Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 67–70
http://www.ssca.org.in/journal

Cramer-Rao Posterior Bounds in the Spirit of van Trees

Malay Ghosh
University of Florida

Received: 07 January 2024; Revised: 06 May 2024; Accepted: 06 May 2024

Abstract

The paper obtains posterior Cramer-Rao bounds for arbitrary parameter vector in the spirit of van Trees. The relationship with the classical Cramer-Rao bound is also discussed.

Key words: Cramer-Rao; Parametric functions; Posterior variances.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

The present short article is a humble tribute to the late Professor C. R. Rao, one of the legendary heroes in the world of statistics. In his professional life, which spanned for nearly eight decades, Professor Rao made a number of pioneering contributions virtually in every single area of statistics, impacting the academic life of several generations of statisticians. It is needless to say that the legacy left behind by Professor Rao will continue its impact on future generations of statisticians as well.

Two major contributions of Professor Rao for which he is most well known in the general scientific community, crossing the boundaries of statisics, are the Cramer-Rao inequality and the Rao-Blackwell Theorem. What is indeed more remarkable is that these results constituted part of the Masters Thesis of Professor Rao. These two fundamental results, accessible even to beginning undrgraduate students in statistics, have far reaching implications, well beyond what possibly was envisaged by their authors. For example, the Rao-Blackwell Theorem, discovered independently by Rao and Blackwell, involves an implicit idea of projection from a certain space of random variables to a second space, spanned by sufficient statistics, resulting thereby in loss reduction under convexity.

The other work, namely the Cramer-Rao inequality, discovered independently by Cramer and Rao, is being used repeatedly by scientists even outside statistics, notably by those working in Quantum Physics, Electrical Engineering and Computer Science.

Early extensions of the Cramer-Rao inequality appear in the articles of Bhattacharyya (1946), Hammersley (1950) and Chapman and Robbins (1951). More recently, there has been a surge of extensions of this inequality, primarily for solving problems in science and engineering, as mentioned in the preceding paragraph.

Corresponding Author: Malay Ghosh
Email: ghoshm@ufl.edu

One very useful extension of the Cramer-Rao inequality appears in the book of Van Trees (2004) who provided a lower bound for the Bayes risk of estimators of one dimensional parameters of interest. This was followed later in a series of articles of Bobrovsky *et al.* (1987), Borovkov and Sakhanenko (1980), Brown and Gajek (1990) and many others. Very important consequences of these results leading to local asymptotic minimaxity in the spirit of Hajek and LeCam are proved in Gill and Levit (1995) and Gassiat *et al.* (2013).

In contrast to the above, Ghosh (1993), obtained Cramer-Rao type bounds for posterior variances. Whereas the original Cramer-Rao inequality is based on the Fisher Information number based on the likelihood and the Bayes risk results involve Fisher information number of both the likelihood and the prior, the lower bound obtained by Ghosh involves the posterior analog of the classical Fisher information number.

The present work extends the work of Ghosh (1993) to the multiparameter case, quite in the spirit of Van Trees (2004) as well as Gill and Levit (1995). In particular, for a vector valued parameter, a lower bound is provided for posterior expected weighted squared norms of the difference of parameter vectors and their posterior means. The technical details are given in the following section.

2. The main results

We begin with the posterior lower bound for the variance-covariance matrix of a vector-valued parameter. Indeed, the same lower bound can be provided for the posterior mean squared error matrix for an arbitrary estimator of a parameter of interest. But the sharpest bound is one for the posterior variance-covariance matrix, since for an arbitrary estimator $\boldsymbol{e}(\boldsymbol{X})$ of a parameter vector $\boldsymbol{\psi}(\boldsymbol{\theta})$,

$$\begin{aligned} & E[(\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{e}(\boldsymbol{X}))(\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{e}(\boldsymbol{X}))^T|\boldsymbol{X}] \\ & = V[\boldsymbol{\psi}(\boldsymbol{\theta})|\boldsymbol{X}] + [E(\boldsymbol{\psi}(\boldsymbol{\theta})|\boldsymbol{X}) - \boldsymbol{e}(\boldsymbol{X}))(E(\boldsymbol{\psi}(\boldsymbol{\theta})|\boldsymbol{X}) - \boldsymbol{e}(\boldsymbol{X}))^T] \\ & \geq V[\boldsymbol{\psi}(\boldsymbol{\theta})|\boldsymbol{X}]. \end{aligned}$$

