THE CRAMÉR-RAO BOUND FOR ESTIMATION-AFTER-SELECTION

Tirza Routtenberg and Lang Tong

School of Electrical and Computer Engineering
Cornell University
Ithaca, NY 14850, New York

ABSTRACT

In many practical parameter estimation problems, a model selection is made prior to estimation. In this paper, we consider the problem of estimating an unknown parameter of a selected population, where the population is chosen from a population set by using a predetermined selection rule. Since the selection step may have an important impact on subsequent estimation, ignoring it could lead to biased-estimation and an invalid Cramér-Rao bound (CRB). In this work, the mean-square-selected-error (MSSE) criterion is used as a performance measure. The concept of Ψ-unbiasedness is introduced for a given selection rule, Ψ, by using the Lehmann-unbiasedness definition. We derive a non-Bayesian Cramér-Rao-type bound on the MSSE of any Ψ-unbiased estimator. The proposed Ψ-CRB is a function of the conditional Fisher information and is a valid bound on the MSSE. Finally, we examine the Ψ-CRB for different selection rules for mean estimation in a linear Gaussian model.

Index Terms— Non-Bayesian parameter estimation, Cramér-Rao bound (CRB), estimation-after-selection, linear Gaussian model, sample mean selection (SMS) rule

1. INTRODUCTION

The problem of estimation-after-selection arises naturally in a verity of practical problems in signal processing and communication. Estimation-after-selection refers to the problem of parameter estimation, in which the estimation is made only after a specific population has been selected from a set of possible populations according to a selection rule. In cognitive radio communications, for example, the parameters of a channel are estimated only after the channel has been identified in the white space. Other applications include multiple radar subset selection problems [1], medical experiments [2], estimation in wireless sensor networks after a sensor node selection [3], and state estimation after bad data detection in power systems [4], [5].

The statistical optimality properties of the preceding selection procedure and the probability of correct selection have been investigated in the literature [6]. Despite the importance of estimation-after-selection, there is no comprehensive *estimation* performance analysis of this problem with arbitrary distributions. For the problem of estimation after a preliminary selection stage, the mean-square-error (MSE) criterion is inappropriate and the conventional Cramér-Rao bound (CRB) is unsuited. In addition, the selection usually induces a bias on any estimator of the unknown parameter in the selected population. For example, no uniformly mean-unbiased estimator exists for estimation-after-selection problems with Gaussian populations and data-dependent selection rules [7].

This research was partially supported by the National Science Foundation under Grant no. CNS-1135844

1.1. Summary of results

In this work, it is assumed that a given selection rule, Ψ, selects one population from a set of populations, where Ψ may not be optimal in any sense. We are interested in the estimation performance of the unknown parameter associated with the selected population. We use the mean square-selected-error (MSSE) criterion and introduce the concept of Ψ-unbiasedness by using the non-Bayesian Lehmann-unbiasedness definition. Then, we develop the appropriate Cramér-Rao-type bound on the MSSE of any Ψ-unbiased estimator. The proposed Ψ-CRB is examined for the randomized (coin-flipping) selection rule and the sample mean selection (SMS) rule. It is shown that the CRB-type lower bound on estimation-after-detection in [8] and [9] can be obtained as a special case of the Ψ-CRB with a single population. Finally, the performance of the Ψ-CRB for different selection rules is examined for the problem of mean estimation in a linear Gaussian model.

1.2. Related works

The estimation-after-selection problem has received considerable attention in mathematical statistics literature for various observation densities and cost functions (see, e.g., [6], [7], [10], [11]). A related but different problem is the estimation of the maximum between unknown parameters [12], in which there is no selection stage. In [8] and [9], non-Bayesian lower bounds on the MSE under hypothesis H_1 are derived, for the problem of estimation-after-binary-detection. The tradeoff between detection and estimation is investigated in [13], in terms of worst-case detection and estimation error probabilities. In [14] and [15], joint detection and estimation of signals in an unknown region of interest (ROI) are investigated for the Bayesian case. In [16], the CRB is derived for the special case of estimation after a model order selection. The relation between model selection and the breakdown phenomena of the maximum-likelihood (ML) estimator are demonstrated in [17].

