CRRAO Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS)

Research Report

Author (s): **B.L.S. Prakasa Rao**

Title of the Report: **Improved Cramér-Rao Type Integral Inequalities or Bayesian Cramér-Rao Bounds**

Research Report No.: **RR2017-06**

Date: **May 24, 2017**

Prof. C R Rao Road, University of Hyderabad Campus,
Gachibowli, Hyderabad-500046, INDIA.
www.crraoaimscs.org

Improved Cramér-Rao Type Integral Inequalities or Bayesian Cramér-Rao Bounds

B.L.S. Prakasa Rao
CR Rao Advanced Institute for Research in Mathematics, Statistics and Computer science, Hyderabad 500046, India

Abstract: New lower bounds on the mean square error for estimators of random parameter are obtained as applications of improved Cauchy-Schwarz inequality due to Walker (Statist. Probab. Lett. 122 (2017), 86-90).

1 Introduction

Cramér-Rao lower bound for the variance of an unbiased estimator of a parameter is well known for its use in statistical literature. There has been a large amount of work to obtain Cramér-Rao type integral inequalities leading to lower bounds for the risks associated with Bayesian estimators. Earlier results in this diretion are due to Schtzenberger (1957) and Gart (1959). Other works in this direction in the statistical literature are due to Borovkov and Sakhanenko (1980), Targhetta (1984, 1988, 1990), Shemyakin (1987), Babrovsky et al. (1987), Brown and Gajek (1990), Prakasa Rao (1992), Ghosh (1993) and Gill and Levit (1995). In engineering literature, this problem is considered under the subject "random parameter estimation". Significant results in this area in the engineering literature are due to Van trees (1968), Ziv and Zakai (1969), Chazan et al. (1975), Miller and Chang (1978), Weinstein and Weiss (1985), Weiss and Weinstein (1985), Brown and Liu (1993) among others. Prakasa Rao (1991) gives a comprehensive survey of results obtained in this area till about 1990. Related results on Cramér-Rao type integral inequalities were obtained in Prakasa Rao (1996, 2000, 2001). In a voluminous work, van Trees and Bell (2007) give a survey of Bayesian bounds for parameter estimation and nonlinear filtering/tracking and edited a volume containing selected papers dealing with Bayesian Cramér-Rao bounds, global Bayesian bounds, hybrid Bayesian bounds, constrained Cramér-rao bounds and their applications to nonlinear dynamic systems.

It is well known that either the Cramér-Rao inequality giving a lower bound for the quadratic risk of an estimator or the Bayesian versions of the Cramér-Rao inequality obtained by several authors are all consequences or applications of the Cauchy-Schwarz inequality for suitable functions of the observations and the parameter. In a recent paper, Walker

(2017) obtained an improved Cauchy-Schwarz inequality. Our aim in this short note is to obtain some Bayesian Cramér-Rao bounds as applications of the improved version of Cauchy-Schwarz inequality. Walker (2017) obtained a generalized Cramér-Rao inequality as an application of the improved Cauchy-Schwarz inequality.

2 Main results

Walker (2017) obtained an improved version of the Cauchy-Schwarz inequality which implies the following probabilistic version.

Theorem 2.1 : *If X and Y are random variables defined on a probabilty space $(\Omega, \mathcal{F}, P)$ with finite second moments, then*

$$|E(XY)|^2 \leq E(X^2)E(Y^2) - (|E(X)|\sqrt{\mathrm{Var}(Y)} - |E(Y)|\sqrt{\mathrm{Var}(X)})^2. \tag{2. 1}$$

As has been pointed out by Walker (2017), the inequality (2.1) is a strict improvement over the Cauchy-Schwarz inequality. This can be seen from the following example due to Walker (2017). Suppose Y is a random variable with mean zero and variance 1 and X is a random variable with mean μ and finite variance σ^2. Then the Cauchy-Schwarz inequality implies

$$[E(XY)]^2 \leq E(X^2)E(Y^2) = (\sigma^2 + \mu^2) \tag{2. 2}$$

where as Theorem 2.1 implies that

$$[E(XY)]^2 \leq \mathrm{Var}(X)E(Y^2) = \sigma^2. \tag{2. 3}$$

It is obvious that the upper bound given by the inequality (2.3) is better than the upper bound given by the inequality (2.2).

Suppose a random variable Y has mean zero but positive variance and X is another random variable with finite variance. Then it follows that

$$|E(XY)|^2 \leq E(X^2)E(Y^2) - |E(X)|^2E(Y^2) = \mathrm{Var}(X)E(Y^2). \tag{2. 4}$$

by Theorem 2.1. Hence

$$E(X^2) \geq (E(X))^2 + \frac{|E(XY)|^2}{E(Y^2)}$$

and we have the following corollary.

