Annals of the Institute of Statistical Mathematics (2020) 72:1237–1256
https://doi.org/10.1007/s10463-019-00725-3

Some information inequalities for statistical inference

K. V. Harsha[1,2] · Alladi Subramanyam[1]

Received: 6 March 2018 / Revised: 15 April 2019 / Published online: 1 August 2019

Abstract
In this paper, we first describe the generalized notion of Cramer–Rao lower bound obtained by Naudts (J Inequal Pure Appl Math 5(4), Article 102, 2004) using two families of probability density functions: the original model and an escort model. We reinterpret the results in Naudts (2004) from a statistical point of view and obtain some interesting examples in which this bound is attained. Further, we obtain information inequalities which generalize the classical Bhattacharyya bounds in both regular and non-regular cases.

Keywords Information inequality · Generalized Cramer–Rao bound · Escort probability distribution · Generalized Bhattacharyya bounds · Deformed exponential family

1 Introduction

For every unbiased estimator T, an inequality of the type

$$\mathrm{Var}_\theta(T) \geq d(\theta) \tag{1}$$

for every θ in the parameter space Θ is called an information inequality and it plays an important role in parameter estimation. The lower bound provided by the family of functions $\{d(\theta) \mid \theta \in \Theta\}$ is said to be sharp if there exists an unbiased estimator T such that

$$\mathrm{Var}_\theta(T) = d(\theta), \quad \forall\, \theta \in \Theta. \tag{2}$$

✉ K. V. Harsha
harshakv@math.iitb.ac.in; kv.harsha1989@gmail.com

Alladi Subramanyam
as@math.iitb.ac.in

1 Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

2 Present Address: Madrid, Spain

The early works of Cramer (1946) and Rao (1945) introduced the Cramer–Rao inequality for regular density functions. For the non-regular density functions, Hammersley (1950) and Chapman and Robbins (1951) introduced an inequality which came to be known as Hammersley–Chapman–Robbins inequality, while Fraser and Guttman (1952) obtained the Bhattacharyya bounds without regularity conditions. See also Lehmann and Casella (1998), page 129 and references therein. Later Vincze (1979) and Khatri (1980) introduced information inequalities by imposing the regularity assumptions on a prior distribution rather than on the model. These ideas have been further investigated by Kshirsagar (2000), Koike (2002) and Qin and Nayak (2008).

Recently in statistical physics, a generalized notion of the Fisher information and a corresponding Cramer–Rao lower bound are introduced by Naudts (2004) using two families of probability density functions: the original model and an escort model. Further, he showed that in the case of a deformed exponential family of probability density functions, there exist an escort family and an estimator whose variance attains the bound. Also from an information geometric point of view, he obtained a dually flat structure of the deformed exponential family.

In this article, concentrating on the statistical aspects of the Naudts's paper we define several information inequalities which generalize the classical Hammersley–Chapman–Robbins bound and Bhattacharyya bounds in both regular and non-regular cases. This is done by imposing the regularity conditions on the escort model rather than on the original model.

In Sect. 2, some preliminary results are stated. Section 3 describes the generalized Cramer–Rao lower bound obtained by Naudts (2004) reinterpreted from a statistical point of view and applied to many examples. In Sect. 4, we obtain a generalized notion of Bhattacharyya bounds in both regular and non-regular cases. We conclude with discussions in Sect. 5.

2 Preliminaries

Let X be a random vector taking values in $A \subseteq \mathbb{R}^n$, and for $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)^\mathsf{T} \in \Theta \subseteq \mathbb{R}^p$, let $P_{f_{\underset{\sim}{\theta}}}$ denote its probability measure. Assume that $P_{f_{\underset{\sim}{\theta}}}$ has density $f(x, \underset{\sim}{\theta})$ with respect to some σ-finite measure μ. To estimate a real-valued function φ of $\underset{\sim}{\theta}$, define a class of estimators as

$$\mathcal{C}_\varphi = \{\delta(X) \mid E_{f_{\underset{\sim}{\theta}}}(\delta(X)) = \varphi(\underset{\sim}{\theta});\ E_{f_{\underset{\sim}{\theta}}}(\delta^2) < \infty,\ \forall\, \underset{\sim}{\theta} \in \Theta\}. \tag{3}$$

Define

$$\mathcal{U}_f = \{U(X) \mid E_{f_{\underset{\sim}{\theta}}}(U) = 0;\ E_{f_{\underset{\sim}{\theta}}}(U^2) < \infty,\ \forall\, \underset{\sim}{\theta} \in \Theta\}. \tag{4}$$

Note that for any $T_1, T_2 \in \mathcal{C}_\varphi$, $T_1 - T_2 \in \mathcal{U}_f$.

Consider a class of functions

$$\Psi = \{S(x, \underset{\sim}{\theta}) \mid E_{f_{\underset{\sim}{\theta}}}(S) = 0;\ 0 < E_{f_{\underset{\sim}{\theta}}}(S^2) < \infty;\ \mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(U, S) = 0, \\ \forall\, U \in \mathcal{U}_f,\quad \forall\, \underset{\sim}{\theta} \in \Theta\}. \tag{5}$$

For a fixed $S \in \Psi$ and for any $T \in \mathcal{C}_\varphi$, $\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T, S)$ depends on T only through $\varphi(\underset{\sim}{\theta})$. That is for any estimators $T_1, T_2 \in \mathcal{C}_\varphi$,

$$\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T_1, S) = \mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T_2, S) \quad \text{since} \quad T_1 - T_2 \in \mathcal{U}_f. \tag{6}$$

Let $S_1(x, \underset{\sim}{\theta}), \ldots, S_m(x, \underset{\sim}{\theta}) \in \Psi$. For each $\alpha = (\alpha_1, \ldots, \alpha_m)^\mathsf{T} \in \mathbb{R}^m$, define

$$\psi(x, \underset{\sim}{\theta}) = \sum_{i=1}^{m} \alpha_i S_i(x, \underset{\sim}{\theta}). \tag{7}$$

Clearly $\psi \in \Psi$. Although ψ depends on α, we avoid the index α for the convenience of writing.

Since ψ, $S_1(x, \underset{\sim}{\theta}), \ldots, S_m(x, \underset{\sim}{\theta}) \in \Psi$, from (6), for all $T \in \mathcal{C}_\varphi$, let

$$\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T, \psi) = \gamma(\underset{\sim}{\theta}) \text{ and } \mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T, S_i) = E_{f_{\underset{\sim}{\theta}}}(T S_i) = \lambda_i(\underset{\sim}{\theta});\quad i = 1, \ldots, m \tag{8}$$

where γ, λ_i are real-valued functions of $\underset{\sim}{\theta}$.

Therefore, $\forall\, T \in \mathcal{C}_\varphi$, the Cauchy–Schwarz inequality

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(T(x)) \geq \frac{(\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T, \psi))^2}{\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(\psi)} = \frac{(\gamma(\underset{\sim}{\theta}))^2}{\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(\psi)} \tag{9}$$

gives a lower bound for the variance of all unbiased estimators of $\varphi(\underset{\sim}{\theta})$.

Now consider

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(\psi) = \mathrm{Var}_{f_{\underset{\sim}{\theta}}}\left(\sum_{i=1}^{m} \alpha_i S_i\right) = \alpha^\mathsf{T} \Sigma \alpha \tag{10}$$

$$(\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T, \psi))^2 = \left(\sum_{i=1}^{m} \alpha_i \lambda_i(\underset{\sim}{\theta})\right)^2 = \alpha^\mathsf{T} M M^\mathsf{T} \alpha \tag{11}$$

where $\Sigma = (\Sigma_{ij}) = (\mathrm{Cov}_f(S_i, S_j))$ is the covariance matrix of $S = (S_1, \ldots, S_m)^\mathsf{T}$ and $M = (\lambda_1(\underset{\sim}{\theta}), \ldots, \lambda_m(\underset{\sim}{\theta}))^\mathsf{T}$.

