, vol. 72, no. 1, pp. 747–757, 2013.

[44] E. Chaumette, J. Galy, A. Quinlan, and P. Larzabal, "A new Barankin bound approximation for the prediction of the threshold region performance of maximum likelihood estimators," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5319–5333, Nov. 2008.

[45] K. Todros and J. Tabrikian, "General classes of performance lower bounds for parameter estimation part I: Non-Bayesian bounds for unbiased estimators," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5045–5063, Oct. 2010.