1.3. Notations

In the rest of this paper, we denote vectors by boldface lowercase letters and matrices by boldface uppercase letters. The operators $(\cdot)^T$ and $(\cdot)^{-1}$ denote the transpose and inverse operators, respectively. The (m,k)th element of the matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{m,k}$. The gradient of a vector, $\frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}}$, is a matrix, with (m,k)th element equal to $\frac{\partial g_m}{\partial \theta_k}$, where $\mathbf{g} = [g_1, \ldots, g_M]^T$. The notation $\mathrm{E}_{\boldsymbol{\theta}}[\cdot]$ represents the non-Bayesian expected value of its argument parametrized by a deterministic parameter $\boldsymbol{\theta}$. The notations $\phi(\cdot)$, $\Phi(\cdot)$, and $\mathbf{1}_A$, denote the standard normal probability density function (pdf), the standard normal cumulative distribution function (cdf), and the indicator function of an event A, respectively. The vector $\mathbf{e}_m \in \mathbb{R}^M$ denotes the mth column of the identity matrix of size $M, \forall m = 1, \ldots, M$.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

2. PROBLEM FORMULATION

Consider a set $M \in \mathbb{N}$ populations, which can represent, for example, M different communication channels. Suppose that $N_m \geq 1$ random observations are drawn from the mth population, where each population is associated pdf, $f_m(\mathbf{y}_m; \theta_m)$, $m = 1, \ldots, M$, in which $\mathbf{y}_m = [y_m[0], \ldots, y_m[N_m - 1]]^T$ and $\theta_m \in \mathbb{R}$ denote the observation vector and the unknown deterministic parameter related to the the mth population, respectively. The joint pdf of the M populations is denoted by $f(\mathbf{x}; \boldsymbol{\theta})$, where $\mathbf{x} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_M^T]^T \in \Omega_{\mathbf{x}}$ is the whole observation vector, $\Omega_{\mathbf{x}}$ is the observation space, and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]^T \in \mathbb{R}^M$. In the sequel, $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \ldots, \hat{\theta}_M]^T \in \mathbb{R}^M$ denotes an estimator of $\boldsymbol{\theta}$.

A *selection rule*, sometimes referred to as a test, is a deterministic function $\Psi : \Omega_{\mathbf{x}} \to \{1, \ldots, M\}$ that indicates the selected population based on the observations, $\mathbf{x}$. We assume that the deterministic subspaces $\mathcal{A}_m \triangleq \{\mathbf{x} : \mathbf{x} \in \Omega_{\mathbf{x}}, \Psi(\mathbf{x}) = m\}$, $m = 1, \ldots, M$ are not function of $\boldsymbol{\theta}$. The conditional joint pdfs of the M populations, conditioned on the event that the selection rule selects the mth population, are denoted by $f(\mathbf{x}|\Psi = m; \boldsymbol{\theta})$, $\forall m = 1, \ldots, M$, and $\Pr(\Psi = m; \boldsymbol{\theta})$ denotes the probability of selection of the mth population.

The estimation-after-selection problem consists of two stages: first, a population, m, is selected and then, the corresponding unknown parameter of the selected population, θ_m, is estimated. In this work, we assume that the population is chosen according to a given selection rule, Ψ, and we discuss the estimation performance for this predetermined selection rule. As a result, the estimation performance criterion should be only a function of the error in the selected population. The model is presented schematically in Fig. 1.

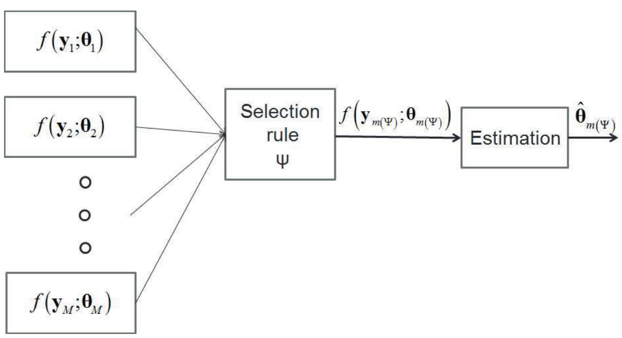

Fig. 1. The abstract model for estimation-after-selection.

