Journal of Multivariate Analysis 101 (2010) 1814–1825

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A van Trees inequality for estimators on manifolds

P.E. Jupp

School of Mathematics and Statistics, University of St Andrews, St Andrews KY16 9SS, UK

ARTICLE INFO

Article history:
Received 8 June 2009
Available online 19 March 2010

AMS 2000 subject classifications:
primary 62F10
62F15
secondary 62B10

Keywords:
Bayes risk
Bias
Cramér–Rao inequality
Fisher information
Hessian
Proper dispersion model
Tensor

ABSTRACT

Van Trees' Bayesian version of the Cramér–Rao inequality is generalised here to the context of smooth loss functions on manifolds and estimation of parameters of interest. This extends the multivariate van Trees inequality of Gill and Levit (1995) [R.D. Gill, B.Y. Levit, Applications of the van Trees inequality: a Bayesian Cramér–Rao bound, Bernoulli 1 (1995) 59–79]. In addition, the intrinsic Cramér–Rao inequality of Hendriks (1991) [H. Hendriks, A Cramér–Rao type lower bound for estimators with values in a manifold, J. Multivariate Anal. 38 (1991) 245–261] is extended to cover estimators which may be biased. The quantities used in the new inequalities are described in differential-geometric terms. Some examples are given.

1. Introduction

Let $\{f(\cdot;\theta) : \theta \in \Theta\}$ be a family of probability density functions on a sample space $\mathcal{X}$, where the parameter space Θ is an open set in $\mathbb{R}^d$. One of the fundamental results in mathematical statistics is that (under weak regularity conditions) any unbiased estimator $\hat{\theta}$ of θ satisfies the multivariate Cramér–Rao inequality

$$E_\theta\left[\left(\hat{\theta}(x)-\theta\right)\left(\hat{\theta}(x)-\theta\right)^T\right] \geq I(\theta)^{-1}, \tag{1.1}$$

i.e. $E_\theta[(\hat{\theta}(x)-\theta)(\hat{\theta}(x)-\theta)^T]-I(\theta)^{-1}$ is positive semi-definite. Here E_θ denotes expectation with respect to $f(\cdot;\theta)$ and $I(\theta)$ denotes the Fisher information matrix at θ. The matrix $E_\theta[(\hat{\theta}(x)-\theta)(\hat{\theta}(x)-\theta)^T]$ is the variance matrix of $\hat{\theta}(x)$.

There are various Bayesian versions of the Cramér–Rao inequality. For the simplest of these, let π be a proper prior distribution on Θ. Then taking the expectation of (1.1) over π and using convexity of the function $X \mapsto X^{-1}$ on the set of positive-definite $d \times d$ matrices yield the *Bayesian Cramér–Rao inequality*

$$E_\pi\left[E_\theta\left[\left(\hat{\theta}(x)-\theta\right)\left(\hat{\theta}(x)-\theta\right)^T\right]\right] \geq E_\pi\left[I(\theta)\right]^{-1}, \tag{1.2}$$

where E_π denotes expectation with respect to π. If Θ is connected then equality holds in (1.2) if and only if the probability density functions have the form

$$f(x;\theta) = \frac{a(x)}{\lambda(\theta)}\exp\left\{-\left(\hat{\theta}(x)-\theta\right)^T\Psi\left(\hat{\theta}(x)-\theta\right)-c(\Psi)\right\} \tag{1.3}$$

E-mail address: pej@st-andrews.ac.uk.

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.03.007

for some positive function a on $\mathcal{X}$, positive-definite matrix Ψ and constant $c(\Psi)$. For a given function a, the family of probability density functions obtained from (1.3) by letting both θ and Ψ vary is a proper multivariate dispersion model in the sense of Jørgensen and Lauritzen [1]. One consequence of (1.2) is that if $\hat{\theta}$ is unbiased then

$$E_\pi\left[\mathrm{var}_\theta\left(\hat{\theta}^i(x)\right)\right] \geq E_\pi\left[I_{ii}(\theta)\right]^{-1} \quad 0 \leq i \leq d,$$

where $I_{ii}(\theta)$ denotes the (i, i)-element of $I(\theta)$. This inequality was given by Gart (see (3.9) of [2]). The Bayesian Cramér–Rao inequality (1.2) can be generalised to the case in which $\hat{\theta}$ may be biased. It is straightforward to show that

$$E_\pi\left[E_\theta\left[\left(\hat{\theta}(x)-\tau(\theta)\right)\left(\hat{\theta}(x)-\tau(\theta)\right)^T\right]\right] \geq E_\pi\left[\tau'(\theta)\right] E_\pi\left[I(\theta)\right]^{-1} E_\pi\left[\tau'(\theta)\right]^T, \tag{1.4}$$

where $\tau(\theta) = E_\theta[\hat{\theta}(x)]$ and $\tau'(\theta) = \mathrm{d}\tau(\theta)/\mathrm{d}\theta$.

