Proc. Indian Acad. Sci. (Math. Sci.) (2020) 130:41
https://doi.org/10.1007/s12044-020-0580-y

Cramér–Rao inequality revisited

B L S PRAKASA RAO

C.R. Rao Advanced Institute of Mathematics, Statistics and Computer Science,
Hyderabad 500 046, India
E-mail: blsprao@gmail.com

Abstract. Among C R Rao's many contributions to statistical inference, one which has been and still is considered to be of extreme importance in the areas of statistics, physics and of signal processing in electrical engineering beside other sciences is an inequality which is now known as the Cramér–Rao inequality. This result was studied in recent years by several other scientists to relax the conditions under which it holds and to generalize it in different directions. Contributions by Bhattacharya (*Sankhya* **8** (1946) 1–14, 201–218, 315–328), Barankin (*Ann. Math. Statist.* **20** (1949) 477–501) and Fabian and Hannan (*Ann. Statist.* **5** (1977) 197–205) are significant in this area. We do not propose to give an extensive survey of results connected with the inequality. Our aim in this communication is to highlight some recent advances.

Keywords. Cramér–Rao inequality; generalized Cramér–Rao inequality; Cramér–Rao type integral inequality.

2020 Mathematics Subject Classification. 62F10, 62F15.

1. Introduction

Calyampudi Radhakrishna Rao (aka) C R Rao was born on September 10, 1920 in Huvanna Hadagali, now in the state of Karnataka. He received M.A. in Mathematics from the Andhra University in 1940, M.A. in Statistics from the Calcutta University in 1943, Ph.D. from the Cambridge University in 1948 and a D.Sc. from the same university in 1965. He is a household name in the family of statisticians, electrical engineers, physicists, econometricians and in general, the scientific community. He has extensively contributed to estimation theory, multivariate analysis, characterization problems, combinatorics, design of experiments, matrix algebra, generalized inverses of matrices, differential geometric methods in statistics and to statistical genetics among others. Among his many contributions to statistical inference, one which has been and still is considered to be of extreme importance in the areas of statistics, physics and of signal processing in electrical engineering beside other sciences is an inequality which is now known as the *Cramér–Rao inequality* (cf. [31]). This result was studied in recent years by several scientists to relax the conditions under which it holds and to generalize it in different directions. Contributions by Bhattacharya [4], Barankin [1] and by Fabian and Hannan [12] are significant in this area. We do not propose to give an extensive survey of results connected with the inequality. As per style minimum three keywords are needed. Only two keywords are present in the article, please provide the keywords. Our aim in this communication is to highlight some recent advances.

This article is part of the "Special Issue in Honour of Professor C R Rao on His Birth Centenary"

Published online: 24 July 2020

One of the basic problems of statistical inference is to obtain information or estimation about an unknown parameter of a population based on an observed sample drawn from the population according to some sampling scheme. A natural question is how to estimate the unknown parameter or a known function of it and how to compare one method of estimation with another. A criterion or criteria have to be specified to judge the performance of the estimators before deciding if one estimator is better than another estimator.

Suppose X is a random variable with distribution function $F(., \theta)$, that is,

$$P(a < X \leq b) = F(b, \theta) - F(a, \theta)$$

whenever $-\infty < a \leq b < \infty$, where the parameter θ is a scalar parameter or a vector parameter. If the parameter θ is known and the function F is known, then the above equation gives the probability that an observation X belongs to the interval $(a, b]$. Typically the parameter θ is unknown and the parameter θ or a known function of the parameter $g(\theta)$ is of interest and has to be estimated from a sample of observations $x_1, \ldots, x_n$ based on the random variable X. Any function T depending on the observations $X_1, \ldots, X_n$ only is called a *statistic*. If the statistic T, evaluated at the observations $x_1, \ldots, x_n$, is chosen as an estimate for the parameter θ, then the $T(x_1, \ldots, x_n)$ is called an *estimate*. It is obvious that there are many different functions T of $X_1, \ldots, X_n$ which can be chosen as estimators for the unknown parameter θ. Not all of them are suitable. An estimator T is said to be an *unbiased* estimator of a function $g(\theta)$ if $E_\theta T = g(\theta)$, that is the expected value of the estimator T, when θ is the true value of the parameter, is equal to the function $g(\theta)$ for all values of parameter θ. If this property does not hold, then the estimator T is a *biased* estimator. The bias of the estimator T is $E(T) - g(\theta)$. One way of measuring the error or the loss involved in associating the random variable T as an estimator of the function $g(\theta)$ is to consider) the difference $T - g(\theta)$. This difference can be positive or negative. The mean squared error of the estimator T is $E_\theta[|T - g(\theta)|^2]$ when θ is the true parameter. For unbiased estimators T of the parametric function $g(\theta)$, the mean squared error is the variance of the estimator T.