Here, both M and Σ depend on $\underset{\sim}{\theta}$. But for the convenience of writing, we suppress the index $\underset{\sim}{\theta}$.

Equation (9) becomes

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(T(x)) \geq \frac{\alpha^\intercal M M^\intercal \alpha}{\alpha^\intercal \Sigma \alpha} \quad \forall\, \alpha \in \mathbb{R}^m \tag{12}$$

which implies

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(T(x)) \geq \sup_{\alpha} \frac{\alpha^\intercal M M^\intercal \alpha}{\alpha^\intercal \Sigma \alpha} = M^\intercal \Sigma^{-1} M \tag{13}$$

where Σ^{-1} is the inverse of the covariance matrix Σ.

For later use, we state the following well known theorem (refer Lehmann and Casella 1998) as

Proposition 1 (Information inequality) *Let X be a random vector with probability density function (pdf) $f(x, \underset{\sim}{\theta})$, where $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)^\intercal \in \Theta \subseteq \mathbb{R}^p$. Consider an estimator $T(X) \in \mathcal{C}_\varphi$, $S_1(x, \underset{\sim}{\theta}), \ldots, S_m(x, \underset{\sim}{\theta}) \in \Psi$ and the functions $\lambda_i : \Theta \to \mathbb{R}$; $i = 1, \ldots, m$ with*

$$E_{f_{\underset{\sim}{\theta}}}(T S_i) = \lambda_i(\underset{\sim}{\theta}); \quad i = 1, \ldots, m. \tag{14}$$

Then the variance of T satisfies the inequality

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(T(x)) \geq M^\intercal \Sigma^{-1} M \tag{15}$$

where $M = (\lambda_1(\underset{\sim}{\theta}), \ldots, \lambda_m(\underset{\sim}{\theta}))^\intercal$ and Σ^{-1} is the inverse of the covariance matrix $\Sigma = (\Sigma_{ij}) = (\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(S_i, S_j))$. The equality in Eq. (15) holds iff

$$S^\intercal \Sigma^{-1} M = a(\underset{\sim}{\theta})(T(x) - \varphi(\underset{\sim}{\theta})) \tag{16}$$

for some function $a(\underset{\sim}{\theta})$ and, $S = (S_1, \ldots, S_m)^\intercal$.

3 Generalized Cramer–Rao type lower bound

Naudts (2004) introduced a generalized notion of the Fisher information by replacing the original model by an escort model at suitable places. Using this, he obtained a generalized Cramer–Rao lower bound. To study the statistical implications of this generalization, first we reinterpret Naudts's generalized bound as follows.

For each $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)^\intercal \in \Theta$, let $P_{g_{\underset{\sim}{\theta}}}$ be a probability measure with density $g(x, \underset{\sim}{\theta})$ with respect to the σ-finite measure μ of Sect. 2. Define

$$\mathcal{U}_g = \{U(X) \mid E_{g_{\underset{\sim}{\theta}}}(U) = 0, \quad \forall\, \underset{\sim}{\theta} \in \Theta\}. \tag{17}$$

Let us make the following assumptions,

(a) The probability measure $P_{g_{\underset{\sim}{\theta}}}$ is absolutely continuous with respect to the probability measure $P_{f_{\underset{\sim}{\theta}}}$ for all $\underset{\sim}{\theta} \in \Theta$. (18)

(b) $\mathcal{U}_f \subseteq \mathcal{U}_g$. (19)

Remark 1 For a simple sufficient condition to check Assumption (18), consider the following.

For each $\underset{\sim}{\theta} \in \Theta$, if $\text{supp}(g_{\underset{\sim}{\theta}}) := \{x \mid g(x, \underset{\sim}{\theta}) > 0\} \subseteq \text{supp}(f_{\underset{\sim}{\theta}})$, then taking $h_{\underset{\sim}{\theta}} = \frac{g_{\underset{\sim}{\theta}}}{f_{\underset{\sim}{\theta}}}$, one has $P_{g_{\underset{\sim}{\theta}}}(A) = \int_A h_{\underset{\sim}{\theta}} \, dP_{f_{\underset{\sim}{\theta}}}$ and hence $P_{g_{\underset{\sim}{\theta}}}$ is absolutely continuous with respect to $P_{f_{\underset{\sim}{\theta}}}$ for all $\underset{\sim}{\theta} \in \Theta$.

Remark 2 If X is a complete statistic for the family $\{f_{\underset{\sim}{\theta}} \mid \underset{\sim}{\theta} \in \Theta\}$ and if there exists a family $\{g_{\underset{\sim}{\theta}} \mid \underset{\sim}{\theta} \in \Theta\}$ satisfying Assumption (18), then $\mathcal{U}_f \subseteq \mathcal{U}_g$.

To see this, note that since X is complete for the family $\{f_{\underset{\sim}{\theta}} \mid \underset{\sim}{\theta} \in \Theta\}$, if $U = U(X) \in \mathcal{U}_f$, then $P_{f_{\underset{\sim}{\theta}}}(U = 0) = 1 \ \forall \ \underset{\sim}{\theta}$. That is,

$$U = 0, \quad \text{a.e. } P_{f_{\underset{\sim}{\theta}}} \text{ (almost everywhere with respect to } P_{f_{\underset{\sim}{\theta}}}) \ \ \forall \ \underset{\sim}{\theta}. \tag{20}$$

From Assumption (18), it follows that

$$U = 0, \quad \text{a.e. } P_{g_{\underset{\sim}{\theta}}} \ \forall \ \underset{\sim}{\theta} \Rightarrow U \in \mathcal{U}_g. \tag{21}$$

Naudts (2004) defined a generalized Fisher information $N(\underset{\sim}{\theta}) = (N_{ij}(\underset{\sim}{\theta}))$ as

$$N_{ij}(\underset{\sim}{\theta}) = \int \partial_i g(x, \underset{\sim}{\theta}) \partial_j g(x, \underset{\sim}{\theta}) \frac{1}{f(x, \underset{\sim}{\theta})} \mathrm{d}x, \quad \partial_i := \frac{\partial}{\partial \theta_i} \text{ and } \quad i, j = 1, \ldots, p. \tag{22}$$

When $f = g$, $N(\underset{\sim}{\theta})$ reduces to the Fisher information $I(\underset{\sim}{\theta})$.

Theorem 1 *Let X be a random vector with pdf $f(x, \underset{\sim}{\theta})$. Let $g(x, \underset{\sim}{\theta})$ be a pdf satisfying Assumptions (18) and (19) with $E_{g_{\underset{\sim}{\theta}}}[T] = \lambda(\underset{\sim}{\theta})$. Assume that*

(a) *$\partial_i g(x, \underset{\sim}{\theta})$ exists and the function $\lambda(\underset{\sim}{\theta})$ is differentiable with respect to θ_i for all $x \in A$ and $\underset{\sim}{\theta} \in \Theta$, where $i = 1, \ldots, p$* (23)

(b) *$N_{ii}(\underset{\sim}{\theta}) > 0$ and $N(\underset{\sim}{\theta})$ is non-singular* (24)

(c) *Partial derivatives of functions of $\underset{\sim}{\theta}$ expressed as integrals with respect to $g(x, \underset{\sim}{\theta})$ can be obtained by differentiating under the integral sign.* (25)

Then for $T(X) \in \mathcal{C}_{\varphi}$, the variance of T satisfies

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(T(X)) \geq M^{\mathsf{T}} N^{-1}(\underset{\sim}{\theta}) M \tag{26}$$

where $M = (\partial_1 \lambda(\underset{\sim}{\theta}), \ldots, \partial_p \lambda(\underset{\sim}{\theta}))^{\mathsf{T}}$.