In this work we are interested in parameter estimation of the unknown deterministic vector $\boldsymbol{\theta}$, where only estimation errors in the selected population are taken into consideration and the selection rule is predetermined. Therefore, for a given selection rule, Ψ, we use the following square-selected-error (SSE) cost (e.g. [2]):

$$C^{(\Psi)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \triangleq \sum_{m=1}^{M} (\hat{\theta}_m - \theta_m)^2 \mathbf{1}_{\{\Psi=m\}}. \quad (1)$$

The corresponding MSSE is given by $\mathrm{E}_{\boldsymbol{\theta}}[C^{(\Psi)}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})]$, where the expectation is parametrized by $\boldsymbol{\theta}$ and taken over the selection rule, Ψ, and the estimator, $\hat{\boldsymbol{\theta}}$. Similarly, the marginal MSSE of a specific population is given by $\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m - \theta_m)^2 | \Psi = m\right]$, for any $m = 1, \ldots, M$.

3. THE Ψ-CRB

In this section, a CRB-type lower bound for estimation-after-selection is derived. The proposed bound is a lower bound on the MSSE of any Lehmann-unbiased estimator, as described in the following.

3.1. Ψ-unbiasedness

The mean-unbiasedness constraint is commonly used in non-Bayesian parameter estimation [18], [19]. However, a mean-unbiased estimator is inappropriate for estimation-after-selection problems, since we are interested only in errors in the selected population (see, e.g., [7]). Lehmann [20] proposed a generalization of the unbiasedness concept, which is based on the considered cost function. In this section, the general Lehmann-unbiasedness is utilized to define the unbiasedness for estimation-after-selection problems.

Definition 1 *The estimator $\hat{\boldsymbol{\theta}}$ is said to be a uniformly unbiased estimator of $\boldsymbol{\theta}$ in the Lehmann sense [20] with respect to (w.r.t.) the scalar nonnegative cost function $C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ if*

$$\mathrm{E}_{\boldsymbol{\theta}}\left[C(\hat{\boldsymbol{\theta}}, \boldsymbol{\eta})\right] \geq \mathrm{E}_{\boldsymbol{\theta}}\left[C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})\right], \quad \forall \boldsymbol{\eta}, \boldsymbol{\theta} \in \Theta, \quad (2)$$

where Θ is the parameter space.

The Lehmann-unbiasedness definition implies that an estimator is unbiased if, on average, it is "closer" to the true parameter, $\boldsymbol{\theta}$, than to any other value in the parameter space, $\boldsymbol{\eta} \in \Theta$. The measure of "closeness" between the estimator and the parameter is the mean of the cost function, $C(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$. For example, it is shown in [20] that under the square-error cost function the Lehmann-unbiasedness in (2) is reduced to the conventional mean-unbiasedness, $\mathrm{E}_{\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}, \forall \boldsymbol{\theta} \in \Omega_{\boldsymbol{\theta}}$. Additional examples for Lehmann-unbiasedness with different cost functions can be found, for example, in [20], [21], and [22]. The following proposition describes the Lehmann-unbiasedness for the estimation-after-selection problem, named Ψ-unbiasedness.

Proposition 1 *An estimator $\hat{\boldsymbol{\theta}} : \Omega_{\mathbf{x}} \to \mathbf{R}^M$ is an unbiased estimator of $\boldsymbol{\theta} \in \mathbf{R}^M$ in the Lehmann sense w.r.t. the SSE cost and the selection-rule Ψ* iff

$$\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m - \theta_m)\mathbf{1}_{\{\Psi=m\}}\right] = 0, \ \forall m = 1, \ldots, M, \ \forall \boldsymbol{\theta} \in \mathbb{R}^M, \quad (3)$$

or equivalently, iff

$$\mathrm{E}_{\boldsymbol{\theta}}\left[\hat{\theta}_m - \theta_m | \Psi = m\right] = 0, \quad \forall \boldsymbol{\theta} \in \mathbb{R}^M, \quad (4)$$

$\forall m = 1, \ldots, M$, *such that* $\Pr(\Psi = m; \boldsymbol{\theta}) \neq 0$.