The problem is how to choose the unbiased estimator T with least variance. This is where C R Rao has made a fundamental contribution which is now known as the Cramér–Rao inequality. Rao obtained what is now known as Cramér–Rao lower bound with applications to statistical inference, econometrics, signal processing, quantum information and several other fields.

In the branch of statistical inference, Cramér–Rao bound (CRB), Cramér–Rao lower bound (CRLB), Cramér–Rao inequality, Frechet–Darmois–Cramér–Rao inequality or some times termed as information inequality expresses a lower bound on the variance of an unbiased estimator of an unknown parameter. This inequality is named in honour of Cramér [8], Rao [31], Frechet [13] and Darmois [9], all of whom independently derived it in 1940's.

On some history: At the young age of 24, C R Rao was teaching a course on estimation to master's students of Calcutta University and was explaining to the class the concept of asymptotic efficiency of an estimator and mentioned Fisher's result that the asymptotic variance of a consistent estimator is bounded below by the reciprocal of Fisher information. One of the students asked whether such a result can be established for finite samples. Rao went back home, worked all night and proved the inequality the next day which is now known as the Cramér–Rao inequality (cf. Ghosh *et al.* [14], p. 160). A paper containing this result was published in 1945 (Rao [31]). Although the paper was completed in 1943, due to delays in publication and suspension of many journals including *Sankhya* during the war period, the results could not be published earlier. Rao sent a reprint of his paper to Professor J. Neyman who named the information bound as the Cramér–Rao inequality (cf. Ghosh *et al.* [14], pp. 151–213). For more details about C R Rao, regarding his life and his association with statistics and statisticians, see [2,10,14,26] and [30].

Cramér–Rao lower bound (CRLB) indicates that the variance of any unbiased estimator of a parameter cannot be lower than the reciprocal of the Fisher information (to be defined later) under some regularity conditions. Any unbiased estimator which achieves this lower bound is said to be *efficient*. If such an estimator exists, then it is termed as the *minimum variance unbiased estimator*. However, in some cases, no unbiased estimator might exist which achieves the Cramér–Rao lower bound.

Dembo *et al.* [11] provided a unified treatment of the Cramér–Rao inequality, the Heisenberg uncertainity principle, entropy inequalities, Fisher information linking many other inequalities in statistics, mathematics, information theory and physics. They showed that the Heisenberg uncertainity principle can be derived as a consequence of the Cramér–Rao inequality (cf. [32]). An interesting result due to Stam [34] is the derivation of Weyl–Heisenberg uncertainty principle in physics using a specific version of Cramér–Rao lower bound. For other applications of the Cramér–Rao lower bound in physics, see [33]. Parthasarathy [17] discussed the philosophy of Cramér–Rao and Bhattacharya inequalities in quantum statistics.

2. The classical Cramér–Rao inequality

Suppose $\theta \in \Theta$ is an unknown scalar parameter which is to be estimated from observations of a random variable X distributed according to some probability density function $f(x,\theta)$ in the sense that

$$P_\theta(X \le x) = \int_{-\infty}^{x} f(u,\theta)\mathrm{d}u, \quad -\infty < u < \infty.$$

Note that

$$\int_{-\infty}^{\infty} f(x,\theta)\mathrm{d}x = 1, \quad \theta \in \Theta.$$

One of the basic assumptions for the validity of the Cramér–Rao inequality is that the integral on the left hand side of the equation given above can be differentiated with respect to the parameter θ under the integral sign. As a consequence, it is as follows.

Assumption A.

$$\int_{-\infty}^{\infty} \frac{\partial f(x,\theta)}{\partial \theta}\mathrm{d}x = 0. \tag{2.1}$$

Let $\hat{\theta}(X)$ be an unbiased estimator of the parameter θ, that is, $E_\theta(\hat{\theta}(X)] = \theta$ for all values of $\theta \in \Theta$. Hence

$$\int_{-\infty}^{\infty} \hat{\theta}(x) f(x,\theta)\mathrm{d}x = \theta, \quad \theta \in \Theta.$$

Suppose that differentiation with respect to theta under the integral sign is permitted in the above equation. Differentiating on both sides of the equation given above with respect to θ, it is as follows.