Throughout this section, we will consider the following set up. Let X be a real or vector-valued random variable with pdf $f(x|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$. We denote by $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ the lower and upper end points of $\boldsymbol{\theta}$. Consider an arbitrary prior $\pi(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$. We will denote the posterior of $\boldsymbol{\theta}$ given X by $\pi(\boldsymbol{\theta}|x)$. Consider an $s(\leq p)$-dimensional function $\psi(\boldsymbol{\theta}) = (\psi_1(\boldsymbol{\theta}), \ldots, \psi_s(\boldsymbol{\theta}))^T$ of $\boldsymbol{\theta}$. We begin with the following lemma.

Lemma 1: Let (i) $\pi(\boldsymbol{\theta}|x) \to 0$ as $\boldsymbol{\theta} \to \boldsymbol{r}_1$ or $\boldsymbol{r}_2$ and (ii) $\psi_i(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|x) \to 0$ as $\boldsymbol{\theta} \to \boldsymbol{r}_1$ or $\boldsymbol{r}_2$ for all $i = 1, \ldots, s$. Then

$$E[\psi(\boldsymbol{\theta})\{\nabla \log \pi(\boldsymbol{\theta}|x)\}^T|x] = -E(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}}|x),$$

where ∇ denotes the gradient operator.

Proof: In view of assumptions (i) and (ii), for all $1 \leq i \leq s$ and $1 \leq j \leq p$, integration by parts yields

$$\begin{aligned} E[\psi_i(\boldsymbol{\theta})\frac{\partial}{\partial\theta_j}\nabla\log\pi(\boldsymbol{\theta}|x)|x] &= \int_{\boldsymbol{r}_1}^{\boldsymbol{r}_2}\psi_i(\boldsymbol{\theta})[(\frac{\partial}{\partial\theta_j}\pi(\boldsymbol{\theta}|x))/\pi(\boldsymbol{\theta}|x)]\pi(\boldsymbol{\theta}|x)d\boldsymbol{\theta} \\ &= -\int_{\boldsymbol{r}_1}^{\boldsymbol{r}_2}\frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\theta_j}\pi(\boldsymbol{\theta}|x)d\boldsymbol{\theta} = -E[\frac{\partial\psi_i(\boldsymbol{\theta})}{\partial\theta_j}|x]. \end{aligned}$$

This proves the result.

We now prove the first main result of this section. The result provides a multiparameter posterior Cramer-Rao type lower bound for a vector-valued function of parameters.

Theorem 1: Assume the conditions of Lemma 1, and assume in addition that $V[\nabla\log\pi(\boldsymbol{\theta}|x)]$ is positive definite. Then

$$V[\boldsymbol{\psi}(\boldsymbol{\theta})|x] \geq E(\frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\theta}}|x)[V(\nabla\log\pi(\boldsymbol{\theta}|x))]^{-1}E[(\frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\theta}})^T|x].$$

Proof: Let $\boldsymbol{u} = \boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|x)$ and $\boldsymbol{v} = \nabla\log\pi(\boldsymbol{\theta}|x)$. Consider the matrix

$$E[\begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix}(\boldsymbol{u}^T\boldsymbol{v}^T)|x] = E\left[\begin{pmatrix} \boldsymbol{u}\boldsymbol{u}^T\boldsymbol{u}\boldsymbol{v}^T \\ \boldsymbol{v}\boldsymbol{u}^T\boldsymbol{v}\boldsymbol{v}^T \end{pmatrix}|x\right],$$

which by construction is non negative definite. This immediately leads to

$$E(\boldsymbol{u}\boldsymbol{u}^T|x) \geq [E(\boldsymbol{u}\boldsymbol{v}^T|x)][E(\boldsymbol{v}\boldsymbol{v}^T|x)]^{-1}[E(\boldsymbol{v}\boldsymbol{u}^T|x)].$$

In view of Assumption (i), $E(\boldsymbol{v}^T|x) = \mathbf{0}$. Also, $E(\boldsymbol{v}\boldsymbol{v}^T|x) = V[\nabla\log\pi(\boldsymbol{\theta}|x))|x]$. The conclusion follows now by applying Lemma 1.