The proof of Proposition 1 can be performed similar to to the one in [20], p. 11, and is omitted due to space limitations. It can be seen that the Lehmann-unbiasedness definition in (3) and (4) is a function of the given selection-rule. Therefore, in the following, an estimator $\hat{\boldsymbol{\theta}}$ is said to be a uniformly Ψ-unbiased estimate of $\boldsymbol{\theta}$ for the given selection-rule Ψ if (3), or equivalently (4)), is satisfied.

3.2. The Ψ-CRB

Calculation of the minimum MSSE among all Ψ-unbiased estimators is usually not tractable and a uniform Ψ-unbiased minimum MSSE estimator may not exist for the commonly-used SMS rule.

Thus, lower bounds on the performance of any Ψ-unbiased estimator are useful for performance analysis and system design. In the following, we develop the Ψ-CRB, which is a lower bound on the MSSE of any Ψ-unbiased estimator.

Let us define the following marginal conditional Fisher information matrices (FIMs):

$$\mathbf{J}_m(\boldsymbol{\theta},\Psi) \triangleq \mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial^T \log f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \times \frac{\partial \log f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|\Psi=m\Bigg], \quad (5)$$

$\forall m = 1,\dots,M$. In addition, we define the following conditions that are a modified version of the well-known CRB regularity conditions (e.g. [23], pp. 440-441).

C.1. The conditional likelihood gradient vector, $\frac{\partial}{\partial\boldsymbol{\theta}}\log f(\mathbf{x}|\Psi = m;\boldsymbol{\theta})$, exists and is finite $\forall\boldsymbol{\theta}\in\mathbb{R}^M$, $\mathbf{x}\in\mathcal{A}_m$, and $\forall m = 1,\dots,M$. Thus, the marginal conditional FIMs, $\mathbf{J}_m(\boldsymbol{\theta},\Psi)$, are well defined $\forall m = 1,\dots,M$. We assume that $\mathbf{J}_m(\boldsymbol{\theta},\Psi)$ is a nonsingular and nonzero matrix $\forall\boldsymbol{\theta}\in\mathbb{R}^M$ and $\forall m = 1,\dots,M$.

C.2. The operations of integration w.r.t. $\mathbf{x}$ and differentiation w.r.t. $\boldsymbol{\theta}$ can be interchanged as follows:

$$\int_{\Omega_{\mathbf{x}}}\frac{\partial}{\partial\boldsymbol{\theta}}\left(g(\mathbf{x},\boldsymbol{\theta})f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})\right)\mathrm{d}\mathbf{x} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\mathrm{E}_{\boldsymbol{\theta}}\left[g(\mathbf{x},\boldsymbol{\theta})|\Psi=m\right],$$

$\forall\boldsymbol{\theta}\in\mathbb{R}^M$ and for any differentiable and measurable function $g(\mathbf{x},\boldsymbol{\theta})$.

Theorem 1 *(Ψ-CRB) Let the regularity conditions C.1-C.2 be satisfied and $\hat{\boldsymbol{\theta}}$ be a Ψ-unbiased estimator of $\boldsymbol{\theta}\in\mathbf{R}^M$ with a finite second moment for a given selection rule, Ψ. Then, the MSSE is bounded by the following Cramér-Rao-type lower bound:*

$$\mathrm{E}_{\boldsymbol{\theta}}\left[C^{(\Psi)}(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})\right] \geq B^{(\Psi)}(\boldsymbol{\theta}), \quad (6)$$

where

$$B^{(\Psi)}(\boldsymbol{\theta}) \triangleq \sum_{m=1}^{M}\Pr(\Psi=m;\boldsymbol{\theta})\left[\mathbf{J}_m^{-1}(\boldsymbol{\theta},\Psi)\right]_{m,m} \quad (7)$$

and $\mathbf{J}_m(\boldsymbol{\theta},\Psi)$ is the conditional FIM defined in (5). Furthermore, the marginal Ψ-CRB on the MSSE of specific population is given by

$$\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m-\theta_m)^2|\Psi=m\right] \geq \left[\mathbf{J}_m^{-1}(\boldsymbol{\theta},\Psi)\right]_{m,m}, \quad (8)$$

$\forall m = 1,\dots,M$.