Proof From Proposition 1, choose $m = p$ functions S_i as

$$S_i = \frac{\partial_i g(x, \underset{\sim}{\theta})}{f(x, \underset{\sim}{\theta})}, \quad i = 1, \ldots, p. \tag{27}$$

From Assumptions (24) and (25),

$$E_{f_{\underset{\sim}{\theta}}}(S_i) = \int \partial_i g(x, \underset{\sim}{\theta}) \mathrm{d}x = 0; \quad \mathrm{Var}_{f_{\underset{\sim}{\theta}}}(S_i) = N_{ii}(\theta) < \infty. \tag{28}$$

Now from Eq. (19), it follows that for all $U \in \mathcal{U}_f$,

$$\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(U, S_i) = \int U(x) \partial_i g(x, \underset{\sim}{\theta}) \mathrm{d}x = \partial_i \left(\int U(x) g(x, \underset{\sim}{\theta}) \mathrm{d}x \right) = 0. \tag{29}$$

Hence, $S_i \in \Psi$ for $i = 1, \ldots, p$.

From Assumption (25), it follows that

$$E_{f_{\underset{\sim}{\theta}}}(T S_i) = \int T \; \partial_i g(x, \underset{\sim}{\theta}) \mathrm{d}x = \partial_i \left(\int T \; g(x, \underset{\sim}{\theta}) \mathrm{d}x \right) = \partial_i \lambda(\underset{\sim}{\theta}) \tag{30}$$

where $\lambda(\underset{\sim}{\theta}) = E_{g_{\underset{\sim}{\theta}}}[T]$ and $i = 1, \ldots, p$.

Applying Proposition 1, the bound in Eq. (26) is obtained.

The fact that $\mathcal{U}_f \subseteq \mathcal{U}_g$ ensures that the bound is same for all unbiased estimators T of $\varphi(\underset{\sim}{\theta})$. □

Now we give some interesting examples in which the Naudts's generalized Cramer–Rao bound is sharp.

Example 1 Suppose $Y_1, \ldots, Y_n$ are independent uniform random variables in $[0, \theta]$, where $\theta > 0$. Then

$$X = \max\{Y_1, \ldots, Y_n\} \sim f(x, \theta) = \frac{n x^{n-1}}{\theta^n}, \quad 0 < x < \theta. \tag{31}$$

Consider an unbiased estimator $T(X) = \frac{(n+k)X^k}{n}$ for θ^k, where k is a positive integer. Then

$$\mathrm{Var}_{f_\theta}(T) = \frac{k^2\theta^{2k}}{n(n+2k)}. \tag{32}$$

Define a pdf $g(x, \theta)$ as

$$g(x,\theta) = \frac{n(n+k)\left(1 - \frac{x^k}{\theta^k}\right)x^{n-1}}{k\theta^n}, \quad 0 < x < \theta. \tag{33}$$

Here, X is a complete statistic for the family $\{f_\theta \mid \theta > 0\}$ (see Example 6.23 (ii), page 42, Lehmann and Casella 1998). Then using Remarks 1 and 2, it follows that $g(x, \theta)$ satisfies Assumptions (18) and (19).

Then the bound in Eq. (26) is obtained as

$$\frac{(\lambda'(\theta))^2}{N(\theta)} = \frac{k^2\theta^{2k}}{n(n+2k)} = \mathrm{Var}_{f_\theta}(T). \tag{34}$$

Thus, the estimator $T(X)$ is an unbiased estimator of θ^k whose variance attains the bound in Eq. (26). When $n = 1, k = 1$, this example reduces to Example 1 given in Naudts (2004). In this case, $\mathrm{Var}_{f_\theta}(T)$ does not attain the Hammersley–Chapman–Robbins lower bound.

Example 2 Suppose $Y_1, \ldots, Y_n$ are independent random variables,

$$Y_1, \ldots, Y_n \sim \exp(-(y-\theta)), \quad y \geq \theta, \theta > 0. \tag{35}$$

Then the random variable $X = \min\{Y_1, \ldots, Y_n\}$ has a pdf

$$f(x,\theta) = n\exp(-n(x-\theta)), \quad x \geq \theta. \tag{36}$$

Now consider an unbiased estimator $T(X) = X - \frac{1}{n}$ of θ. Then

$$\mathrm{Var}_{f_\theta}(T) = \frac{1}{n^2}. \tag{37}$$

The pdf $g(x, \theta)$ which optimizes the bound in Eq. (26) is

$$g(x,\theta) = n^2(x-\theta)\exp(-n(x-\theta)), \quad x \geq \theta. \tag{38}$$

Here, X is a complete statistic for the family $\{f_\theta \mid \theta > 0\}$see Example 6.23 (iii), page 43, Lehmann and Casella 1998). Then using Remarks 1 and 2, it follows that $g(x, \theta)$

satisfies Assumptions (18) and (19). Note that $E_{g_\theta}[T] = \lambda(\theta) = \frac{1}{n} + \theta$ and the bound in Eq. (26) is obtained as

$$\frac{(\lambda'(\theta))^2}{N(\theta)} = \frac{1}{n^2} = \mathrm{Var}_{f_\theta}(T). \tag{39}$$

Example 3 **Location family**
Let $f(x)$ and $g(x)$ be two density functions on $x \in D' \subseteq \mathbb{R}$ satisfying Assumptions (18) and (19). Now let X be a random variable with density function $f(x,\theta) = f(x-\theta)$, $\theta \in \mathbb{R}$ and $x \in D \subseteq \mathbb{R}$. Let $g(x,\theta) = g(x-\theta)$. Let $T(X)$ be an unbiased estimator for $\varphi(\theta)$. Let $E_g(T) = \lambda(\theta)$. Then from Eq. (16), the optimality condition for the bound in Eq. (26) is given by

$$\frac{\partial_\theta g(x,\theta)}{f(x,\theta)} = a(\theta)(T(x) - \varphi(\theta)) \tag{40}$$

for some function $a(\theta)$. In this case

$$\partial_\theta g(x,\theta) = -g'(x-\theta) \tag{41}$$

where g' denotes the derivative of $g(x)$ with respect to x. Then Eq. (40) becomes

$$g'(x-\theta) = a(\theta)(\varphi(\theta) - T(x))f(x,\theta). \tag{42}$$

Let $\theta = 0$ and $x_0 \in D'$, then

$$g(x) = a(0)\left(\varphi(0)\int_{x_0}^{x} f(x)\mathrm{d}x - \int_{x_0}^{x} T(x)f(x)\mathrm{d}x\right) \tag{43}$$

$$= a(0)h(x) \tag{44}$$

where

$$h(x) = \varphi(0)\int_{x_0}^{x} f(x)\mathrm{d}x - \int_{x_0}^{x} T(x)f(x)\mathrm{d}x < \infty \tag{45}$$

can be computed since $f(x), T(x), \varphi(0)$ are given.