Assumption B.

$$\int_{-\infty}^{\infty} \hat{\theta}(x) \frac{\partial f(x,\theta)}{\partial \theta} \mathrm{d}x = 1, \quad \theta \in \Theta. \tag{2.2}$$

Assumption C. The set $\{x : f(x, \theta > 0\}$ does not depend on the parameter $\theta \in \Theta$.

Hereafter we call Assumptions A, B and C as regularity conditions. Under such assumptions, equations (2.1) and (2.2) imply that

$$\int_{-\infty}^{\infty} (\theta - \hat{\theta}(X)) \frac{\partial f(x,\theta)}{\partial \theta} \mathrm{d}x = 1, \quad \theta \in \Theta. \tag{2.3}$$

Under the Assumption C, equation (2.3) can be written in an alternate way in the form

$$\int_{-\infty}^{\infty} (\theta - \hat{\theta}(X)) \frac{\partial \log f(x,\theta)}{\partial \theta} f(x,\theta) \mathrm{d}x = 1, \quad \theta \in \Theta. \tag{2.4}$$

Recall the Cauchy–Schwartz inequality in its simplest form: for any two square integrable functions $g(.)$ and $h(.)$,

$$\int_{-\infty}^{\infty} [g(x)h(x)]^2 \mathrm{d}x \leq \int_{-\infty}^{\infty} g^2(x) \mathrm{d}x \int_{-\infty}^{\infty} h^2(x) \mathrm{d}x$$

equality occuring if and only if the functions $g(.)$ and $h(.)$ are linearly related. Let us choose $g(x) = (\hat{\theta}(x) - \theta)[f(x,\theta)]^{1/2}$ and $h(x) = \frac{\partial \log f(x,\theta)}{\partial \theta}[f(x,\theta)]^{1/2}$ and apply the Cauchy–Schwartz inequality. As a consequence of equation (2.4), it follows that

$$1 \leq \int_{-\infty}^{\infty} (\hat{\theta}(x) - \theta)^2 f(x,\theta) \mathrm{d}x \int_{-\infty}^{\infty} \left[\frac{\partial \log f(x,\theta)}{\partial \theta}\right]^2 f(x,\theta) \mathrm{d}x. \tag{2.5}$$

Let

$$I(\theta) = \int_{-\infty}^{\infty} \left[\frac{\partial \log f(x,\theta)}{\partial \theta}\right]^2 f(x,\theta) \mathrm{d}x. \tag{2.6}$$

Suppose that $0 < I(\theta) < \infty, \theta \in \Theta$. The function $I(\theta)$ is called the *Fisher information*. An alternate way of writing the inequality (2.7) is

$$1 \leq \mathrm{Var}(\hat{\theta}(X)) I(\theta)$$

leading to the Cramér–Rao inequality

$$\mathrm{Var}(\hat{\theta}(X)) \geq \frac{1}{I(\theta)} \tag{2.7}$$

for unbiased estimator $\hat{\theta}(X)$ of a scalar parameter θ.

Suppose the estimator $\hat{\theta}(X)$ is possibly biased. Let $b(\theta) = E_\theta[\hat{\theta}(X)] - \theta$. The function $b(\theta)$ is called the *bias* of the estimator $\hat{\theta}(X)$. Suppose the function $b(\theta)$ is differentiable with respect to θ with derivative $b'(\theta)$ and few other regularity conditions also hold. It can be shown that

$$E(\hat{\theta}(X) - \theta)^2 \geq \frac{[1 + b'(\theta)]^2}{I(\theta)}. \tag{2.8}$$

The function $M(\hat{\theta}(X)) = E_\theta[(\hat{\theta}(X) - \theta)^2]$ is called the *mean squared error* (MSE) of the estimator $\hat{\theta}(X)$.

3. Cramér–Rao inequality in the multidimensional case

Suppose X is a random variable with the probability density function $f(x, \theta)$ which depends on a d-dimensional vector parameter $\theta = (\theta_1, \ldots, \theta_d)^T \in R^d$, where α^T denotes the transpose of a vector α. Let

$$I_{jk} = E_\theta\left[\frac{\partial \log f(X, \theta)}{\partial \theta_j} \frac{\partial \log f(X, \theta)}{\partial \theta_k}\right]$$

and

$$I(\theta) = ((I_{jk}))_{d\times d},$$

assuming that all the elements of the matrix $I(\theta)$ are finite. The matrix $I(\theta)$ is called the *Fisher information matrix*. Suppose the matrix $I(\theta)$ is non-singular. It is easy to see that the matrix $I(\theta)$ is symmetric. Let $T(X) = (T_1(X), \ldots, T_d(X))^T$ be a statistic. Suppose that $E_\theta[T_j(X)] = \psi_j(\theta)$, $j = 1, \ldots, d$ exist. Let $\psi(\theta) = (\psi_1(\theta), \ldots, \psi_d(\theta)^T$. Suppose that $\psi_i(\theta)$ is differentiable with respect to θ_j for $i, j = 1, \ldots, d$. Let

$$\frac{\partial \psi(\theta)}{\partial \theta} = \left(\left(\frac{\partial \psi_i(\theta)}{\partial \theta_j}\right)\right)_{d\times d}.$$