Corollary 2.1: *If X and Y are random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ with finite second moments and if $E(Y) = 0$, , then*

$$E(X^2) \geq (E(X))^2 + \frac{|E(XY)|^2}{E(Y^2)}. \tag{2. 5}$$

We now discuss some applications of the inequality derived in Corollary 2.1.

Let Z be a random variable defined on a probability space $(\Omega, \mathcal{F}, P_\theta)$ where $\theta \in \Theta \subset R$. Suppose that the parameter θ has a prior density $\lambda(\theta)$ with respect to the Lebesgue measure on R. Let us consider a function $\psi(z, \theta)$ such that $E_\theta[\psi(Z, \theta)|Z] = 0$ where $E_\theta(\psi(Z, \theta)|Z)$ denotes the expectation of the random variable $\psi(Z, \theta)$ with respect to the posterior distribution of θ given Z. Let $E(\psi(Z, \theta))$ denote the expectation of the random variable $\psi(Z, \theta)$ with respect to the joint distribution of the random vector (Z, θ). Then, for any random variable $\ell(Z)$, with $E[|\ell(Z)|] < \infty$,

$$E(\ell(Z)\psi(Z, \theta)) = E[E(\ell(Z)\psi(Z, \theta)|Z)] = E[\ell(Z)E(\psi(Z, \theta)|Z)] = 0 \tag{2. 6}$$

and hence

$$E((\theta - \ell(Z))\psi(Z, \theta)) = E(\theta\psi(Z, \theta)). \tag{2. 7}$$

Applying the inequality given in Corollary 2.1 for the random variable $X = \theta - \ell(Z)$ and for the random variable $Y = \psi(Z, \theta)$ with conditional mean zero given the random variable Z and finite second moment, we obtain that

$$E([\theta - \ell(Z)]^2) \geq (E[\theta - \ell(Z)])^2 + \frac{(E[\theta\psi(Z, \theta)])^2}{E([\psi(Z, \theta)]^2)}. \tag{2. 8}$$

Special cases

(i) Suppose we choose $\psi(Z, \theta) = \theta - E(\theta|Z)$. It is obvious that $E[\psi(Z, \theta)|Z] = 0$. Applying the inequality (2.8), we get that

$$\begin{aligned}
E([\theta - \ell(Z)]^2) &\geq (E[\theta - \ell(Z)])^2 + \frac{(E[\theta\psi(Z, \theta)])^2}{E([\psi(Z, \theta)]^2)} \\
&= (E[\theta - \ell(Z)])^2 + \frac{(E[\theta(\theta - E(\theta|Z))])^2}{E([\theta - E(\theta|Z)]^2)} \\
&= (E[\theta - \ell(Z)])^2 + \frac{(E[(\theta - E(\theta|Z))(\theta - E(\theta|Z))])^2}{E([\theta - E(\theta|Z)]^2)} \\
&= (E[\theta - \ell(Z)])^2 + E([\theta - E(\theta|Z)]^2).
\end{aligned}$$

(ii) Let $\lambda(.)$ denote a prior density of the parameter θ and suppose that $f(z,\theta)$ is the probability density function of a random variable Z given the parameter θ. Then the joint density of the random vector (Z,θ) is $g(z,\theta) = f(z,\theta)\lambda(\theta)$. Let $\pi(\theta|z)$ denote the posterior density function of the parameter θ given the observation z. Let $I(\theta)$ denote the Fisher information in the observation Z given the parameter θ. Suppose we choose

$$\psi(z,\theta) = \frac{\partial \log(\pi(\theta|z))}{\partial\theta}.$$

Observe that $E[\psi(Z,\theta)|Z] = 0$ and it is easy to check that

$$E([\theta - \ell(Z)]^2) \geq (E[\theta - \ell(Z)])^2 + \frac{(E[(\theta - \ell(Z))\psi(Z,\theta)])^2}{E([\psi(Z,\theta)]^2)}. \tag{2. 9}$$

Let

$$I(\lambda) = E[(\frac{\partial \log \lambda(\theta)}{\partial\theta})^2]$$

and

$$I(\theta) = E[(\frac{\partial \log f(Z,\theta)}{\partial\theta})^2|\theta].$$

Applying the inequality given by Corollary 2.1, we get that

$$\begin{aligned} E([\theta - \ell(Z)]^2) &\geq (E[\theta - \ell(Z)])^2 + \frac{(E[(\theta - \ell(Z))\psi(Z,\theta)])^2}{E([\psi(Z,\theta)]^2)} \\ &= (E[\theta - \ell(Z)])^2 + \frac{(E[\theta\psi(Z,\theta)])^2}{E([\psi(Z,\theta)]^2)} \\ &= (E[\theta - \ell(Z)])^2 + \frac{(E[\theta\psi(Z,\theta)])^2}{E(I(\theta)) + I(\lambda)}. \end{aligned}$$