Now $a(0)$ can be solved from the normalization condition $\int_{D'} g(x)\mathrm{d}x = 1$ as

$$a(0) = \frac{1}{\int_{D'} h(x)\mathrm{d}x} \quad \text{if} \quad \int_{D'} h(x)\mathrm{d}x < \infty. \tag{46}$$

Thus, the optimizing family $g(x,\theta) = g(x-\theta)$ is obtained.

Example 4 **Scale family**
Let $f(x)$ and $g(x)$ be two density functions on $x \in D' \subseteq \mathbb{R}$ satisfying Assumptions (18) and (19). Now let

$$X \sim f(x,\theta) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right) \quad x \in D \subseteq \mathbb{R}, \theta > 0 \tag{47}$$

and

$$g(x,\theta)=\frac{1}{\theta}g\left(\frac{x}{\theta}\right). \tag{48}$$

Let $T(X)$ be an unbiased estimator for $\varphi(\theta)$. Let $E_g(T)=\lambda(\theta)$. Then from Eq. (16),

$$\frac{\partial_\theta g(x,\theta)}{f(x,\theta)}=a(\theta)(T(x)-\varphi(\theta)) \tag{49}$$

for some function $a(\theta)$.

$$\frac{-x}{\theta^3}g'(x/\theta)-\frac{1}{\theta^2}g(x/\theta)=a(\theta)(T(x)-\varphi(\theta))f(x,\theta) \tag{50}$$

where g' denotes the derivative of function $g(x)$ with respect to x.

Let $\theta=1$. Then we have

$$xg'(x)+g(x)=a(1)(\varphi(1)-T(x))f(x). \tag{51}$$

Let $x_0\in D'$. Integrating the above equation from x_0 to x, we get

$$xg(x)-x_0g(x_0)=a(1)\int_{x_0}^{x}(\varphi(1)-T(x))f(x)\,\mathrm{d}x \tag{52}$$

$$=a(1)(h(x)-h(x_0)) \tag{53}$$

where

$$h(x)-h(x_0)=\int_{x_0}^{x}(\varphi(1)-T(x))f(x)\,\mathrm{d}x. \tag{54}$$

Thus, we get

$$xg(x)=a(1)h(x)\Rightarrow g(x)=a(1)k(x) \tag{55}$$

for some function $k(x)$.

Now $a(1)$ can be solved from the normalization condition of the function $\int_{D'}g(x)\mathrm{d}x=1$ as

$$a(1)=\frac{1}{\int_{D'}k(x)\mathrm{d}x}\quad\text{if}\quad\int_{D'}k(x)\mathrm{d}x<\infty. \tag{56}$$

Thus, the optimizing family $g(x,\theta)=\frac{1}{\theta}g(\frac{x}{\theta})$ is obtained.

Example 5 Let X be a random variable distributed according to the Gamma distribution $f(x,\theta)$ with a scale parameter $\theta>0$ and a known shape parameter $\alpha>0$ given by

$$f(x,\theta)=\frac{1}{\Gamma(\alpha)}\frac{x^{\alpha-1}e^{-x/\theta}}{\theta^\alpha},\quad x>0. \tag{57}$$

Here, $\{f_\theta \mid \theta > 0\}$ is a one-dimensional exponential family with a canonical statistic X. Hence, X is a complete statistic for the family $\{f_\theta \mid \theta > 0\}$ (see Theorem 6.22, page 42, Lehmann and Casella 1998 for the details). Let $T(X) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+k)} X^k$, where k is an integer such that $k \neq 0$ and $2k + \alpha > 0$. Then T is an unbiased estimator of θ^k with $E_{f_\theta}(T^2) < \infty$.

$$\mathrm{Var}_{f_\theta}(T) = \left[\frac{\Gamma(\alpha)\Gamma(2k+\alpha)}{(\Gamma(\alpha+k))^2} - 1\right]\theta^{2k}. \tag{58}$$

Consider a pdf $g(x, \theta)$ such that $\mathrm{Var}_{f_\theta}(T)$ attains the bound in Eq. (26) as follows.

For $k > 0$,

$$g(x,\theta) = \frac{1}{c}\frac{e^{-x/\theta}}{\theta}\left[\sum_{i=0}^{k-1} s_i \left(\frac{x}{\theta}\right)^{\alpha+k-(i+2)}\right], \quad c = \sum_{i=0}^{k-1} s_i \Gamma(\alpha + k - (i+1)) \tag{59}$$

where $x > 0$, $s_i = \prod_{j=1}^{i}(\alpha + k - j)$; $i = 1, \ldots, k-1$ and $s_0 = 1$.

For $k < 0$ and $k \neq -1$,

$$g(x,\theta) = \frac{1}{c}\frac{e^{-x/\theta}}{\theta}\left[\sum_{i=1}^{k_1} s_i \left(\frac{x}{\theta}\right)^{\alpha-(i+1)}\right], \quad c = \sum_{i=0}^{k_1} s_i \Gamma(\alpha - i) \tag{60}$$

where $x > 0$, $k_1 = -k$, $s_i = \prod_{j=1}^{i-1}(\alpha - j)$; $i = 2, \ldots, k_1$ and $s_0 = 1$.

For $k = -1$,

$$g(x,\theta) = \frac{1}{\Gamma(\alpha-1)}\frac{x^{\alpha-2}e^{-x/\theta}}{\theta^{\alpha-1}}, \quad x > 0. \tag{61}$$

This is an interesting special case as the variance of $T = 1/X$ does not attain the Bhattacharyya bounds of any order, while it attains the bound in Eq. (26). Note that in all cases using Remarks 1 and 2, it follows that $g(x, \theta)$ satisfies Assumptions (18) and (19).

Example 6 Let X be a random variable distributed according to the Normal distribution $\mathcal{N}(0, \theta^2)$ given by

$$f(x,\theta) = \frac{1}{\sqrt{2\pi}\theta} e^{\frac{-x^2}{2\theta^2}}, \quad x \in \mathbb{R} \quad \text{and} \quad \theta > 0. \tag{62}$$

Consider an unbiased estimator $T(X) = \frac{X^4}{3}$ for θ^4. Then $\mathrm{Var}_{f_\theta}(T) = \frac{32\theta^8}{3}$. Consider a pdf

$$g(x,\theta) = \frac{1}{\sqrt{2\pi}\theta}\left(\frac{3}{4} + \frac{x^2}{4\theta^2}\right) e^{\frac{-x^2}{2\theta^2}}, \quad x \in \mathbb{R}. \tag{63}$$

In this example, X is not complete for the family $\{f_\theta \mid \theta > 0\}$. However, Assumptions (18) and (19) are satisfied as shown in Lemma 1 of Appendix.

We have

$$N(\theta) = \frac{6}{\theta^2} \quad \text{and} \quad \lambda(\theta) = E_{g_\theta}(T) = 2\theta^4. \tag{64}$$

Thus, the bound in Eq. (26) is obtained as

$$\frac{(\lambda'(\theta))^2}{N(\theta)} = \frac{32\theta^8}{3} = \mathrm{Var}_{f_\theta}(T). \tag{65}$$

Thus, the variance of T attains Naudts's bound with optimizing family $g(x, \theta)$. Note that $f(x, \theta)$ belongs to exponential family and $T(X)$ is a second-degree polynomial in the canonical statistic X^2. Hence, its variance attains the Bhattacharyya bound of order 2. Thus, the 'first-order' bound obtained using g is equal to the second-order Bhattacharyya bound.