The following inequality gives a lower bound for the covariance matrix $\mathrm{Cov}_\theta(T(X))$ of the statistic $T(X)$ as an estimator of the function $\psi(\theta)$:

$$\mathrm{Cov}_\theta(T(X)) \geq \frac{\partial \psi(\theta)}{\partial \theta}\left([I(\theta)]^{-1}\frac{\partial \psi(\theta)}{\partial \theta}\right)^T \tag{3.1}$$

where for two square matrices A and B of the same order, the inequality $A \geq B$ indicates that the matrix $A - B$ is positive semi-definite.

4. Improvements on the Cauchy–Schwartz inequality

For real-valued random variables U and V with finite second moment, Cauchy–Schwartz inequality states that

$$E(U^2)E(V^2) \geq [E(UV)]^2, \tag{4.1}$$

equality in (4.1) if and only if U and V are linearly related, that is, there exist constants a, b such that $aV + bU$ is a constant almost surely. If the random variable V is not a constant almost surely, then $E(V^2) > 0$ and the inequality (4.1) can be written in the form

$$E(U^2) \geq \frac{[E(UV)]^2}{E(V^2)}. \tag{4.2}$$

This inequality is invariant under non-zero constant multiplication but is not invariant under translation. If U is replaced by cU and V is replaced by dV where c and d are non-zero constants, the inequality (4.2) does not change. However, if U is replaced by $U - a$ and V is replaced by $V - b$ where a and b are constants, then the inequality (4.2) changes to the inequality

$$E[(U-a)^2 \geq \frac{[E[(U-a)(V-b)]^2}{E[(V-b)^2)]} \tag{4.3}$$

which implies that

$$E(u^2) \geq (E[U])^2 - (E(U)-a)^2 + \frac{[E[(U-a)(V-b)]^2}{E[(V-b)^2)]}. \tag{4.4}$$

This inequality is different from the inequality stated in (4.2). The inequality (4.4) gives a family of lower bounds for $E(U^2)$ and it can be shown that the lower bound is optimum when $a = E(U)$ and $b = E(V)$ leading to the inequality

$$E(U^2) \geq (E(U))^2 + \frac{[\mathrm{Cov}(U,V)]^2}{\mathrm{Var}(V)} \tag{4.5}$$

or equivalently,

$$\mathrm{Var}(U) \geq \frac{[\mathrm{Cov}(U,V)]^2}{\mathrm{Var}(V)}. \tag{4.6}$$

This inequality is sharp [5,6].

Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ be two sequences of real numbers. Cauchy–Schwartz inequality implies that

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right)$$

with equality occuring if and only if there exists a constant c such that $x_i = cy_i, i = 1, \ldots, n$. Walker [36] improved this inequality. Let

$$\bar{x} = n^{-1}\sum_{i=1}^{n} x_i, \bar{y} = n^{-1}\sum_{i=1}^{n} y_i$$

and

$$V_x = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2, V_y = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2,$$

Walker's inequality states that

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) - n(|\bar{x}|\sqrt{V_y} - |\bar{y}|\sqrt{V_x})^2 \tag{4.7}$$

which is clearly an improvement over the Cauchy–Schwartz inequality. Probabilistic version of this inequality is that for any two random variables U and V with finite second moment,

$$[E(UV)]^2 \leq E(U^2)E(V^2) - [|E(U)|\sqrt{\mathrm{Var}(V)} - |E(V)|\sqrt{\mathrm{Var}(U)}]^2 \tag{4.8}$$

which is a strict improvement over the inequality (4.1) whenever $E(U) \neq 0$ or $E(V) \neq 0$.