Proof: According to Cauchy-Schwarz inequality:

$$\mathrm{E}_{\boldsymbol{\theta}}\left[g^2(\mathbf{x},\boldsymbol{\theta})\middle|\Psi=m\right] \geq \frac{\mathrm{E}^2_{\boldsymbol{\theta}}\left[g(\mathbf{x},\boldsymbol{\theta})h(\mathbf{x},\boldsymbol{\theta})|\Psi=m\right]}{\left[h^2(\mathbf{x},\boldsymbol{\theta})|\Psi=m\right]}, \quad (9)$$

for any measurable functions $g(\mathbf{x},\boldsymbol{\theta})$ and $h(\mathbf{x},\boldsymbol{\theta})$ with finite moment. By substituting $g(\mathbf{x},\boldsymbol{\theta}) = \hat{\theta}_m - \theta_m$ and $h(\mathbf{x},\boldsymbol{\theta}) = \frac{\partial\log f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\mathbf{J}_m^{-1}(\boldsymbol{\theta},\Psi)\mathbf{e}_m$ in (9) and under condition C.1, one obtains

$$\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m-\theta_m)^2\middle|\Psi=m\right] \geq \frac{\left(\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m-\theta_m)\frac{\partial\log f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|\Psi=m\right]\mathbf{J}_m^{-1}(\boldsymbol{\theta},\Psi)\mathbf{e}_m\right)^2}{\mathbf{e}_m^T\mathbf{J}_m(\boldsymbol{\theta},\Psi)\mathbf{e}_m}, \quad (10)$$

for any estimator with $\left[(\hat{\theta}_m-\theta_m)^2\middle|\Psi=m\right] < \infty$. By using integration by parts and assuming regularity condition C.2, it can be verified that

$$\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m-\theta_m)\frac{\partial\log f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\middle|\Psi=m\right] = \frac{\partial}{\partial\boldsymbol{\theta}}\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m-\theta_m)\middle|\Psi=m\right]+\mathbf{e}_m = \mathbf{e}_m, \quad (11)$$

$\forall m = 1,\dots,M$, where the last equality is obtained by using the Ψ-unbiasedness conditions from (4). By substituting (11) in (10), we obtain

$$\mathrm{E}_{\boldsymbol{\theta}}\left[(\hat{\theta}_m-\theta_m)^2\middle|\Psi=m\right] \geq \mathbf{e}_m^T\mathbf{J}_m^{-1}(\boldsymbol{\theta},\Psi)\mathbf{e}_m, \quad (12)$$

$\forall m = 1,\dots,M$, which is the marginal MSSE bound in (8). Then, by multiplying (12) by $\Pr(\Psi=m;\boldsymbol{\theta})$ and taking the sum of over $m = 1,\dots,M$, we obtain the Ψ-CRB in (6).

It should be noted that for the special case of independent populations, i.e. $f(\mathbf{x};\boldsymbol{\theta}) = \prod_{m=1}^M f_m(\mathbf{y}_m;\theta_m)$, the classical FIM, $\mathbf{J}(\boldsymbol{\theta})$, is a diagonal matrix and, in terms of the CRB, the unknown parameters of the different populations are decoupled from each other. However, the conditional FIMs are not necessary diagonal since the selection step may create a dependency and coupling between the parameters over the different populations.

The following lemma presents two alternative formulations of the conditional FIMs.

Lemma 1 *Assuming that regularity conditions C.1-C.2 are satisfied and*

1. *The second derivative w.r.t. $\boldsymbol{\theta}$ of $f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})$ exists and is bounded and continuous $\forall\mathbf{x}\in\mathcal{A}_m$.*
2. *The integral $\int_{\mathcal{A}_m} f(\mathbf{x}|\Psi=m;\boldsymbol{\theta})\,\mathrm{d}\mathbf{x}$ is twice differentiable under the integral sign $\forall m = 1,\dots,M$, $\boldsymbol{\theta}\in\mathbb{R}^M$.*

