Revisiting the van Trees inequality in the spirit of Hájek and Le Cam

Elisabeth Gassiat

Laboratoire de mathématiques
Université Paris-Sud
Orsay, France

elisabeth.gassiat@math.u-psud.fr

&

David Pollard

Yale University
New Haven, CT, United States

david.pollard@yale.edu

&

Gilles Stoltz

GREGHEC (HEC Paris, CNRS)
Jouy-en-Josas, France

stoltz@hec.fr

First version: July 12, 2001.
This version: September 10, 2013.

Abstract

This paper revisits the multi-dimensional van Trees inequality in an intrinsic form and under minimal assumptions, in the spirit of Hájek and Le Cam. We prove that the van Trees inequality is a Cramér-Rao inequality for some Bayesian location model. We add to the known long series of applications of this inequality a simple way to prove local asymptotic minimax (LAM) lower bounds for the quadratic risk in parametric and semi-parametric settings.

Introduction: To be totally reshuffled and toned down

1. Introduction

A long story made short. Once upon a time (in April 2001), a student (Gilles Stoltz) had to get a grade for a graduate course on the asymptotic theory of statistical estimation. The lecturer (David Pollard, visiting Paris for a semester) had the intuition that some developments and improvements around the van Trees inequality (van Trees, 1968) could be obtained in the spirit of Hájek and Le Cam. More precisely, he wrote the following statement for a take-home examination:

> On the van Trees inequality: Replace assumptions of Gill and Levit [1995] by analogous assumptions of Hellinger differentiability. Try to deduce the van Trees inequality from the information inequality (the Cramér-Rao bound) for a parametric family $m_\alpha(x, \theta) = q_\alpha(\theta)\, f_{\theta-\alpha}(x)$, where q is a density with respect to Lebesgue mesure on Θ with compact support, and $q_\alpha = q(\,\cdot - \alpha)$. Prove a rigorous theorem, if you can. Illustrate the application of the theorem by adapting one of the examples of Gill and Levit [1995]. Even better: Use the theorem to prove a rigorous efficiency result under differentiability-in-quadratic-mean assumptions.

This statement was the sparkle of discussions and iterations, of some, separate and joint, work. The research programme was completed in three months, with an unexpected additional finding: it became clear that some direct proof of the van Trees inequality based on an ad hoc information inequality (and exploiting some separation of the variables x and θ) would require much lighter assumptions than its derivation via the Cramér-Rao bound. An account of this is already mentioned in the July 2001 version of Pollard [2001; 2005].

Right after (in July 2001) David Pollard had to go back home to New Heaven, Connecticut, but he did not so without warmly encouraging Gilles Stoltz to polish and publish the above-mentioned results. Time passed, Gilles Stoltz completed a PhD thesis in machine learning and would have given up writing up the present paper if a third researcher (Elisabeth Gassiat) had not been around. She has been presenting (from September 2004 on) in her own graduate lectures at Université Paris-Sud, Orsay, the derivations of the van Trees inequality in the spirit of Hájek and Le Cam together with a simple and direct application to local asymptotic minimax (LAM) lower bounds for the quadratic risk in parametric and semi-parametric settings. The three of us regularly came back to the van Trees inequality over the years, progressively simplifying its (direct and indirect) proofs and relaxing the needed assumptions. We unfortunately did not finish in time for 2010 and could not honor the memory of Lucien Le Cam for the 10th anniversary of his passing away.

The van Trees inequality: pointwise versus $\mathbb{L}_2$–type assumptions. The aim of this paper is to revisit the van Trees inequality, originally introduced and proved by van Trees [1968], in the framework of the theory of Hájek and Le Cam. We do so to further illustrate the elegancy and neatness gained by working in this framework, compared to assuming some pointwise regular behavior of the densities. Almost all articles devoted to the van Trees inequality, the ones of Bobrovsky et al. [1987], Gill and Levit [1995], Letac [2008], and Jupp [2010], considered such pointwise assumptions, with the notable exception of Lenstra [2005], whose results we discuss below.