This can be seen from the following example.
Suppose V is a standard normal random variable and U is a random variable with finite mean μ and finite variance σ^2. Then the standard Cauchy–Schwartz inequality given in (4.1) implies that

$$[E(UV)]^2 \leq E(U^2)E(V^2) = E(U^2) = \sigma^2 + \mu^2, \tag{4.9}$$

where, as the improved Cauchy–Schwartz inequality stated in (4.9) implies that

$$\begin{aligned}[E(UV)]^2 \leq {} & E(U^2)E(V^2) - [|E(U)|\sqrt{(}\mathrm{Var}(V)) \\ & -|E(V)|\sqrt{(}\mathrm{Var}(U))^2 = \sigma^2,\end{aligned} \tag{4.10}$$

since $E(V) = 0, E(V^2) = 1, E(U) = \mu$ and $E(U^2) = \mu^2 + \sigma^2$. It is clear that the inequality (4.10) gives a better upper bound than the inequality (4.9).

Ibragimov [15] extended the Cramér–Rao inequality for general loss functions.

5. Generalized Cramér–Rao inequality

Bercher [3] obtained a generalized Cramér–Rao inequality. Suppose $f(x, \theta)$ and $g(x, \theta)$ are two probability density functions of a random vector X taking values in a set $\mathcal{X} \subset R^k$ and $\theta \in \Theta \subset R^n$. Let $\hat{\theta}(X)$ be an estimator for θ based on X. Let

$$B_f(\theta) = E_f[\hat{\theta}(X) - \theta] = \int_{\mathcal{X}} (\hat{\theta}(x) - \theta) f(x, \theta)\mathrm{d}x. \tag{5.1}$$

The function $B_f(\theta)$ is the bias of the estimator $\hat{\theta}(X)$ when the random vector X has the probability density function $f(x, \theta)$. Consider another probability density function $g(x, \theta)$ which can be a weighted version of the function $f(x, \theta)$, that is, $f(x, \theta) = h(x, \theta) f(x, \theta)$ or an escort density function, that is, $g(x, \theta) = [f(x, \theta)]^q$ for some q which can be a tuning parameter with applications in statistical physics. Let $\beta \geq 1$ and suppose the risk of the estimator $\hat{\theta}(X)$ is measured through a power function $||u||^\beta$, $u \in R^d$. Let

$$E_g[||\hat{\theta}(X) - \theta||^\beta] = \int_{\mathcal{X}} ||\hat{\theta}(x) - \theta||^\beta g(x, \theta)\mathrm{d}x. \tag{5.2}$$

Here $||\cdot||$ denotes the Euclidean norm on R^d. Under some standard regularity conditions, Bercher [3] proved that

$$(E_g[||\hat{\theta}(X) - \theta||^\alpha])^{1/\alpha}(I_\beta[f|g; \theta])^{1/\beta} \geq |n + \nabla_\theta B_f(\theta|,$$

where α and β are conjugates, that is, $\alpha^{-1} + \beta^{-1} = 1, \alpha \geq 1$ and $I_\beta f|g; \theta]$ is the (β, g)-Fisher information defined by

$$I_\beta[f|g; \theta] = \int_{\mathcal{X}} ||\frac{\nabla_\theta f(x, \theta)}{g(x, \theta)}||^\beta g(x, \theta)\mathrm{d}x \tag{5.3}$$

is the *generalized Fisher information* of order β contained in the probability density function f with respect to the probability density function g.

Kelbert and Mozgunov [16] derived a generalized Cramér–Rao inequality for weighted covariance matrices. The need for weight function in statistics arises when the observations cannot be considered as equivalent or when the estimation of the parameter is sensitive in a neighbourhood of some value.

In a recent work, Cianchi *et al.* [7] developed an unified approach for establishing a broad class of Cramér–Rao inequalities for the location parameter families, that is, for density functions of the type $f(x,\theta) = f(x-\theta)$. Suppose the function $f(x)$ is differentiable with derivative $f'(x)$ and X is a random variable with probability density function $f(x-\theta)$. Suppose $E(X) = \theta$, that is, the random variable X is an unbiased estimator of the parameter θ. In this case, the Fisher information is

$$E\left[\left(\frac{f'(X)}{f(X)}\right)^2\right]$$

and the Cramér–Rao inequality reduces to

$$E(X^2) \geq \frac{1}{E\left[\left(\frac{f'(X)}{f(X)}\right)^2\right]}. \tag{5.4}$$

It can also be shown that equality occurs in the above inequality if and only if the probability density function f is Gaussian. Cianchi *et al.* [7] obtained a general version of the inequality given above. Suppose $1 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Further, suppose that X is a random variable with the probability density function $f(x)$ which is continuously differentiable such that $E[|X|]^p < \infty$ and $E[|\frac{f'(X)}{f(X)}|^q] < \infty$. Then

$$(E|X|^p)^{1/p} \geq \frac{1}{\left(E\left[|\frac{f'(X)}{f(X)}|^q\right]\right)^{1/q}} \tag{5.5}$$

with equality holding if and only if $f(x) = \alpha \mathrm{e}^{-\beta|x|^p}$ for some positive constants α and β. This inequality is termed as L^p *Cramér–Rao inequality* in Cianchi *et al.* [7]. Inequalities of Weyl-type and information inequalities generalizing the concept of Fisher information are presented in [20,22–24].