Then, the marginal conditional FIMs in (5) satisfy

$$\mathbf{J}_m(\boldsymbol{\theta},\Psi) = \mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial^T\log f(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\log f(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\middle|\Psi=m\right] - \frac{\partial^T\log\Pr(\Psi=m;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\log\Pr(\Psi=m;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \quad (13)$$

and

$$\mathbf{J}_m(\boldsymbol{\theta},\Psi) = -\mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial^T\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}}\log f(\mathbf{x};\boldsymbol{\theta})\middle|\Psi=m\right] + \frac{\partial^T\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}}\log\Pr(\Psi=m;\boldsymbol{\theta}), \quad (14)$$

$\forall m = 1,\dots,M$, $\forall\boldsymbol{\theta}\in\mathbb{R}^M$, and for any selection rule Ψ.

Proof: The proof appears in [24] and is similar to the results in [8]. It is omitted due to space limitations.

3.3. Special cases

1) Randomized selection rule (coin-flipping): The randomized selection rule, Ψ_{rand}, satisfies

$$\Pr(\Psi_{\text{rand}}=m;\boldsymbol{\theta}) = p_m, \quad \forall m = 1,\dots,M,$$

where the probabilities, $p_m\in[0,1]$, $\forall m = 1,\dots,M$, are fixed, independent of $\mathbf{x}$ and $\boldsymbol{\theta}$. By substituting the randomized selection rule in (3), the Ψ-unbiasedness is reduced to

$$\mathrm{E}_{\boldsymbol{\theta}}\left[\hat{\theta}_m-\theta_m\right] = 0, \quad \forall\boldsymbol{\theta}\in\mathbb{R}^M, \quad (15)$$

$\forall m = 1, \dots, M$ with $p_m \neq 0$. Similarly, the Ψ-CRB in (7) in this case is reduced to

$$B^{(\Psi_{\text{rand}})}(\boldsymbol{\theta}) = \sum_{m=1}^{M} p_m \left[\mathbf{B}(\boldsymbol{\theta})\right]_{m,m}, \tag{16}$$

where

$$\mathbf{B}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial^T \log f(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log f(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]$$

is the conventional CRB. Therefore, for the randomized selection rule, the Ψ-unbiasedness in (15) is the classical mean-unbiasedness and the proposed Ψ-CRB in (16) is a linear combination of the diagonal elements of the conventional CRB.

2) SMS rule: For selecting the population with the largest (or smallest) mean, the SMS rule, Ψ_{SMS}, selects the population with the largest (or smallest) sample mean. The conditions under which this procedure has statistical optimality properties from the decision perspective have been investigated in many works (see, e.g., [6]). The bound on the MSSE for Ψ_{SMS} is important, primarily, because of the widespread use of the SMS rule in realistic signal processing problems.

3) Relation to estimation-after-detection: For the special case of single population, i.e. $M = 1$ and $\boldsymbol{\theta} = \theta_1 = \theta$, our model is reduced to the estimation-after-data-censoring model. In this case, Ψ is not a selection rule but a function that restricts the set of observations available for parameter estimation. That is, we use the observations only if $\Psi = 1$. For example, the data censoring can be performed based on a binary detection step, which decides if a signal is present or not [8], [9]. By using (5) and substituting $M = 1$ in (8), we obtain the following marginal Ψ-CRB on the MSSE for this case:

$$\mathrm{E}_\theta\left[(\hat{\theta}-\theta)^2|\Psi=1\right] \geq$$
$$\mathrm{E}_\theta^{-1}\left[\frac{\partial^T \log f(\mathbf{x}|\Psi=1;\theta)}{\partial \theta}\frac{\partial \log f(\mathbf{x}|\Psi=1;\theta)}{\partial \theta}\middle|\Psi=1\right]. \tag{17}$$

The special case of the Ψ-CRB in (17) is identical to the conditional CRB for estimation after a binary detection [8], [9]. Therefore, the proposed Ψ-CRB can be interpreted as a generalization of the conditional CRB for multiple populations.