(iii) Let $\lambda(.)$ denote the prior density of the parameter θ and suppose that $f(z,\theta)$ is the probability density function of the random variable Z given the parameter θ. Then the joint density of the random vector (Z,θ) is $g(z,\theta) = f(z,\theta)\lambda(\theta)$. We will now obtain an improved version of the van Trees inequality (cf. van Trees (1968), Gill and Levit (1995)). Let

$$\psi(z,\theta) = \frac{\partial \log(f(z,\theta)\lambda(\theta))}{\partial\theta}.$$

Assuming that the prior density $\lambda(\theta)$ converges to zero as θ tends to the boundary of the set Θ , it follows that

$$\int_\Theta \frac{d[f(z,\theta)\lambda(\theta)]}{d\theta} d\theta = [f(z,\theta)\lambda(\theta)]_{\partial\Theta} = 0 \tag{2. 10}$$

and

$$\begin{aligned}\int_{\Theta}\theta\frac{d[f(z,\theta)\lambda(\theta)]}{d\theta}d\theta &= [\theta f(z,\theta)\lambda(\theta)]_{\partial\Theta}-\int_{\Theta}f(z,\theta)\lambda(\theta)d\theta\\ &= -\int_{\Theta}f(z,\theta)\lambda(\theta)d\theta.\end{aligned}$$

Using the above equations, it follows that

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{\Theta}(\theta-\ell(z))\frac{d[f(z,\theta)\lambda(\theta)]}{d\theta}d\theta dz &= \int_{-\infty}^{\infty}\int_{\Theta}f(z,\theta)\lambda(\theta)d\theta dz\\ &= 1.\end{aligned}$$

Observe that $E[\psi(Z,\theta)|Z]=0$. Applying Corollary 2.1, we get that

$$\begin{aligned}E([\theta-\ell(Z)]^2) &\geq (E[\theta-\ell(Z)])^2+\frac{(E[(\theta-\ell(Z))\psi(Z,\theta)])^2}{E([\psi(Z,\theta)]^2)}\\ &= (E[\theta-\ell(Z)])^2+\frac{(E[\theta\psi(Z,\theta)])^2}{E([\psi(Z,\theta)]^2)}\\ &= (E[\theta-\ell(Z)])^2+\frac{(E[\theta\psi(Z,\theta)])^2}{E(I(\theta))+I(\lambda)}.\end{aligned}$$

(iv) Let $\lambda(.)$ denote the prior density of the parameter θ and suppose that $f(z,\theta)$ is the probability density function of the random variable Z given the parameter θ. Then the joint density of the random vector (Z,θ) is $g(z,\theta)=f(z,\theta)\lambda(\theta)$. Define the likelihood ratio given by

$$L(z;\theta_1,\theta_2)=\frac{g(z,\theta_1)}{g(z,\theta_2)}.$$

For any fixed $h\neq 0$, and $0<s<1$, define

$$\psi(z,\theta)=L^s(z;\theta+h,\theta)-L^{1-s}(z;\theta-h,\theta).$$

Following Weiss and Weinstein (1985), it follows that

$$E[\ell(Z)\psi(Z,\theta)]=0$$

and

$$E[\theta\psi(Z,\theta)]=-hE[L^{1-s}(Z;\theta-h,\theta)]$$

(cf. Weiss and Weinstein (1985)). As an application of Corollary 2.1, we get that

$$E([\theta-\ell(Z)]^2)\geq(E[\theta-\ell(Z)])^2+\frac{h^2(E[L^{1-s}(Z;\theta-h,\theta)])^2}{E[\psi(Z,\theta)^2]}. \tag{2. 11}$$

Following arguments given in Weiss and Weinstein (1985), it follows that

$$E([\theta - \ell(Z)]^2) \geq (E[\theta - \ell(Z)])^2 + \frac{h^2 e^{2\mu(s,h)}}{e^{\mu(2s,h)} + e^{\mu(2s-1),h)} - 2e^{\mu(s,2h)}} \tag{2. 12}$$

where

$$\begin{aligned}\mu(s,h) &= \log E[L^s(Z;\theta+h,\theta)] \\ &= \log[\int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} [g(z,\theta+h)]^s [g(z,\theta)]^{1-s} d\theta]\end{aligned}$$

where $g(z,\theta)$ is the joint probability density function of the random vector (Z,θ).

In a similar fashion, it is possible to improve other lower bounds for the risk of Bayesian estimators using Corollary 2.1 as applications of the improved Cauchy-Schwarz inequality due to Walker (2017) and also obtain similar Bayesian bounds for functions of a parameter.