But the work of Hájek and Le Cam is nowadays the reference for the asymptotic theory of statistical estimation, see also Ibragimov and Has'minskii [1981]. In this theory, smoothness of parametric models is considered in the $\mathbb{L}_2$–sense, by considering square roots of densities as the basic objects. In terms of efficiency, one good way of generalizing the Cramér-Rao inequality to sequences of estimators is via lower bounds for the local asymptotic minimax (LAM) risk.

In this respect, the van Trees inequality, which may be seen as some Bayesian version of the Cramér-Rao inequality suited for biased statistics, was proved by Gill and Levit [1995] to lead to a simple efficiency result for sequences of regular estimators only (see Section 6.1).

Our contributions. First of all, we give an elegant direct proof of the van Trees inequality in a Hájek–Le Cam setting, with somewhat minimal assumptions (almost all of them merely ensure that the quantities at hand are well-defined). We only assume smoothness through differentiability in quadratic mean and absolute continuity of the function of the parameter to estimate. Also, our aim is to give a unifying statement that applies to various settings, where the parameter space Θ is compact (as in Gill and Levit, 1995) or is not compact ($\Theta = \mathbb{R}$ or $\Theta = \mathbb{R}^p$, as in Lenstra, 2005)—see Theorem 1 and Corollary 3. A detailed comparison of our version to the version by Gill and Levit [1995] can be found in Section 4.3: we prove that our assumptions are strictly milder than theirs. In the same vein, the version of van Trees obtained by Lenstra [2005] in a Hájek–Le Cam setting is not satisfactory yet as it still requires some pointwise regularity of the density functions and Theorem 1 proves that none of them was necessary.

In a second part, we indicate that the van Trees inequality is exactly a Cramér-Rao bound for a well-defined location model, not just something close to a Cramér-Rao bound (e.g., some Bayesian Cramér-Rao bound).

Finally, we show how the van Trees inequality yields an elementary proof of local asymptotic minimax (LAM) lower bounds for positive quadratic risk functions in parametric and semi-parametric contexts; these bounds are valid for all sequences of estimators (not just regular ones).

Outline. In Section 2 we recall some basic definitions and results of the Hájek–Le Cam setting (including the information equality and the Cramér-Rao bound). We state our minimal and elegant version of the van Trees inequality in Section 4, discuss variations around it, and compare it to previous results in the literature. Section 5 explains that the van Trees inequality is exactly a Cramér-Rao bound for some well-defined location model based on Bayesian mixtures of densities. We propose our application to efficiency, namely, local asymptotic minimax (LAM) lower bounds, in Section 6. Finally, Section 7 and Appendix A gather some technical material (in particular, the detailed proofs of the main results).

2. Setup and notation

We consider a statistical model $(\mathbb{P}_\theta)$, defined on a measurable space $(\mathcal{X}, \mathcal{A})$ and indexed by $\theta \in \Theta$, where Θ is an open subset of $\mathbb{R}^p$, with $p \geqslant 1$. We assume that it is dominated by a sigma-finite measure μ, that is, for all $\theta \in \mathbb{R}^p$, the probability $\mathbb{P}_\theta$ is absolutely continuous with respect to μ, with density denoted by f_θ. We denote by $\mathbb{E}_\theta$ the expectation with respect to $\mathbb{P}_\theta$.

In the sequel, unless stated otherwise, the norms will refer to the Euclidian norms, denoted by $\|\cdot\|$ in finite-dimensional real vector spaces, and by $\|\cdot\|_\mu$ in $\mathbb{L}_2(\mu)$; in the latter case, for all functions $g \in \mathbb{L}_2(\mu)$, possibly vector-valued,

$$\|g\|_\mu = \sqrt{\int_{\mathcal{X}} \|g\|^2 \, \mathrm{d}\mu}\,.$$

In particular, we will denote by $\|\cdot\|_1$ the ℓ^1–norms; for a matrix $A = [a_{i,j}]_{i,j}$, we define

$$\|A\|_1 = \sum_{i,j} |a_{i,j}|\,.$$

We denote by $\xi_\theta = \sqrt{f_\theta}$ the square roots of the densities.