4. LINEAR GAUSSIAN MODEL

Consider the following observation model

$$\begin{cases} y_1[n] = \theta_1 + w_1[n], & n = 0,\dots,N-1 \\ y_2[n] = \theta_2 + w_2[n], & n = 0,\dots,N-1 \end{cases}, \tag{18}$$

where $\mathbf{w}[n] = [w_1[n], w_2[n]]^T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, that is, the noise vectors are independent and identically normally distributed with zero-mean and known covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2}^2 \\ \sigma_{1,2}^2 & \sigma_2^2 \end{bmatrix}$. Suppose we are interested in the selection of the population with the largest mean, θ_1, θ_2. It can be verified that for this case

$$\frac{\partial^T}{\partial \boldsymbol{\theta}} \frac{\partial \log f(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -N\boldsymbol{\Sigma}^{-1}. \tag{19}$$

Thus, the conventional unconditional FIM is given by $\mathbf{J}(\boldsymbol{\theta}) = N\boldsymbol{\Sigma}^{-1}$. Thus, for the general case by substituting (19) in (14), one obtains

$$\mathbf{J}_m(\boldsymbol{\theta}, \Psi) = N\boldsymbol{\Sigma}^{-1} + \frac{\partial^T \partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \log \Pr(\Psi = m; \boldsymbol{\theta}), \tag{20}$$

$m = 1, 2$ for any given arbitrary selection rule Ψ. By substituting (20) in and (6), we obtain the proposed bound for any selection rule, Ψ. In particular, by substituting (20) in (16), we obtain the Ψ-CRB for randomized selection rule:

$$B^{(\Psi_{\text{rand}})}(\boldsymbol{\theta}) = \frac{1}{N}\left(p_1\sigma_1^2 + p_2\sigma_2^2\right).$$

The SMS rule for this case selects the population which provides the larger sample mean, i.e. $\Psi_{\text{SMS}} = \arg\max_{m=1,2}\left\{\hat{\theta}_m\right\}$, where

$$\hat{\theta}_m \triangleq \frac{1}{N}\sum_{n=0}^{N-1} y_m[n], \; m = 1, 2$$

is the ML estimator. Since the ML estimators are jointly Gaussian random variables with means θ_1, θ_2 and covariance matrix $\frac{1}{N}\boldsymbol{\Sigma}$, we obtain

$$\Pr(\Psi_{\text{SMS}} = m; \boldsymbol{\theta}) = \Pr(\hat{\theta}_m - \hat{\theta}_l > 0; \boldsymbol{\theta}) = \Phi\left(\frac{\theta_m - \theta_l}{\sigma}\right), \tag{21}$$

$\forall m, l = 1, 2, m \neq l$ where $\sigma^2 \triangleq \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{1,2}^2}{N}$. Then, the probability in (21) is used in order to calculate $\frac{\partial^T \partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \log \Pr(\Psi = m; \boldsymbol{\theta})$ and obtain the analytical term of the Ψ-CRB with the SMS rule.

In the simulations we used the following selection rules: SMS, randomized selection rule with $p_1 = p_2 = \frac{1}{2}$, and randomized selection rule with $p_1 = 1$ and $p_2 = 0$, denoted by Ψ_{SMS}, $\Psi_{\text{rand},1/2}$, and $\Psi_{\text{rand},1}$, respectively. The performance of the ML estimator, together with these selection rules, was evaluated using $5,000$ Monte-Carlo simulations and compared to the corresponding Ψ-CRBs. The parameters were set to $\theta_1 = 11$, $\theta_2 = 10.75$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$. It can be observed that for low signal-to-noise ratios (SNR), $\Psi_{\text{SMS}} \to \Psi_{\text{rand},1/2}$ and the corresponding MSSE/ proposed CRB approaches the randomized Ψ-CRB in (16) with $p_1 = p_2 = \frac{1}{2}$. Similarly, for high SNR, $\Psi_{\text{SMS}} \to \Psi_{\text{rand},1}$ and the SMS Ψ-CRB converges to the randomized Ψ-CRB in (16) with $p_1 = 1$, since the first population has the largest mean. In addition, it can be seen that the Ψ-CRBs can be used a benchmarks for the MSSE performance of the ML estimator with the different selection rule.