References:

Babrovsky, B.Z., Mayer-Wolf, E. and Zakai, M. (1987) Some classes of global Cramér-Rao bounds, *Ann. Statist.*, **15**, 1421-1438.

Borovkov, A.A. and Sakhanenko, A.I. (1980) On estimates for the average quadratic risk, *Probab. Math, Statist.*, **1**, 185-195. (In Russian).

Brown, L.D. and Gajek, L. (1990) Information inequalities for the Bayes risk, *Ann. Statist.*, **18**, 1578-1594.

Brown, L.D. and Liu, R.C. (1993) Bounds on the Bayes and minimax risk for signal parameter estimation, *IEEE Trans. Inform. Theory*, **IT-39**, 1386-1394.

Chazan, D., Ziv,J. and Zakai, M. (1975) Improved lower bounds on signal parameter estimation, *IEEE Trans. Inform. Theory*, **IT-21**, 90-93.

Gart, John J. (1959) An extension of Cramér-Rao inequality, *Annals. Math. Stat.*, **30**, 367-380.

Ghosh, Malay. (1993) Cramér-Rao bounds for posterior variances, *Statist. Probab. Lett.*, **17**, 173-178.

Gill, Richard D. and Levit, Borris Y. (1995) Application of the van Trees inequality: a Batesian Cramér-Rao bound, *Bernoulli*, **1**, 59-79.

Miller, R. and Chang, C. (1978) A modified Cramér-Rao bound and its applications, *IEEE Trans. Inform. Theory*, **IT-24**, 398-400.

Prakasa Rao, B.L.S. (1991) On Cramér-Rao type integral inequalities, *Calcutta Statist. Assoc. Bull. H.K. Nandi Mem. spec. Vol.*, **40**, 183-205; Reprinted in *Bayesian Bounds for Parameter estimation and Nonlinear Filtering/Tracking*, Ed. H. van Trees and K.L. bell, IEEE Press, Wiley-Interscience, John Wiley and Sons, New York, pp. 900-922.

Prakasa Rao, B.L.S. (1992) Cramér-Rao type integral inequalities for functions of multidimensional parameter, *Sankhya Series A*, **54**, 53-73.

Prakasa Rao, B.L.S. (1996) Remarks on Cramér-Rao type integral inequalities for randomly censored data, In *Analysis of Censored Data*, IMS Lecture Notes No. 27, Ed. H.L. Koul and J.V. Deshpande, Institute of Mathematical Statistics, USA, pp. 160-176.

Prakasa Rao, B.L.S. (2000) Cramér-Rao type integral inequalities in Banach spaces, In *Perspectives in Statistical Sciences*, Ed. A.K. Basu, J.K. Ghosh, P.K. Sen and B.K. Sinha, Oxford University Press, New Delhi, pp. 245-260.

Prakasa Rao, B.L.S. (2001) Cramér-Rao type integral inequalities for general loss functions, *TEST*, **10**, 105-120.

Schutzenberger, M.P. (1957) A generalization of the Frechet-Cramér inequality to the case of Bayes estimation, *Bull. Amer. Math. Soc.*, **63**, 142.

Shemyakin, M.L. (1987) Rao-Cramér type integral inequalities for estimates of a vector parameter, *Theory Probab. Appl.*, **32**, 426-434.

Targhetta, M. (1984) On Bayesian analogues to Bhattacharya's lower bounds, *Arab Gulf J. Sci. Res.*, **2**, 583-590.

Targhetta, M. (1988) On the attainment of a lower bound for the Bayes risk in estimating a parametric function, *Statistics*, **19**, 233-239.

Targhetta, M. (1990) A note on the mixing problem and the Schutzenberger inequality, *Metrika*, **37**, 155-161.

van Trees, Harry L. (1968) *Detection, Estimation and Modulation Theory Part 1*, Wiley, New York.

van Trees, Harry L. and Bell, Kristine L. (2007) *Bayesian Bounds for Parameter estimation and Nonlinear Filtering/Tracking*, IEEE Press, Wiley-Interscience, John Wiley and Sons, New York.

Walker, S.G. (2017) A self-improvement to the Cauchy-Schwarz inequality, *Statist. Probab. Lett.*, **122**, 86-90.

Weinstein, E. and Weiss, A. (1985) Lower bounds on the mean square estimation error, *Proceedings of the IEEE*, **73**, 1433-1434.

Weiss, A. and Weinstein, E. (1985) A lower bound on the mean square error in random parameter estimation, *IEEE Trans. Inform. Theory*, **IT-31**, 680-682.

Ziv, J. and Zakai, M. (1969) Some lower bounds on signal parameter estimation, *IEEE Trans. Inform. Theory*, **IT-15**, 386-391.