6. Cramér–Rao type integral inequality

Suppose $\hat{\theta}(X)$ is an estimator of a scalar parameter $\theta \in \Theta = [a, b]$ based on an observation of a random variable X taking values in a set X with probability density function $f(x,\theta)$. Suppose further the parameter θ has a prior density $\lambda(\theta)$ on the parameter space Θ. It is of interest to measure the risk, called the Bayes risk, associated with the estimator $\hat{\theta}(X)$. Let $E_\theta(.)$ denote the conditional expectation given the parameter θ and $E(.)$ denote the expectation with respect to the joint distribution of the random vector (X,θ). Assume that $f(x,\theta)\lambda(\theta) \to 0$ as $\theta \to a$ and as $\theta \to b$. It can be shown that

$$E[(\hat{\theta}(X) - \theta)^2] \geq \frac{1}{E(I(\theta)) + I(\lambda)}, \tag{6.1}$$

where

$$\begin{aligned} E(I(\theta)) &= E\left(\int_{\mathcal{X}} \left[\frac{\partial f(x,\theta)}{\partial\theta}\right]^2 f(x,\theta)\mathrm{d}x\right) \\ &= \int_\Theta \left[\int_{\mathcal{X}} \left[\frac{\partial f(x,\theta)}{\partial\theta}\right]^2 f(x,\theta)\mathrm{d}x\right] \lambda(\theta)\mathrm{d}\theta \\ &= \int_\Theta \int_{\mathcal{X}} \left[\frac{\partial f(x,\theta)}{\partial\theta}\right]^2 f(x,\theta)\lambda(\theta)\mathrm{d}x\mathrm{d}\theta. \end{aligned} \tag{6.2}$$

and

$$I(\lambda) = \int_{\Theta} \left[\frac{\partial \lambda(\theta)}{\partial \theta} \right]^2 \lambda(\theta) \mathrm{d}\theta. \tag{6.3}$$

The function $I(\lambda)$ is the Fisher information corresponding to the prior density function $\lambda(.)$ and $I(\theta)$ is the conditional Fisher information, given the parameter θ. All the computations given above can be justified by Fubini theorem. The inequality (6.1) is termed as the Cramér–Rao integral inequality or some times known as Bayesian Cramér–Rao inequality or Bayesian Cramér–Rao lower bound. It is again a consequence of the Cauchy–Schwartz inequality. This bound has been found to be a useful tool for obtaining the minimax risk in nonparametric statistics and found application in the subject of nonlinear filtering and signal processing (cf. [35]). We will give a sketch of the proof. Note that

$$\int_a^b \frac{\partial f(x,\theta)\lambda(\theta)}{\mathrm{d}\theta} \mathrm{d}\theta = [f(x,\theta)\lambda(\theta)]_a^b = 0, \tag{6.4}$$

since $f(x,\theta)\lambda(\theta) \to 0$ as $\theta \to a$ and as $\theta \to b$, by assumption. Therefore,

$$\int_a^b \hat{\theta}(x) \frac{\partial f(x,\theta)\lambda(\theta)}{\mathrm{d}\theta} \mathrm{d}\theta = 0. \tag{6.5}$$

Furthermore, integrating by parts, it follows that

$$\begin{aligned} \int_a^b \theta \frac{\partial f(x,\theta)\lambda(\theta)}{\mathrm{d}\theta} \mathrm{d}\theta &= [\theta f(x,\theta)\lambda(\theta)]_a^b - \int_a^b f(x,\theta)\lambda(\theta)\mathrm{d}\theta \\ &= -\int_a^b f(x,\theta)\lambda(\theta)\mathrm{d}\theta. \end{aligned} \tag{6.6}$$

Subtracting (6.6) from (6.5), we get that

$$\int_a^b (\hat{\theta}(x) - \theta)) \frac{\partial f(x,\theta)\lambda(\theta)}{\mathrm{d}\theta} \mathrm{d}\theta = \int_a^b f(x,\theta)\lambda(\theta)\mathrm{d}\theta.$$