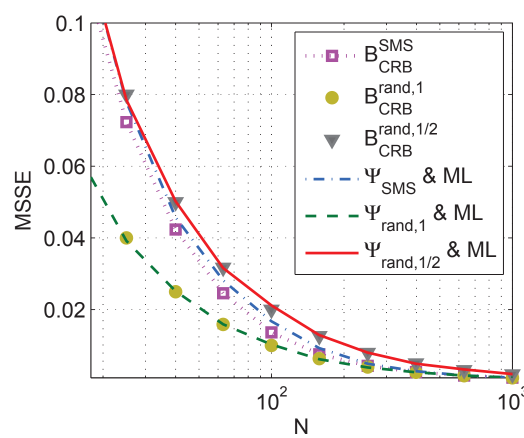

Fig. 2. The Ψ-CRBs and the performance of the ML estimator for estimation-after-selection in linear Gaussian model with different selection rules.

5. REFERENCES

[1] H. Godrich, A. Petropulu, and H. Poor, "Sensor selection in distributed multiple-radar architectures for localization: A knapsack problem formulation," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 247–260, Jan. 2012.

[2] J. Bowden and E. Glimm, "Unbiased estimation of selected treatment means in two-stage trials," *Biometrical Journal*, vol. 50, no. 4, pp. 515–527, 2008.

[3] T. Zhao and A. Nehorai, "Distributed sequential Bayesian estimation of a diffusive source in wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1511–1524, Apr. 2007.

[4] E. Handschin, F. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power Apparatus and Systems*, vol. 94, no. 2, pp. 329–337, Mar. 1975.

[5] J. Kim and L. Tong, "On topology attack of a smart grid: Undetectable attacks and countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1294–1305, July 2013.

[6] M. L. Eaton, "Some optimum properties of ranking procedures," *The Annals of Mathematical Statistics*, vol. 38, no. 1, pp. 124–137, Feb. 1967.

[7] A. Cohen and H. B. Sackrowitz, "Estimating the mean of the selected population," *In: Gupta, S. S., Berger, J. O., eds. Statistical Decision Theory and Related Topics-III.*, vol. 1, pp. 247–270, 1982.

[8] E. Chaumette, P. Larzabal, and P. Forster, "On the influence of a detection step on lower bounds for deterministic parameter estimation," *IEEE Trans. Signal Processing*, vol. 53, no. 11, pp. 4080–4090, Nov. 2005.

[9] E. Chaumette and P. Larzabal, "Cramér-Rao bound conditioned by the energy detector," *IEEE Signal Processing Letters*, vol. 14, no. 7, pp. 477–480, July 2007.

[10] K. Sarkadi, "Estimation after selection," *Studia Scientarium Mathematicarum Hungarica*, pp. 341–350, 1967.

[11] J. D. Gibbons, I. Olkin, and M. Sobel, *Selecting and Ordering Populations. A New Statistical Methodology*. New York: John Wiley and Sons., 1977.

[12] H. van Hasselt, "Estimating the maximum expected value: an analysis of (nested) cross validation and the maximum sample average," *ArXiv, preprint arXiv:1302.7175*, Feb. 2013.

[13] B. Baygun and A. Hero, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 688–703, May 1995.

[14] E. Bashan, R. Raich, and A. Hero, "Optimal two-stage search for sparse targets using convex criteria," *IEEE Trans. Signal Processing*, vol. 56, no. 11, pp. 5389–5402, Nov. 2008.

[15] E. Bashan, G. Newstadt, and A. Hero, "Two-stage multiscale search for sparse targets," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2331–2341, May 2011.

[16] S. Sando, A. Mitra, and P. Stoica, "On the Cramér-Rao bound for model-based spectral analysis," *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 68–71, Feb. 2002.

[17] N. Arkind and B. Nadler, "Parametric joint detection-estimation of the number of sources in array processing," in *SAM 2010*, Oct. 2010, pp. 269–272.

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

[19] S. Kay and Y. C. Eldar, "Rethinking biased estimation," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, May 2008.

[20] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer Texts in Statistics, 2005.

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

[22] ——, "Performance bounds for constrained parameter estimation," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop, SAM 2012*, June 2012, pp. 513–516.

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

[24] T. Routtenberg and L. Tong, "Performance analysis and sequential sampling polices for estimation-after-selection," in preparation.