Definition 1 (Differentiability in $\mathbb{L}_2$)**.** *The dominated statistical model $(\mathbb{P}_\theta)_{\theta\in\Theta}$ is differentiable in $\mathbb{L}_2(\mu)$ at $\theta_0 \in \Theta$ if there exists a p–dimensional vector-valued function $\dot{\xi}_{\theta_0} \in \mathbb{L}_2(\mu)$ such that*

$$\left\| \xi_{\theta_0} - \xi_\theta - (\theta_0 - \theta)^{\mathrm{T}} \, \dot{\xi}_{\theta_0} \right\|_\mu = o\big(\|\theta_0 - \theta\|\big) \qquad \textit{as} \quad \theta \to \theta_0.$$

The vector-valued function $\dot{\xi}_{\theta_0} = \big(\dot{\xi}_{\theta_0,i}\big)_{i\in\{1,\ldots,p\}}$ is called the derivative of the model at θ_0. The Fisher information $\mathcal{I}(\theta_0)$ of the model at θ_0 is then defined by

$$\mathcal{I}(\theta_0) = 4 \int_{\mathcal{X}} \dot{\xi}_{\theta_0} \dot{\xi}_{\theta_0}^{\mathrm{T}} \, \mathrm{d}\mu\,. \tag{1}$$

We denote the components of θ by $\theta = (\theta_1, \ldots, \theta_p)$. For $i \in \{1, \ldots, p\}$, we refer to the $(p-1)$–dimensional vector of all components of θ but the i–th one as θ_{-i}, so that $\theta = (\theta_i, \theta_{-i})$. The standard definition of absolute continuity for functions $\mathbb{R} \to \mathbb{R}$ can be generalized to functions $\mathbb{R}^p \to \mathbb{R}$ in several ways. In this paper, the convenient extension is the following.

Definition 2 (Absolute continuity)**.** *Let $D \subseteq \mathbb{R}^p$ be an open domain. A function $\varphi : D \to \mathbb{R}$ is absolutely continuous if for almost all $\theta_{-i} \in \mathbb{R}^{p-1}$, the function $\theta_i \mapsto \varphi\big(\theta_i, \theta_{-i}\big)$ is absolutely continuous on the open domain $D(\theta_{-i}) = \big\{\theta_i \in \mathbb{R} : \ (\theta_i, \theta_{-i}) \in D\big\}$ whenever the latter is non empty.*

For $s \geqslant 1$, a vector-valued function $\psi = (\psi_j)_{1\leqslant j\leqslant s} : D \to \mathbb{R}^s$ is absolutely continuous if all its components $\psi : D \to \mathbb{R}$ are absolutely continuous.

In particular, the gradient

$$\nabla\varphi = \left[\frac{\partial\varphi}{\partial\theta_i}\right]_{i\in\{1,\ldots,p\}} \tag{2}$$

of an absolutely continuous function $\varphi : \mathbb{R}^p \to \mathbb{R}$ exists at almost all $\theta \in \mathbb{R}^p$. Moreover, for all $i \in \{1, \ldots, p\}$, almost all $\theta_{-i} \in \mathbb{R}^{p-1}$, and all real numbers a and b such that $(a, b) \subseteq D(\theta_{-i})$,

$$\int_a^b \frac{\partial\varphi}{\partial\theta_i}\big(\theta_i, \theta_{-i}\big) \, \mathrm{d}\theta_i = \varphi\big(b, \theta_{-i}\big) - \varphi\big(a, \theta_{-i}\big)\,.$$

By convention, we let the gradient $\nabla\varphi$ equal $[0]$ at points θ where it was not defined.