Example 7 **Poisson distribution**
Let $X_1, \ldots, X_n$ are i.i.d random variables from Poisson distribution

$$f(x, \theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, \ldots \quad \text{and} \quad \theta > 0. \tag{66}$$

Consider the joint pdf

$$f(x_1, \ldots, x_n, \theta) = \frac{\theta^{n\bar{x}} e^{-n\theta}}{x_1! \ldots x_n!}, \quad \text{where} \quad \bar{x} = \frac{x_1 + \ldots + x_n}{n}. \tag{67}$$

Here, $\bar{X}$ is a complete statistic for the family $\{f_\theta \mid \theta > 0\}$ (see Theorem 6.22, page 42, Lehmann and Casella 1998). Consider an unbiased estimator $T(X) = \bar{X}(\bar{X} - \frac{1}{n})$ for θ^2. The variance of T attains the bound in Eq. (26) if we choose the pdf

$$g(x_1, \ldots, x_n, \theta) = \frac{1}{2} \frac{\theta^{n\bar{x}} e^{-n\theta}}{x_1! \ldots x_n!} + \frac{\bar{x}}{2} \frac{\theta^{n\bar{x}-1} e^{-n\theta}}{x_1! \ldots x_n!}. \tag{68}$$

Using Remarks 1 and 2, it follows that $g(x_1, \ldots, x_n, \theta)$ satisfies Assumptions (18) and (19). $\mathrm{Var}_{f_\theta}(T)$ attains the Bhattacharyya bound of order 2, while it attains 'first-order' Naudts's bound.

Example 8 Let $X_1, \ldots, X_n$ are i.i.d uniform random variables in $[0, \theta]$, where $\theta > 0$. Then the joint pdf is

$$f(x_1, \ldots, x_n, \theta) = \frac{1}{\theta^n} \Pi_{i=1}^n \mathbf{1}_{\{0 \le x_i \le \theta\}} \tag{69}$$

where $\mathbf{1}$ denotes the indicator function.

The statistic $T = \max\{X_1, \ldots, X_n\}$ is a sufficient statistic with $E_{f_\theta}(T) = \frac{n}{n+1}\theta$, and $\mathrm{Var}_{f_\theta}(T)$ attains the bound in Eq. (26) if we choose the pdf

$$g(x_1, \ldots, x_n, \theta) = \frac{n+1}{\theta^n}\left(1 - \frac{t}{\theta}\right); \quad 0 \le t \le \theta \tag{70}$$

where $t = \max\{x_1, \ldots, x_n\}$.

The pdf $g(x_1, \ldots, x_x, \theta)$ can be written as

$$g(x_1, \ldots, x_x, \theta) = Z\left(\frac{n+1}{\theta^n} - \frac{(n+1)t}{\theta^{n+1}} - 1\right) \tag{71}$$

where the Z is a function defined by $Z(u) = [1+u]_+$, with $[v]_+ = \max\{v, 0\}$, and $F(u) = u - 1$ is the inverse function of Z.

Such family $\{g(x, \theta)|\theta \in \Theta\}$ is called a deformed exponential family with a deformed logarithm function F and deformed exponential function Z (refer Harsha and Subrahamanian Moosath 2015; Matsuzoe and Henmi 2013; Naudts 2004 for more details). From Proposition 5.2, Naudts (2004), it can be easily seen that $f(x, \theta)$ is the F-escort distribution so that the variance of the sufficient statistic T attains the Naudts's bound.

Remark 3 Deformed exponential family is a generalization of exponential family in which the deformed logarithm of the density function is a linear function of the statistic T. In an exponential family, the statistic T is sufficient and complete under some conditions. As in exponential family, T is sufficient in deformed exponential family also. For statistical applications, the definition of deformed exponential family should include the requirement that T is a complete statistic.

In the above example, g is a deformed exponential family, while this is not the case in most of the other examples. However, $\mathrm{Var}_{f_\theta}(T)$ attains the bound given by Naudts (2004).

4 Generalized Bhattacharyya bounds

In this section, we obtain an information inequality which generalizes the Bhattacharyya bound given by Fraser and Guttman (1952). This is defined using the divided difference of a density function $g(x, \theta)$ satisfying the conditions (18) and (19). We begin by recalling the definition of the divided difference formula.

4.1 One-parameter case

Definition 1 Let $h(\theta)$ be a scalar function of $\theta \in \Theta \subseteq \mathbb{R}$. Let $k \ge 1$ be a positive integer. Let us define the divided difference of the function h at $k+1$ nodes $\theta^0, \ldots, \theta^k$. We have $k+1$ data points,

$$(\theta^0, h(\theta^0)), \ldots, (\theta^k, h(\theta^k)). \tag{72}$$

Define the first divided difference of h as

$$\underset{\theta^{\nu+1}}{\Delta} h(\theta^{\nu}) := \frac{h(\theta^{\nu+1}) - h(\theta^{\nu})}{\theta^{\nu+1} - \theta^{\nu}}; \quad \nu = 0, \dots k-1. \tag{73}$$

Second divided difference is given by

$$\underset{\theta^{\nu+1},\theta^{\nu+2}}{\Delta^2} h(\theta^{\nu}) := \underset{\theta^{\nu+2}}{\Delta}\ \underset{\theta^{\nu+1}}{\Delta} h(\theta^{\nu}) = \frac{\underset{\theta^{\nu+2}}{\Delta} h(\theta^{\nu+1}) - \underset{\theta^{\nu+1}}{\Delta} h(\theta^{\nu})}{\theta^{\nu+2} - \theta^{\nu}} \tag{74}$$
$$\text{where } \nu = 0, \dots k-2.$$

In general, for $j \geq 2$, the jth divided difference is defined as

$$\underset{\theta^{\nu+1},\dots,\theta^{\nu+j}}{\Delta^j} h(\theta^{\nu}) := \underset{\theta^{\nu+j}}{\Delta} \cdots \underset{\theta^{\nu+1}}{\Delta} h(\theta^{\nu}) = \frac{\underset{\theta^{\nu+2},\dots,\theta^{\nu+j}}{\Delta^{j-1}} h(\theta^{\nu+1}) - \underset{\theta^{\nu+1},\dots,\theta^{\nu+j-1}}{\Delta^{j-1}} h(\theta^{\nu})}{\theta^{\nu+j} - \theta^{\nu}}$$
$$\text{where } \nu = 0 \dots k-j. \tag{75}$$

In particular for $\nu = 0$, the ith divided difference is given by

$$\underset{\theta^1,\dots,\theta^i}{\Delta^i} h(\theta^0) = \sum_{j=0}^{i} \frac{h(\theta^j)}{\prod_{l \neq j}(\theta^j - \theta^l)}, \quad i = 1, \dots k. \tag{76}$$

Choose and fix θ^0 in Θ. For convenience, we write $g_x(\theta)$ instead of $g(x,\theta)$. Let $T(X)$ be an unbiased estimator of a real-valued function $\varphi(\theta)$ of θ. Then consider ith divided difference (for $\nu = 0$) of the density $g(x,\theta)$ on $k+1$ nodes of $\theta^0, \dots, \theta^k$, where $i = 1, \dots, k$. Define

$$S_i = \frac{1}{f(x,\theta^0)} \underset{\theta^1,\dots,\theta^i}{\Delta^i} g_x(\theta^0), \quad i = 1, \dots k. \tag{77}$$

We now give a lower bound for the variance of T using these functions.