A more interesting Bayesian inequality is the *van Trees inequality*

$$E_\pi\left[E_\theta\left[\left(\hat{\theta}(x)-\theta\right)\left(\hat{\theta}(x)-\theta\right)^T\right]\right] \geq \left\{E_\pi\left[I(\theta)\right] + \mathcal{I}(\pi)\right\}^{-1} \tag{1.5}$$

([3], pp. 72, 84), where the estimator $\hat{\theta}$ need not be unbiased and $\mathcal{I}(\pi)$ is the *information in the prior* π, defined as

$$\mathcal{I}(\pi) = E_\pi\left[\left(\frac{\partial \log \lambda(\theta)}{\partial\theta}\right)^T \frac{\partial \log \lambda(\theta)}{\partial\theta}\right], \tag{1.6}$$

λ being the density of π with respect to Lebesgue measure. Letac [4] has shown that equality holds in (1.5) if and only if the family $\{f(\cdot;\theta) : \theta \in \Theta\}$ is a general exponential family with canonical statistic $\{E_\pi[I(\theta)] + \mathcal{I}(\pi)\}^{-1}\hat{\theta}(x)$.

In the case $\Theta = \mathbb{R}$, inequality (1.5) was given by Schützenberger [5]. A careful version of (1.5) under very weak assumptions was given (for $\Theta = \mathbb{R}$) by Lenstra [6]. Some applications and generalisations have been considered by Bobrovsky et al. [7] and by Brown and Gajek [8]. Other generalisations and various applications to engineering are presented in the anthology by van Trees and Bell [9].

If $\hat{\theta}$ is unbiased then the van Trees inequality (1.5) follows immediately from the Bayesian Cramér–Rao inequality (1.2). On the other hand, Letac [10] has pointed out that if $\hat{\theta}$ is allowed to be biased then (1.5) is not a consequence of the generalisation (1.4) of (1.2).

Gill and Levit [11], Section 4, generalised the van Trees inequality (1.5) to parameters of interest and weighted quadratic loss as follows. Let Ω be an open set in $\mathbb{R}^p$, G be a positive-definite symmetric $p \times p$ matrix-valued function on Θ, and A be a $p \times d$ matrix-valued function on Θ. The *prior divergence* of A is the p-vector field $\mathrm{div}_\pi A$ on Ω with components

$$(\mathrm{div}_\pi A)_a = \frac{1}{\lambda}\sum_{i=1}^{d}\frac{\partial}{\partial\theta^i}\left(A_a^i\lambda\right),$$

where $\theta = (\theta^1, \ldots, \theta^d)$ and A_a^i (for $i = 1, \ldots, d$ and $a = 1, \ldots, p$) are the components of A. The *G-weighted information in the prior* π *given by* A is the scalar $\mathcal{I}_G(A, \pi)$ defined as

$$\mathcal{I}_G(A,\pi) = E_\pi\left[(\mathrm{div}_\pi A)^T(\theta) G(\theta)^{-1}(\mathrm{div}_\pi A)(\theta)\right].$$

Gill and Levit showed that, under mild regularity conditions, for any estimator $\hat{\phi}$ of $\phi(\theta)$,

$$E_\pi\left[E_\theta\left[\left(\hat{\phi}(x)-\phi(\theta)\right)^T G(\theta)^{-1}\left(\hat{\phi}(x)-\phi(\theta)\right)\right]\right] \geq \frac{\left\{E_\pi\left[\mathrm{tr}\left(A(\theta)G(\theta)^{-1}\frac{\partial\phi(\theta)}{\partial\theta}\right)\right]\right\}^2}{E_\pi\left[\mathrm{tr}\left(G(\theta)^{-1}A(\theta)^T I(\theta)A(\theta)\right)\right] + \mathcal{I}_G(A,\pi)}. \tag{1.7}$$

Inequality (1.7) has been used by Gill and Levit [11] to derive minimax convergence rates in some non-parametric and semi-parametric problems and by Gill [12] to obtain an asymptotic lower bound on Bayes risk in quantum statistical estimation.

A reasonable general setting for parametric inference involves parameter spaces that are differential manifolds. Thus it is striking that the inequality (1.7) has the implicit restrictions that (i) both the full parameter space Θ and the space Ω of interest parameters are subsets of Euclidean spaces, (ii) the loss function is quadratic. The objectives of this note are (a) to give a version of inequality (1.7) for arbitrary smooth loss functions on manifolds, (b) to provide geometric interpretations of the quantities that arise. These objectives are achieved by using the derivative of the loss function, i.e. by working in cotangent spaces.

Section 2 presents the intrinsic van Trees inequality for general smooth loss functions. Various geometric Cramér–Rao inequalities are given in Section 3. Some examples are considered in Section 4.