Example 1 (Location model)**.** Fix an absolutely continuous density function q on $\mathbb{R}^p$, with almost-sure gradient denoted by ∇q, such that $\|\nabla q\|^2 \, \mathbb{I}_{\{q>0\}}/q$ is integrable with respect to the Lebesgue measure λ. We consider the statistical model $(\mathbb{Q}_\alpha)$ indexed by $\alpha \in \mathbb{R}^p$ and formed by the probability distributions $\mathbb{Q}_\alpha$ with density function $q(\,\cdot\, - \alpha)$ with respect to λ. A standard result (which is stated in a more general way as Proposition 2 in appendix) is that this model is differentiable in $\mathbb{L}_2(\lambda)$ at all points of $\mathbb{R}^p$, with derivative at $\alpha_0 \in \mathbb{R}^p$ equal to

$$-\frac{1}{2}\frac{\nabla q(\,\cdot\, - \alpha_0)}{\sqrt{q(\,\cdot\, - \alpha_0)}}\,\mathbb{I}_{\left\{q(\,\cdot\,-\alpha_0)>0\right\}}\,. \tag{3}$$

Consequently, the Fisher information of this model is independent of $\alpha_0 \in \mathbb{R}^p$ and equals

$$\mathcal{I}_q \stackrel{\text{def}}{=} \int_{\mathbb{R}} \nabla q \, \nabla q^{\mathrm{T}} \, \frac{\mathbb{I}_{\{q>0\}}}{q} \, \mathrm{d}\lambda\,.$$

3. Reminder of some basic facts around the information (in)equality

We recall versions of the information (in)equalities for one-dimensional and then for multidimensional statistics (which will always be denoted, respectively, by T and S). The material that follows is completely standard and we state these results only to make this paper self-contained. Lemma 2 and the proof techniques of Corollary 1 will indeed be key elements in the proofs of Section 4, while the bound of Section 3.3 is stated for the sake of later comparisons.

3.1. An information equality for one-dimensional statistics

We consider a statistical model satisfying the following assumption.

Assumption ($\mathbb{L}_2$.Diff.θ_0)**.** *The model* $(\mathbb{P}_\theta)_{\theta\in\Theta}$*, where* $\Theta \subseteq \mathbb{R}^p$ *is an open set, is dominated by* μ *and is differentiable in* $\mathbb{L}_2(\mu)$ *at* $\theta_0 \in \Theta$.

Lemma 1 (Information equality)**.** *Under Assumption* $\mathbb{L}_2$.Diff.θ_0*, for all statistics* $T : \mathcal{X} \to \mathbb{R}$ *such that* T *is locally bounded in* $\mathbb{L}_2(\mathbb{P}_\theta)$ *around* θ_0*, i.e., such that there exists an open neighborhood* U *of* θ_0 *with*

$$M_{U,T} \stackrel{\text{def}}{=} \sup_{\theta\in U} \, \mathbb{E}_\theta\big[T^2\big] < \infty\,,$$

the expectation function $\gamma_T : \theta \in U \longmapsto \gamma_T(\theta) = \mathbb{E}_\theta\big[T\big]$ *is well-defined and is differentiable at* θ_0*, with gradient*

$$\nabla\gamma_T(\theta_0) = 2\int_{\mathcal{X}} \xi_{\theta_0}\dot{\xi}_{\theta_0} T \, \mathrm{d}\mu\,.$$

The proof is extracted from Pollard [2001; 2005] and is provided in appendix for the sake of completeness.