Remark 1: In the particular case when $\boldsymbol{\psi}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ so that $\frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\theta}} = \boldsymbol{I}_p$, one gets the inequality $V(\boldsymbol{\theta}|x) \geq [V(\nabla\log\pi(\boldsymbol{\theta}|x))|x]^{-1}$. The classical Cramer-Rao inequality says that for unbiased estimators $\boldsymbol{T}(X)$ of a parameter vector $\boldsymbol{\theta}$, $V[\boldsymbol{T}(X)|\boldsymbol{\theta}] \geq \boldsymbol{I}^{-1}(\boldsymbol{\theta})$, where $\boldsymbol{I}(\boldsymbol{\theta})$ denotes the Fisher Information matrix. $V(\nabla\log\pi(\boldsymbol{\theta}|x))$ is the posterior analog of the classical Fisher Information matrix. It may be noted that while $\boldsymbol{I}(\boldsymbol{\theta}) = V[(\frac{\partial\log f(X|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})|\boldsymbol{\theta}]$, $V[\nabla\log\pi(\boldsymbol{\theta}|x)|x] = V[(\frac{\partial\log\pi(\boldsymbol{\theta}|x}{\partial\boldsymbol{\theta}})|x]$.

Remark 2: Equality holds in Theorem 1 when $\boldsymbol{\psi}(\boldsymbol{\theta}) - E[\boldsymbol{\psi}(\boldsymbol{\theta})|x]$ and $\nabla\log\pi(\boldsymbol{\theta}|x)$ are linearly related. As a simple example, consider $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n|\boldsymbol{\theta}$ are iid $\mathrm{N}(\boldsymbol{\theta},\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ known, and the prior $\pi(\boldsymbol{\theta})$ is $\mathrm{N}(\boldsymbol{\mu},\boldsymbol{\Lambda})$. Then the posterior $\pi(\boldsymbol{\theta}|x)$ is $\mathrm{N}((\boldsymbol{I}-\boldsymbol{B})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{\mu},(\boldsymbol{I}-\boldsymbol{B})\boldsymbol{\Sigma}/n)$, where $\boldsymbol{B} = n^{-1}\boldsymbol{\Sigma}(n^{-1}\boldsymbol{\Sigma}+\boldsymbol{\Lambda})^{-1}$. Then

$$\nabla\log\pi(\boldsymbol{\theta}|x) = [(\boldsymbol{I}-\boldsymbol{B})\boldsymbol{\Sigma}/n)^{-1}(\boldsymbol{\theta}-((\boldsymbol{I}-\boldsymbol{B})\boldsymbol{x}+\boldsymbol{B}\boldsymbol{\mu})).$$

Then $\nabla\log\pi(\boldsymbol{\theta}|x)$ is linearly related to $\boldsymbol{\theta}$ and accordingly $V(\boldsymbol{\theta}|x) = [V(\nabla\log\pi(\boldsymbol{\theta}|x))|x]^{-1}$.

Remark 3: It is important to point out that while the classical Cramer-Rao inequality is based on $\nabla\log f(x|\boldsymbol{\theta})$, the van Tress inequality is based on $\nabla\log(f(x|\boldsymbol{\theta})\pi(\boldsymbol{\theta}))$. Ours is based on $\nabla\log\pi(\boldsymbol{\theta}|x)$ instead.

Our next result provides a lower bound for $E[||\boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|x)||^2|x] = E[\mathrm{tr}(V(\boldsymbol{\psi}(\boldsymbol{\theta})|x))]$.