Integrating over x on both sides of the above equation, we get that

$$\int_{\mathcal{X}} \int_a^b (\hat{\theta}(x) - \theta) \frac{\partial f(x,\theta)\lambda(\theta)}{\mathrm{d}\theta} \mathrm{d}x = \int_{\mathcal{X}} \int_a^b f(x,\theta)\lambda(\theta)\mathrm{d}\theta\mathrm{d}x = 1$$

or equivalently,

$$\int_{\mathcal{X}} \int_a^b (\hat{\theta}(x) - \theta) \frac{\partial \log(f(x,\theta)\lambda(\theta))}{\mathrm{d}\theta} f(x,\theta)\lambda(\theta)\mathrm{d}x\mathrm{d}\theta = 1.$$

Applying the Cauchy–Schwartz inequality, it follows that

$$\left[\int_{\mathcal{X}} \int_a^b (\hat{\theta}(x) - \theta)^2 f(x,\theta)\lambda(\theta)\mathrm{d}x\mathrm{d}\theta\right] \left[\int_{\mathcal{X}} \int_a^b \left(\frac{\partial \log(f(x,\theta)\lambda(\theta))}{\mathrm{d}\theta})^2 f(x,\theta\right) \lambda(\theta)\mathrm{d}x\mathrm{d}\theta\right] \geq 1$$

or equivalently,

$$E[(\hat{\theta}(X) - \theta)^2] \geq \frac{1}{[\int_{\mathcal{X}} \int_a^b (\frac{\partial \log(f(x,\theta)\lambda(\theta))}{d\theta})^2 f(x,\theta)\lambda(\theta)\mathrm{d}x\mathrm{d}\theta]}. \tag{6.7}$$

A bit of algebra shows that

$$\int_{\mathcal{X}}\int_a^b \left(\frac{\partial \log(f(x,\theta)\lambda(\theta))}{\mathrm{d}\theta}\right)^2 f(x,\theta)\lambda(\theta)\mathrm{d}x\mathrm{d}\theta = E(I(\theta)) + I(\lambda), \qquad (6.8)$$

leading to the the Cramér–Rao type integral inequality

$$E[(\hat{\theta}(X)-\theta)^2] \geq \frac{1}{E(I(\theta))+I(\lambda)} \qquad (6.9)$$

which gives a lower bound on the Bayes risk when the parameter θ has a prior density $\lambda(\theta)$.

Improved Cramér–Rao type integral inequalities based on the improved Cauchy–Schwartz inequality due to Walker [36] are given in [27–29]. In a voluminous work, Van Trees and Bell [35] gave a survey of Bayesian lower bound for parameter estimation and nonlinear filter/tracking and edited a volume containing selected papers dealing with Cramér–Rao bounds, global Bayesian bounds, hybrid Bayesian bounds, constrained Cramér–Rao bounds and their applications to nonlinear dynamic systems. Cramér–Rao type integral inequalities for multidimensiomal parameter, for censored data, for general loss functions and for random elements taking values in a Banach space are obtained in [18–20,23–25].

Acknowledgements

The author would like to dedicate this paper to his teacher, Professor C R Rao during his centenary year. This work was supported under the scheme “INSA Senior Scientist” at the C R Rao Advanced Institute of Mathematics, Statistics and Computer Science, Hyderabad, India.

References

[1] Barankin E W, Locally best unbiased estimates, *Ann. Math. Statist.* **20** (1949) 477–501
[2] Bera A, ET interview with C.R. Rao, *Econometric Theory* **19** (2003) 329–398
[3] Bercher J F, On multidimensional generalized Cramér–Rao inequalities, uncertainty relations and characterizations of generalized q-Gaussian distributions, *J. Phys. A: Math. Theor.* **46** (2013) 095303, 18 pp
[4] Bhattacharya A, On some analogues of the amount of information and their use in statistical estimation, *Sankhya* **8** (1946) 1–14, 201–218, 315–328
[5] Blyth C, Necessary and sufficient conditions for inequalities of Cramér–Rao type, *Ann. Statist.* **2** (1974) 464–473
[6] Blyth C and Roberts D M, On inequalities of Cramér–Rao type and admissibility proofs, Proc. Sixth Berkeley Symp. Math. Statist. Prob. (1972) (University of California Press) vol. 1, pp. 17–30
[7] Cianchi A, Lutwak E, Yang D and Zhang G, A unified approach to Cramér–Rao type inequalities, *IEEE Trans. Inf. Theory* **60** (2014) 643–650
[8] Cramér H, Mathematical Methods of Statistics (1946) (New Jersey: Princeton University Press)
[9] Darmois G, Sur les limites de le dispersion de certaines estimations, *Rev. Int. Inst. Statist.* 13 (1945) 9–15
[10] Degroot M H (1987) A conversation with C. R. Rao, *Statist. Sci.* **2** (1987) 53–67
[11] Dembo A, Cover T M and Thomas J A, Information-theoretic inequalities, *IEEE Trans. Inf. Theory* **37** (1991) 1501–1518

