Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Constrained Cramér–Rao Lower Bound in Errors-In Variables (EIV) models: Revisited

A. Al-Sharadqah [a,*], K.C. Ho [b]

[a] Department of Mathematics, and Interdisciplinary Research Institute for the Sciences (IRIS), California State University Northridge, Northridge, CA 91330-8313, USA
[b] Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211, USA

ARTICLE INFO

Article history:
Received 8 February 2017
Received in revised form 12 September 2017
Accepted 17 October 2017
Available online xxxx

Keywords:
Constrained Cramér–Rao Lower Bound
Errors-In-Variables models
Functional model

ABSTRACT

The Constrained Cramér–Rao Lower Bound (CCRB) works only for an unbiased estimator. The CCRB of Stoica and Ng (1998) is revisited and generalized. The bound is applied to two applications in the nonlinear EIV models.

In a curve fitting problem, the goal is to estimate the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ that defines a family of curves represented by the implicit equation

$$P(x, y; \boldsymbol{\theta}) = 0. \tag{1}$$

For example in the linear fitting problem, the two parameters to be estimated are (α, β). The two parameters describe the straight-line $y = \alpha + \beta x$. Similarly, the parameters can be the coordinates of the circle center $\mathbf{c} = (a, b)^T$ and the radius R for a single circle fitting. In the concentric circles fitting problem, the model often contains the common center $\mathbf{c} = (a, b)$ and the radii R_1 and R_2 where $R_1 < R_2$.

The estimation requires a number of measurement points and they are represented by the observation vector $\boldsymbol{\mathcal{X}} = (\mathbf{s}_1^T, \ldots, \mathbf{s}_n^T)^T = \tilde{\boldsymbol{\mathcal{X}}} + \mathbf{e}$, where $\mathbf{s}_i^T = (x_i, y_i)$ is the ith observation, $\tilde{\boldsymbol{\mathcal{X}}} = (\tilde{\mathbf{s}}_1^T, \ldots, \tilde{\mathbf{s}}_n^T)$ is the true value of $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\mathcal{E}} = (\mathbf{e}_1^T, \ldots, \mathbf{e}_n^T)^T$ is the noise vector and $\mathbf{e}_i$ is the measurement noise for point $i = 1, 2, \ldots, n$.

We can exploit Eq. (1) in many ways to obtain $\boldsymbol{\theta}$. For instance, the least-squares approach uses the noisy measurements in Eq. (1) and it minimizes the sum of the residual squared errors. The Errors-In-Variables (EIV) approach, which we are going to focus on, takes a completely different formulation. It solves the fitting problem by treating the true point vector $\tilde{\boldsymbol{\mathcal{X}}}$ as an unknown but fixed (non-random) parameter. Essentially the true points of the measurements are considered as nuisance variables. Each true point is expected to satisfy Eq. (1) perfectly. The set of implicit equations, one for each true point, acts as constraints for the estimation. Such model is known in the EIV literature as the *functional model*, and it is adopted in the application oriented community especially in computer vision. Employing the EIV approach has an additional $2n$ ('incidental') parameters and there are a total of $2n + p$ unknowns to be found.

As a standard measure in statistical performance, the efficiency of any unbiased estimator can be determined by the Cramèr–Rao bound (CRB) obtained by taking the inverse of the Fisher information matrix (FIM). The CRB is widely used in

* Corresponding author.
E-mail addresses: ali.alsharadqah@csun.edu (A. Al-Sharadqah), hod@missouri.edu (K.C. Ho).

https://doi.org/10.1016/j.spl.2017.10.009
0167-7152/

statistics for any unbiased estimator without any restriction on the parameters space that is continuous. However, when a parameter space is imposed with some constraints, then the FIM might not be easy to play with.