Theorem 2 *Let $g(x,\theta)$ be a density function satisfying conditions (18) and (19) with $E_{g_{\theta^0}}[T] = \lambda(\theta^0)$. For $T(X) \in \mathcal{C}_{\varphi}$, the variance of T satisfies*

$$\operatorname{Var}_{f_{\theta^0}}(T(X)) \geq \sup_{\theta^1,\dots,\theta^k} M^{\mathsf{T}} \Sigma^{-1} M \tag{78}$$

where $M = \left(\underset{\theta^1}{\Delta} \lambda(\theta^0), \dots, \underset{\theta^1,\dots,\theta^i}{\Delta^k} \lambda(\theta^0) \right)^{\mathsf{T}}$, $\underset{\theta^1,\dots,\theta^i}{\Delta^i} \lambda(\theta^0)$ is the ith divided difference of λ, $i = 1, \dots, k$ and $\Sigma = (\Sigma_{ij})$ is the covariance matrix of the column vector $S = (S_1, \dots, S_k)^{\mathsf{T}}$.

Proof Note that

$$E_{f_{\theta^0}}[S_i] = \int \underset{\theta^1,\ldots,\theta^i}{\Delta^i} g_x(\theta^0)\mathrm{d}x \tag{79}$$

$$= \int \left(\sum_{j=0}^{i} \frac{g_x(\theta^j)}{\prod_{l \neq j}(\theta^j - \theta^l)} \right) \mathrm{d}x \tag{80}$$

$$= \sum_{j=0}^{i} \frac{1}{\prod_{l \neq j}(\theta^j - \theta^l)} = 0. \tag{81}$$

Also we have

$$E_{f_{\theta^0}}[TS_i] = \int T(x) \underset{\theta^1,\ldots,\theta^i}{\Delta^i} g_x(\theta^0)\mathrm{d}x \tag{82}$$

$$= \int T(x) \left(\sum_{j=0}^{i} \frac{g_x(\theta^j)}{\prod_{l \neq j}(\theta^j - \theta^l)} \right) \mathrm{d}x \tag{83}$$

$$= \sum_{j=0}^{i} \frac{1}{\prod_{l \neq j}(\theta^j - \theta^l)} \int T(x) g_x(\theta^j)\mathrm{d}x \tag{84}$$

$$= \sum_{j=0}^{i} \frac{\lambda(\theta^j)}{\prod_{l \neq j}(\theta^j - \theta^l)} = \underset{\theta^1,\ldots,\theta^i}{\Delta^i} \lambda(\theta^0) \tag{85}$$

where $\underset{\theta^1,\ldots,\theta^i}{\Delta^i} \lambda(\theta^0)$ is the ith divided difference of the function λ.

Hence, it follows that $S_i \in \Psi, \quad i = 1, \ldots, k$.

Apply Proposition 1 for S_i to obtain the bound in Eq. (78) with $M = \left(\underset{\theta^1}{\Delta} \lambda(\theta^0), \ldots, \underset{\theta^1,\ldots,\theta^i}{\Delta^k} \lambda(\theta^0) \right)^{\mathsf{T}}$. □

Remark 4 It follows from Eq. (77) that other choices of functions S_i are possible. For example, one can choose first differences corresponding to $k+1$ distinct points in Θ. When $g = f$, the bound so obtained is the one given by Kshirsagar (2000). See also Koike (2002).

Now let us define

$$S_i = \frac{g^i(x,\theta)}{f(x,\theta)}, \quad i = 1, \ldots k. \tag{86}$$

where $g^i(x,\theta)$ denotes the ith derivative of $g(x,\theta)$ with respect to θ.

Theorem 3 *Let X be a random vector with pdf $f(x,\theta)$. Let $g(x,\theta)$ be a pdf satisfying* (18) *and* (19) *with $E_{g_\theta}[T] = \lambda(\theta)$. Assume that*

(a) *$g(x, \theta)$ and the function $\lambda(\theta)$ are k-times differentiable for all $x \in A$ and $\theta \in \Theta$.* (87)

(b) $\mathrm{Var}_{f_\theta}(S_i) > 0$, *and the covariance matrix $\Sigma = (\Sigma_{ij})$ of the column vector $S = (S_1, \ldots, S_k)^\intercal$ is non-singular.* (88)

(c) *Derivatives of functions of θ expressed as integrals with respect to $g(x, \theta)$ can be obtained by differentiating under the integral sign.* (89)

Then for $T(X) \in \mathcal{C}_\varphi$, the variance of T satisfies

$$\mathrm{Var}_{f_\theta}(T(X)) \geq M^\intercal \Sigma^{-1} M \tag{90}$$

where $M = (\lambda^1(\theta), \ldots, \lambda^k(\theta))^\intercal$, $\lambda^i(\theta)$ is the ith derivative of λ, $i = 1, \ldots, k$.

Proof Note that $E_{f_\theta}[S_i] = 0$ and $E_{f_\theta}[T S_i] = \lambda^i(\theta)$, where $i = 1, \ldots, k$. Hence, $S_i \in \Psi$ from the given assumptions and now apply Proposition 1 for S_i to obtain the bound in Eq. (90) with $M = (\lambda^1(\theta), \ldots, \lambda^k(\theta))^\intercal$. □

Remark 5 Assuming appropriate regularity conditions and considering the limiting case of the bounds in Theorem 2, a bound possibly better than the bound in Eq. (90) results. For the case $g = f$, Koike (2002) gives an example to show that the bound (obtained by these limiting arguments) is indeed sharper.

Remark 6 When $g = f$, Eq. (78) reduces to the Bhattacharyya bounds of order k given by Fraser and Guttman (1952), and for $k = 1$, it gives the Hammersley–Chapman–Robbins bound. When $k = 1$, Eq. (90) reduces to the Naudts's generalized Cramer–Rao bound, and when $g = f$, it reduces to the classical Bhattacharyya bounds of order k in regular case.

4.2 Multiparameter case

Let $X \sim f(x, \underset{\sim}{\theta})$, where $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)^\intercal \in \Theta \subseteq \mathbb{R}^p$. Let $T(X)$ be an unbiased estimator of a real-valued function $\varphi(\underset{\sim}{\theta})$ of $\underset{\sim}{\theta}$. Let $g(x, \underset{\sim}{\theta})$ be a density function parameterized by $\underset{\sim}{\theta}$ satisfying Assumptions (18) and (19). Let the expectation of $T(X)$ with respect to $g(x, \theta)$ is $\lambda(\underset{\sim}{\theta})$, a real-valued function of $\underset{\sim}{\theta}$, i.e., $E_{g_{\underset{\sim}{\theta}}}(T) = \lambda(\underset{\sim}{\theta})$.

Let $k \geq 1$ be an integer. Let $\mathbf{i} = (i_1, \ldots, i_p)$ such that $i_j \geq 0$, $0 < i_1 + \cdots + i_p \leq k$. Assume that the density function $g(x, \underset{\sim}{\theta})$ and function $\lambda(\underset{\sim}{\theta})$ have all partial derivatives with respect to $\theta_1, \ldots, \theta_p$ of order up to k, and kth-order partial derivatives are continuous. Define

$$\partial^{\mathbf{i}} := \frac{\partial^{|\mathbf{i}|}}{\partial\theta^{i_1} \ldots \partial\theta^{i_p}} \quad \text{where} \quad |\,\mathbf{i}\,| := i_1 + \ldots + i_p. \tag{91}$$

Define functions

$$S^{\mathbf{i}} := \frac{1}{f(x, \underset{\sim}{\theta})} \partial^{\mathbf{i}} g(x, \underset{\sim}{\theta}); \quad \lambda^{\mathbf{i}}(\underset{\sim}{\theta}) := \partial^{\mathbf{i}} \lambda(\underset{\sim}{\theta}). \tag{92}$$

We have the following theorem.