Theorem 2: $E[\mathrm{tr}(V(\boldsymbol{\psi}(\boldsymbol{\theta})|x)] \geq E[\{\mathrm{tr}(\frac{\partial\boldsymbol{\psi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})\}^2|x]/E[||\nabla \log \pi(\boldsymbol{\theta})|x||^2|x]$.
Proof: In view of Lemma 1, it follows that

$$\mathrm{tr}E[\{\boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|x)\}\{\nabla \log \pi(\boldsymbol{\theta}|x)\}^T|x] = -\mathrm{tr}E[(\frac{\partial\boldsymbol{\psi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})|x] = -E[\mathrm{tr}(\frac{\partial\boldsymbol{\psi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})|x].$$

The above is equivalent to

$$-E[\mathrm{tr}(\frac{\partial\boldsymbol{\psi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})|x] = E[\{(\nabla \log \pi(\boldsymbol{\theta}|x)^T(\boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|x))\}|x].$$

Now an application of the Cauchy-Schwarz inequality yields

$$[E\mathrm{tr}(\frac{\partial\boldsymbol{\psi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})|x]^2 \leq E[||\nabla \log \pi(\boldsymbol{\theta}|x)||^2|x]E\{(||\boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|x)||^2)|x\}.$$

This yields the result. Remark 4. The above result can be generalized easily. Note that for an arbitrary positive definite matrix $\boldsymbol{B}$, and two random vectors $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$,

$$E(\boldsymbol{Z}_1^T\boldsymbol{Z}_2) = E(\boldsymbol{Z}_1^T\boldsymbol{B}^{-1/2}\boldsymbol{B}^{1/2}\boldsymbol{Z}_2) \leq E(\boldsymbol{Z}_1^T\boldsymbol{B}^{-1}\boldsymbol{Z}_1)E(\boldsymbol{Z}_2^T\boldsymbol{B}\boldsymbol{Z}_2).$$

Writing $\boldsymbol{Z}_1 = \boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|x)$ and $\boldsymbol{Z}_2 = \nabla \log \pi(\boldsymbol{\theta}|x)$, one gets the weighted squared error posterior risk

$$E[(\boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|x))^T\boldsymbol{B}^{-1}(\boldsymbol{\psi}(\boldsymbol{\theta}) - E(\boldsymbol{\psi}(\boldsymbol{\theta})|xa))] \geq [E\mathrm{tr}(\frac{\partial\boldsymbol{\psi}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}})|x]^2/E[(\nabla \log \pi(\boldsymbol{\theta}|x))^T\boldsymbol{B}\nabla \log \pi(\boldsymbol{\theta}|x)|x].$$

References

Bhattacharyya, A. (1946). On some analogues of the amount of information and their use in statistical estimation. *Sankhyā: The Indian Journal of Statistics*, **1**, 1–14.

Bobrovsky, B.-Z., Mayer-Wolf, E., and Zakai, M. (1987). Some classes of global Cramér-Rao bounds. *The Annals of Statistics*, **1**, 1421–1438.

Borovkov, A. and Sakhanenko, A. (1980). On estimates of the expected quadratic risk (in russian). *Probability and Mathematical Statistics*, **1**, 185–95.

Brown, L. D. and Gajek, L. (1990). Information inequalities for the Bayes risk. *The Annals of Statistics*, **18**, 1578–1594.

Chapman, D. G. and Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *The Annals of Mathematical Statistics*, **1**, 581–586.

Gassiat, E., Pollard, D., and Stoltz, G. (2013). Revisiting the van Trees inequality in the spirit of Hajek and Le Cam. *Unpublished Manuscript*, .

Ghosh, M. (1993). Cramer-Rao bounds for posterior variances. *Statistics & Probability Letters*, **17**, 173–178.

Gill, R. D. and Levit, B. Y. (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, **1**, 59–79.

Hammersley, J. M. (1950). On estimating restricted parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, **12**, 192–240.

Van Trees, H. L. (2004). *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. Wiley, New York.