Evaluating the CRB for the geometric fitting problems, such as circle or ellipse fitting, has a long history. For the circle fitting problem, for instance, the CRB was derived in 1995 by Chan and Thomas (1995). Kanatani (1998) derived a general CRB for arbitrary curves for any unbiased estimator, then independently Zelniker and Clarkson presented another proof in 2006, see Zelniker and Clarkson (2006) for more details. That is, let $\hat{\boldsymbol{\theta}}$ be an unbiased estimator of $\boldsymbol{\theta}$ satisfying Eq. (1) and let $\mathbf{V}$ be its covariance matrix. If $\mathbf{e}_i$'s are identically and independently distributed and $\mathbf{e}_i \sim \mathbb{N}(\mathbf{0}, \sigma^2\mathbf{I}_2)$, then there is a symmetric positive semi-definite matrix $\mathbf{V}_{\min}$ such that $\mathbf{V} \geq \sigma^2\mathbf{V}_{\min}$. Here the notation $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. From now on, we will also denote the zero and the identity matrix of size k by $\mathbf{0}_k$ and $\mathbf{I}_k$, respectively. In addition, we denote $\mathbf{A}_{k\times m}$ to be a matrix of size $k \times m$, while $\mathbf{A}_k$ to be a square matrix of size k.

One obvious disadvantage surrounds the CRB is that it can be applied only to unbiased estimators! However, all estimators for the geometric fitting problems are, essentially, biased. The magnitude of bias is particularly significant when the noise level of the observation is large.

In the early 2000's, Chernov and Lesort (2004) realized that Kanatani's formula does not work for any (practical) estimator in curve fitting problems in EIV models because all existing estimators are biased. To resolve this problem, they treated the CRB derivation problem as one with a minimal requirement called geometric consistency. That is, an estimator $\hat{\boldsymbol{\theta}}$ is called a geometrically consistent estimator of the true parameter vector $\tilde{\boldsymbol{\theta}}$, if, $\mathrm{plim}_{\boldsymbol{\mathcal{X}}\to\tilde{\boldsymbol{\mathcal{X}}}}\hat{\boldsymbol{\theta}}(\boldsymbol{\mathcal{X}}) = \tilde{\boldsymbol{\theta}}$.

Precisely geometrically consistent estimator means that whenever $\sigma = 0$, i.e. when the true points are observed without noise, then the estimator returns the true parameter vector, i.e. finds the true curve. Geometrically, it means that if there is a model curve that interpolates the data points, then the algorithm finds it. With some degree of informality, one can assert that whenever $\mathrm{plim}_{\boldsymbol{\mathcal{X}}\to\tilde{\boldsymbol{\mathcal{X}}}}\hat{\boldsymbol{\theta}}(\boldsymbol{\mathcal{X}}) = \tilde{\boldsymbol{\theta}}$ holds, the estimate $\hat{\boldsymbol{\theta}}(\boldsymbol{\mathcal{X}})$ is *consistent* in the limit $\sigma \to 0$. This is regarded as a minimal requirement for any sensible fitting algorithm. For example, if the observed points lie on one circle, then every circle fitting algorithm finds that circle uniquely. Kanatani (2005) remarks that algorithms which fail to follow this property "are not worth considering". Therefore, this property is essential for any meaningful estimator in geometric estimation problems.

Chernov and Lesort (2004) assumed that estimators must have this property in order to derive the CRB. They also assumed that the observation noises $\mathbf{e}_i$ are isotropic. Accordingly, they employed the first-order analysis for any geometrically consistent estimator and derived the minimal possible lower bound, to the first leading term. That is, there is a symmetric positive semi-definite matrix $\mathbf{V}_{\min}$, such that, *the leading term of the variance matrix*, say $\mathbf{V}$, of any geometrically consistent estimator $\hat{\boldsymbol{\theta}}$ satisfying the condition $\mathrm{plim}_{\boldsymbol{\mathcal{X}}\to\tilde{\boldsymbol{\mathcal{X}}}}\hat{\boldsymbol{\theta}}(\boldsymbol{\mathcal{X}}) = \tilde{\boldsymbol{\theta}}$ has its natural bound $\sigma^2\mathbf{V}_{\min}$, i.e., $\mathbf{V} \geq \sigma^2\mathbf{V}_{\min}$. This bound is the leading term of the variance and it coincides with Kanatani's formula. The CRB derived by Kanatani can be applied only to unbiased estimators. Therefore, they called it the *Kanatani–Cramér–Rao bound* (KCR) after Kanatani. Since then, the KCR bound has been used as a measure of the efficiency for any meaningful estimator for the curve fitting problem.