Theorem 4 *For $T(X) \in \mathcal{C}_{\varphi}$, the variance of T satisfies*

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(T(X)) \geq M^{\mathsf{T}} \Sigma^{-1}(\underset{\sim}{\theta}) M \tag{93}$$

where Σ is the covariance matrix of the column vector $S = (S^{\mathbf{i}})$ containing all possible $S^{\mathbf{i}}$ and $M = (\lambda^{\mathbf{i}}(\underset{\sim}{\theta}))$ is a column vector containing all possible $\lambda^{\mathbf{i}}(\underset{\sim}{\theta})$.

Proof Note that

$$E_{f_{\underset{\sim}{\theta}}}(S^{\mathbf{i}}) = \int \partial^{\mathbf{i}} g(x, \underset{\sim}{\theta}) \mathrm{d}x = 0 \tag{94}$$

$$\mathrm{Cov}_{f_{\underset{\sim}{\theta}}}(T, S^{\mathbf{i}}) = \int T(x)\, \partial^{\mathbf{i}} g(x, \theta)\, \mathrm{d}x = \lambda^{\mathbf{i}}(\underset{\sim}{\theta}). \tag{95}$$

Hence, for all $\mathbf{i}$, we have $S^{\mathbf{i}} \in \Psi$. Now apply Proposition 1 for $S^{\mathbf{i}} \in \Psi$ to obtain the bound. □

Remark 7 If $\mid \mathbf{i} \mid = 1$, the bound in Eq. (93) reduces to Naudts's bound in vector parameter.

Note 1 If the density $g(x, \underset{\sim}{\theta})$ is not regular, we can obtain an information inequality by replacing the partial derivatives by the corresponding divided difference formula. This is done as follows.

Consider a scalar function $h(\underset{\sim}{\theta})$ of $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)$. Let $k \geq 1$ be an integer. Let us consider $k+1$ nodes of $\underset{\sim}{\theta}$ say, $\underset{\sim}{\theta}^0 = (\theta_1^0, \ldots, \theta_p^0), \ldots, \underset{\sim}{\theta}^k = (\theta_1^k, \ldots, \theta_p^k)$. Define

$$\underset{\sim}{\theta}^{\nu}_{i} = (\theta_1^{\nu}, \ldots, \theta_i^{\nu+1}, \ldots \theta_p^{\nu}). \tag{96}$$

Define the first divided difference of h as

$$\underset{\theta_i^{\nu+1}}{\Delta} h(\underset{\sim}{\theta}^{\nu}) = \frac{h(\underset{\sim}{\theta}^{\nu}_{i}) - h(\underset{\sim}{\theta}^{\nu})}{\theta_i^{\nu+1} - \theta_i^{\nu}} \tag{97}$$

where $\nu = 0, \ldots, k-1$, $j = 1$, $i = 1, \ldots, p$.

In general, for $j \geq 2$, define jth divided difference of h as

$$\underset{\theta_i^{\nu+1},\ldots,\theta_i^{\nu+j}}{\Delta^j} h(\underset{\sim}{\theta}^{\nu}) := \underset{\theta_i^{\nu+j}}{\Delta} \cdots \underset{\theta_i^{\nu+1}}{\Delta} h(\underset{\sim}{\theta}^{\nu}) = \frac{\underset{\theta_i^{\nu+2},\ldots,\theta_i^{\nu+j}}{\Delta^{j-1}} h(\underset{\sim}{\theta}^{\nu+1}) - \underset{\theta_i^{\nu+1},\ldots,\theta_i^{\nu+j-1}}{\Delta^{j-1}} h(\underset{\sim}{\theta}^{\nu})}{\theta_i^{\nu+j} - \theta_i^{\nu}}$$
$$\text{where } \nu = 0 \ldots k-j;\; j = 1,\ldots,k;\; i = 1,\ldots,p. \tag{98}$$

In many cases, one may be interested in estimating a vector-valued function $\Phi(\underset{\sim}{\theta})$ of $\underset{\sim}{\theta}$. Let $\mathsf{T} = (T_1, \ldots, T_r)^{\mathsf{T}}$ be an unbiased estimator of $\Phi(\underset{\sim}{\theta}) = (\varphi_1(\underset{\sim}{\theta}), \ldots, \varphi_r(\underset{\sim}{\theta}))^{\mathsf{T}}$, where $r \leq p$. That is $E_{f_{\underset{\sim}{\theta}}}(T_i) = \varphi_i(\underset{\sim}{\theta})$, $i = 1, \ldots, r$. Let us consider $\mathsf{S} = (S_1, \ldots, S_m)^{\mathsf{T}}$, where functions $S_i \in \Psi$, $i = 1, \ldots, m$. Let us assume that the covariance matrix Σ of $(r+m) \times 1$ vector (T, S) is positive definite. We have

$$\Sigma = \begin{bmatrix} \Sigma_{\mathsf{T}} & \Sigma_{\mathsf{TS}} \\ \Sigma_{\mathsf{ST}} & \Sigma_{\mathsf{S}} \end{bmatrix} \tag{99}$$

where Σ_{T} is the covariance matrix of $r \times 1$ vector T, Σ_{S} is the covariance matrix of $m \times 1$ vector S and Σ_{TS} is the covariance matrix between T and S. If the covariance matrix Σ_{S} is invertible, the Shur complement of Σ_{S} in Σ is given by $\Sigma_{\mathsf{T}} - \Sigma_{\mathsf{TS}}\Sigma_{\mathsf{S}}^{-1}\Sigma_{\mathsf{ST}}$. It is easy to see that $\Sigma_{\mathsf{T}} - \Sigma_{\mathsf{TS}}\Sigma_{\mathsf{S}}^{-1}\Sigma_{\mathsf{ST}}$ is positive definite since Σ is positive definite. This can be written as

$$\Sigma_{\mathsf{T}} - \Sigma_{\mathsf{TS}}\Sigma_{\mathsf{S}}^{-1}\Sigma_{\mathsf{ST}} \succ 0. \tag{100}$$

Equivalently, one can write

$$\Sigma_{\mathsf{T}} \succ \Sigma_{\mathsf{TS}}\Sigma_{\mathsf{S}}^{-1}\Sigma_{\mathsf{ST}} \tag{101}$$

which means that $\Sigma_{\mathsf{T}} - \Sigma_{\mathsf{TS}}\Sigma_{\mathsf{S}}^{-1}\Sigma_{\mathsf{ST}}$ is positive definite.

The above inequality can be interpreted as follows. Consider a linear estimator $\alpha^{\mathsf{T}}T$ which is unbiased for $\alpha^{\mathsf{T}}\Phi(\underset{\sim}{\theta})$. Then

$$\mathrm{Var}_{f_{\underset{\sim}{\theta}}}(\alpha^{\mathsf{T}}T) \geq \alpha^{\mathsf{T}}J(\underset{\sim}{\theta})\alpha, \tag{102}$$

where $J(\theta) = \Sigma_{\mathsf{TS}}\Sigma_{\mathsf{S}}^{-1}\Sigma_{\mathsf{ST}}$.

If $\Sigma_{\mathsf{T}} = J(\theta)$, then variance of $\alpha^{\mathsf{T}}T$ attains this bound. That is, $\alpha^{\mathsf{T}}T$ is the minimum variance unbiased estimator for $\alpha^{\mathsf{T}}\Phi(\underset{\sim}{\theta})$ for any α.