[12] Fabian V and Hannan J, On the Cramér–Rao inequality, *Ann. Statist.* **5** (1977) 197–205
[13] Frechet M, Sur le extension de certaines evaluations statistiques au cas de petitis echantillons, *Rev. Int. Inst. Statist.* **11** (1943) 182–205
[14] Ghosh J K, Mitra S K and Parthasarathy K R Glimpses of India's Statistical Heritage (1992) (New Delhi: Wiley Eastern)
[15] Ibragimov I A, A generalization of the Cramér–Rao inequality, *J. Math. Sci. (New York)* **93** (1999) 349–353
[16] Kelbert M and Mozgunov P, Generalization of Cramér–Rao and Bhattacharya inequalities for the weighted covariance matrix, *Math. Commun.* **22** (2017) 25–40
[17] Parthasarathy K R, On the Philosophy of Cramér–Rao–Bhattacharya Inequalities in Quantum Statistics, Technical Report (2009) (Delhi: Indian Statistical Institute)
[18] Prakasa Rao B L S, On Cramér–Rao type integral inequalities, *Cal. Stat. Assoc. Bull.*, **40** (1991) 183–205, reprinted in "Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking", edited by H Van Trees and K L Bell (2007) (New York: IEEE Press)
[19] Prakasa Rao B L S, Cramér–Rao type integral inequalities for estimators of functions of multidimensional parameter, *Sankhya*, **54** (1992) 53–73
[20] Prakasa Rao B L S, Remarks on Cramér–Rao type integral inequalities for randomly censored data, in: Analysis of Censored Data, IMS Lecture Notes No. 27, (eds.) by H L Koul and J V Deshpande (1996) pp. 163–176
[21] Prakasa Rao B L S, Inequalities of Weyl-type and information inequalities, *Indian J. Pure Appl. Math.* **29** (1998) 681–694
[22] Prakasa Rao B L S, Cramér–Rao type integral inequalities in Banach spaces, in: Perspectives in Statistical Sciences, (eds.) A K Basu, J K Ghosh, P K Sen and B K Sinha (2000a) (Oxford: Oxford University Press) pp. 245–260
[23] Prakasa Rao B L S, Some covariance identities and inequalities and their applications, *Proc. Indian National Sci. Acad. – A: Physical Sci.* **66A** (2000b) 537–543
[24] Prakasa Rao B L S, Sequential Cramér–Rao type integral inequlities, Special issue on Statistical Methods in Honour of Prof. C. R. Rao, Proceedings Andhra Pradesh Akademi of Sciences (2000c) vol. 5, pp. 33–38
[25] Prakasa Rao B L S, Cramér–Rao type integral inequalities for general loss functions, *Test*, **10** (2001) 105–120
[26] Prakasa Rao B L S, C. R. Rao: A Life in Statistics, in the series "Living Legends in Indian Science", *Curr. Sci.* **107** (2014) 895–901
[27] Prakasa Rao B L S, Improved Cramér–Rao type integral inequalities or Bayesian Cramér–Rao bounds, *J. Indian Soc. Probab. Stat.* **19** (2018a) 1–7
[28] Prakasa Rao B L S, Improved sequential Cramér–Rao type integral inequalities, *Sequential Anal.* **37** (2018b) 59–68
[29] Prakasa Rao B L S, Improved Cramér–Rao type integral inequality for randomly censored data, *J. Iranian Stat. Soc.* **17** (2018c) 1–12
[30] Rajarama Bhat B, Wise decisions under uncertainty, *Resonance* **18** (2013) 1127–1132
[31] Rao C R, Information and the accuracy attainable in the estimation of statistical parameters, *Bull. Cal. Math. Soc.* **37** (1945) 81–89
[32] Richards D, Maximum likelihood estimation and Cramér–Rao inequality, Fourth IIA-Penn State Astrostatistics School (2013)
[33] Roy Frieden B, Physics from Fisher Information (1998) (Cambridge: Cambridge University Press)
[34] Stam A J, Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Inf. Control* **2** (1959) 101–112
[35] Van Trees H and Bell K L, Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking (2007) (New York: IEEE Press)
[36] Walker S, A self-improvement to the Cauchy–Schwartz inequality, *Statist. Probab. Lett.* **122** (2017) 86–89