The isotropic assumption about the noise is intuitively realistic for computer vision applications. For edge detection, pattern recognition, and computer vision, detecting (observing) any point $\mathbf{s}_i$ does not give any information about other points and hence the errors are independent. It is also reasonable to assume that the errors $\mathbf{e}_i$'s have the same covariance matrix for all points, because we use the same algorithm for edge detection. These statistical assumptions, however, are easily violated in several real life situations. In fact, reasonable assumptions about the noise depend on the particular application itself. Several studies in the literature have recently used other specific assumptions, such as the Markov Chain assumption between observations for correlated data that appears in archaeology (Chernov and Sapirstein, 2008) and others. This motivates us enough to generalize their work. We will generalize the CRB in order to be applicable to a wide range of data including correlated data.

Compared to the previous work, we took a different approach to evaluate the bound under the EIV model. In particular, the parameter space is expanded to $2n + p$ variables and Eq. (1) is utilized as a constraint between each true data point and the geometric parameters. Such interpretation allows us to start from the work done by Stoica and Ng (1998, Theorem 1) for Constrained CRB (CCRB) that can be applied to only unbiased estimator. Then we derive the constrained CR bound for a biased estimator. Therefore, our results do not require the condition $\mathrm{plim}_{\boldsymbol{\mathcal{X}}\to\tilde{\boldsymbol{\mathcal{X}}}}\hat{\boldsymbol{\theta}}(\boldsymbol{\mathcal{X}}) = \tilde{\boldsymbol{\theta}}$.

To state Stoica and Ng's results (Stoica and Ng, 1998), let us first denote the probability density function of the random vector $\boldsymbol{\mathcal{X}}$ parameterized by $\boldsymbol{\Theta}$ by $f(\boldsymbol{\mathcal{X}}; \boldsymbol{\Theta})$, and the likelihood function by L. The score function is $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\Theta}) = \frac{\partial \log L(\boldsymbol{\Theta};\boldsymbol{\mathcal{X}})}{\partial \boldsymbol{\Theta}}$, and as such, the FIM becomes $\mathbf{J}_m = \mathbb{E}(\boldsymbol{\Delta}\boldsymbol{\Delta}^T)$. Now, if we define $\mathbf{P} = \left(\tilde{P}_1, \ldots, \tilde{P}_n\right)^T$, where $\tilde{P}_i = P(\tilde{\mathbf{s}}_i, \tilde{\boldsymbol{\theta}})$ then the constraints $\mathbf{P} = \mathbf{0}_{n\times 1}$ imposes n-continuously differentiable constraints given by $\mathbf{F}_{n\times m} = \mathbf{0}_{n\times m}$, where the ith row of $\mathbf{F}$ is $\frac{\partial P(\tilde{\mathbf{s}}_i, \boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}^T}\Big|_{\boldsymbol{\Theta}=\tilde{\boldsymbol{\Theta}}}$, i.e.,

$$\left(\frac{\partial \tilde{P}_i}{\partial \tilde{\mathbf{s}}_1^T}, \ldots, \frac{\partial \tilde{P}_i}{\partial \tilde{\mathbf{s}}_n^T}, \frac{\partial \tilde{P}_i}{\partial \tilde{\theta}_1}, \ldots, \frac{\partial \tilde{P}_i}{\partial \tilde{\theta}_p}\right). \tag{2}$$

The $n \times m$ gradient matrix $\mathbf{F}$ is assumed to have full row rank, i.e. $\mathrm{rank}(\mathbf{F}) = n$ (note here that $m = 2n + p$). This means that there exists a matrix $\mathbf{U} = \mathbf{U}(\tilde{\boldsymbol{\Theta}}) \in \mathbb{R}^{m\times(m-n)}$ such that

$$\mathbf{F}\mathbf{U} = \mathbf{0}_{n\times(m-n)}, \qquad \mathbf{U}^T\mathbf{U} = \mathbf{I}_{m-n}. \tag{3}$$

Please cite this article in press as: Al-Sharadqah A., Ho K.C., Constrained Cramér–Rao Lower Bound in Errors-In Variables (EIV) models: Revisited. Statistics and Probability Letters (2017), https://doi.org/10.1016/j.spl.2017.10.009.