Remark 8 An information bound provides insight into the statistical model as it shows that the error in estimation cannot be made arbitrarily small. It is of interest, therefore, to know whether a bound is sharp. The presence of a complete sufficient statistic is useful in this context. Conversely, if a bound is sharp, the *uniform minimum variance unbiased estimator* (UMVUE) is automatically identified. This fact may be useful even when a complete sufficient statistic exists. This is so because the Lehmann–Scheffe theorem gives only the existence of the UMVUE. Attainment of a bound is also relevant in the geometric interpretation of a statistical model as discussed in Sect. 5.

5 Discussions

In Proposition 5.2, Naudts (2004), it is shown that if g is a deformed exponential family with a statistic T, then there exists an escort family f such that the variance of T attains the Naudts's bound. Considering this from a statistical perspective, let f be the original model and assume that there exists a deformed exponential family $\mathcal{S} = \{g(x;\theta) \mid \theta \in \Theta \subseteq \mathbb{R}\}$ with a canonical statistic T given by

$$g(x;\theta) = Z(\theta T(x) - \phi(\theta)) \quad \text{or} \quad F(g(x;\theta)) = \theta T(x) - \phi(\theta) \tag{103}$$

where $F : (0, \infty) \longrightarrow \mathbb{R}$ is a smooth function satisfying $F'(x) > 0$ and $F''(x) < 0$, Z is the inverse function of F and $\phi(\theta)$ is chosen such that g is a probability density function.

Then assume that f is the F-escort distribution of g defined by

$$f(x,\theta) = \frac{1}{F'(g)h_F(\theta)}; \quad h_F(\theta) = \int \frac{1}{F'(g(x;\theta))}\mathrm{d}x. \tag{104}$$

Then it is easy to see that T is an unbiased estimator of the expectation parameter $\eta = E_{f_\theta}(T)$ and the variance of T attains the Naudts's bound.

For the geometric interpretation, we first observe that if g is an exponential family, then the original model f is equal to g. Then the estimator T is an unbiased estimator of the expectation parameter, and its variance attains the Cramer–Rao lower bound. The exponential family has a dually flat structure. The expectation parameter is the dual coordinate in the dually flat α-geometry by Amari (1985) with $\alpha = 1$. When g is a deformed exponential family with dually flat χ-geometric structure, then $E_{f_\theta}(T)$ is the dual coordinate (refer Amari et al. 2012; Harsha and Subrahamanian Moosath 2015 for more details). As observed above, T is an unbiased estimator for $E_{f_\theta}(T)$ and its variance attains the Naudts's lower bound. In the context of statistical inference, the χ-geometry of the deformed exponential family seems to provide a useful generalization of 1-geometry of an exponential family. It would be an interesting problem to construct a differential geometric framework for the parameter estimation problem in a deformed exponential family.

Acknowledgements The first author is supported by the Postdoctoral Fellowship from Indian Institute of Technology Bombay, Mumbai, India.

Appendix

Lemma 1 *Let X be a random variable with pdf*

$$f(x,\theta)=\mathcal{N}(0,\theta^2)=\frac{1}{\sqrt{2\pi}\theta}e^{\frac{-x^2}{2\theta^2}},\quad x\in\mathbb{R}\ \text{ and }\ \theta>0. \tag{105}$$

Consider a pdf

$$g(x,\theta)=\frac{1}{\sqrt{2\pi}\theta}\left(\frac{3}{4}+\frac{x^2}{4\theta^2}\right)e^{\frac{-x^2}{2\theta^2}},\quad x\in\mathbb{R}. \tag{106}$$

Then the family $\{g_\theta \mid \theta>0\}$ *satisfies Assumptions* (18) *and* (19).

Proof Since $\mathrm{supp}(g_\theta)=\mathrm{supp}(f_\theta)$, P_{g_θ} is absolutely continuous with respect to P_{f_θ} for all $\theta\in\Theta$. Hence, $\{g_\theta \mid \theta>0\}$ satisfies Assumption (18).

To show $\mathcal{U}_f\subseteq\mathcal{U}_g$, let $U(X)\in\mathcal{U}_f$. Then we have for all $\theta\in\Theta$,

$$E_{f_\theta}(U)=\int_{x\in\mathbb{R}}U(t)\ \exp(-t^2/2\theta^2)\ \mathrm{d}t=0. \tag{107}$$

By differentiating (107) both sides with respect to θ, we get

$$\int_{x\in\mathbb{R}}U(t)\ t^2\ \exp(-t^2/2\theta^2)\ \mathrm{d}t=0. \tag{108}$$

Now from Eqs. (107) and (108), it follows that

$$E_{g_\theta}(U)=\frac{3}{4\sqrt{2\pi}\theta}\int_{x\in\mathbb{R}}U(t)\ \exp(-t^2/2\theta^2)\ \mathrm{d}t \tag{109}$$

$$+\frac{3}{4\sqrt{2\pi}\theta^3}\int_{x\in\mathbb{R}}U(t)\ t^2\ \exp(-t^2/2\theta^2)\ \mathrm{d}t \tag{110}$$

$$=0. \tag{111}$$

Hence, $\{g_\theta \mid \theta>0\}$ satisfies Assumption (19). □

References

Amari, S. (1985). *Differential-geometrical methods in statistics*, Lecture notes in statistics, Vol. 28. New York: Springer.

Amari, S., Ohara, A., Matsuzoe, H. (2012). Geometry of deformed exponential families: Invariant, dually flat and conformal geometries. *Physica A*: *Statistical Mechanics and Its Applications*, *391*, 4308–4319.

Chapman, D. G., Robbins, H. (1951). Minimum variance estimation without regularity assumptions. *Annals of Mathematical Statistics*, *22*(4), 581–586.

Cramer, H. (1946). A contribution to the theory of statistical estimation. *Scandinavian Actuarial Journal*, *1*, 85–94.

Fraser, D. A. S., Guttman, I. (1952). Bhattacharyya bounds without regularity assumptions. *Annals of Mathematical Statistics*, *23*(4), 629–632.

Hammersley, J. M. (1950). On estimating restricted parameters. *Journal of the Royal Statistical Society Series B*, *12*(2), 192–240.

Harsha, K. V., Subrahamanian Moosath, K. S. (2015). Dually flat geometries of the deformed exponential family. *Physica A*: *Statistical Mechanics and its Applications*, *433*, 136–147.

Khatri, C. G. (1980). Unified treatment of Cramér–Rao bound for the nonregular density functions. *Journal of Statistical Planning and Inference*, *4*, 75–79.
Koike, K. (2002). On the inequality of Kshirsagar. *Communications in Statistics*: *Theory and Methods*, *31*, 1617–1627.
Kshirsagar, A. M. (2000). An extension of the Chapman–Robbins inequality. *Journal of the Indian Statistical Association*, *38*, 355–362.
Lehmann, E. L., Casella, G. (1998). *Theory of point estimation*, 2nd ed. New York: Springer.
Matsuzoe, H., Henmi, M. (2013). Hessian structures on deformed exponential families. In F. Nielsen, F. Barbaresco (Eds.), *Geometric science of information*, Lecture notes in computer science, Vol. 8085, pp. 275–282. Berlin: Springer.
Naudts, J. (2004). Estimators, escort probabilities, and ϕ-exponential families in statistical physics. *Journal of Inequalities in Pure and Applied Mathematics*, *5*(4), Article 102.
Qin, M., Nayak, T. K. (2008). Kshirsagar-type lower bounds for mean squared error of prediction. *Communications in Statistics*: *Theory and Methods*, *37*, 861–872.
Rao, C. R. (1945). Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, *37*, 81–89.
Vincze, I. (1979). On the Cramer–Frechet–Rao inequality in the non-regular case. In *Contributions to statistics*, The Jaroslav Hajek Memorial Volume, pp. 253–262. Prague: Academia.