3.2. A multidimensional information (in)equality

Via the choice of the statistic $T = 1$ a.s., one sees that for a model satisfying Assumption $\mathbb{L}_2$.Diff.θ_0,

$$\int_{\mathcal{X}} \xi_{\theta_0}\dot{\xi}_{\theta_0} \, \mathrm{d}\mu = \big[0\big]_{i\in\{1,\ldots,p\}}\,. \tag{4}$$

We now consider a vector-valued statistic $S : \mathcal{X} \to \mathbb{R}^s$, where $s \geqslant 1$ is an integer. The components of S are referred to as $(S_j)_{j\in\{1,\ldots,s\}}$. When the expectation function $\gamma_S : \theta \mapsto \gamma_S(\theta) = \mathbb{E}_\theta[S]$ is well-defined around θ_0 and is differentiable at $\theta_0 \in \Theta$, we denote, after an abuse of notation, by $\nabla\gamma_S(\theta_0)$

the concatenation of the s gradients of the functions $\theta \mapsto \mathbb{E}_\theta[S_j] = \gamma_{S_j}(\theta)$ at θ_0. That is, $\nabla\gamma_S(\theta_0)$ is a $p \times s$ matrix, whose element $(i,j) \in \{1,\ldots,p\} \times \{1,\ldots,s\}$ equals

$$\nabla\gamma_S(\theta_0)_{i,j} = \frac{\partial \gamma_{S_j}}{\partial \theta_i}(\theta_0)\,.$$

Lemma 1 can be generalized in a multidimensional fashion as follows; it suffices to apply it component-wise to each S_j, together with (4). For two vectors $u = (u_i)_{i\in\{1,\ldots,p\}}$ and $v = (v_j)_{j\in\{1,\ldots,s\}}$, we denote by $u \otimes v = uv^{\mathrm{T}} = [u_i v_j]_{(i,j)\in\{1,\ldots,p\}\times\{1,\ldots,s\}}$ their tensor product, which is a $p \times s$ matrix. (In the whole article, vectors are column-vectors.)

Lemma 2 (Multidimensional information equality)**.** *Under Assumption $\mathbb{L}_2$.Diff.θ_0, for all statistics $S : \mathcal{X} \to \mathbb{R}^s$ such that $\|S\|$ is locally bounded in $\mathbb{L}_2(\mathbb{P}_\theta)$ around θ_0, i.e., such that there exists an open neighborhood U of θ_0 with*

$$M_{U,S} \stackrel{\text{def}}{=} \sup_{\theta\in U} \mathbb{E}_\theta\big[\|S\|^2\big] < \infty\,,$$

the expectation function $\gamma_S : \theta \in U \longmapsto \gamma_S(\theta) = \mathbb{E}_\theta\big[S\big]$ is well-defined and is differentiable at θ_0, with gradient

$$\nabla\gamma_S(\theta_0) = 2\int_{\mathcal{X}} \xi_{\theta_0}\dot{\xi}_{\theta_0} \otimes S \,\mathrm{d}\mu = 2\int_{\mathcal{X}} \xi_{\theta_0}\dot{\xi}_{\theta_0} \otimes \big(S - \mathbb{E}_{\theta_0}[S]\big)\,\mathrm{d}\mu\,.$$

This information equality entails the following information inequality, known as the Cramér-Rao bound. We denote the fact that a symmetric matrix M is positive semi-definite by $M \succcurlyeq 0$.

Corollary 1 (Multidimensional information inequality; also known as the Cramér-Rao bound)**.** *Under the assumptions of Lemma 2, the following matrices are positive semi-definite,*

$$\begin{bmatrix} \mathrm{Var}_{\theta_0}(S) & \nabla\gamma_S(\theta_0)^{\mathrm{T}} \\ \nabla\gamma_S(\theta_0) & \mathcal{I}(\theta_0) \end{bmatrix} \succcurlyeq 0 \tag{5}$$

and thus, whenever $\mathcal{I}(\theta_0)$ is definite,

$$\mathrm{Var}_{\theta_0}(S) - \nabla\gamma_S(\theta_0)^{\mathrm{T}}\,\mathcal{I}(\theta_0)^{-1}\,\nabla\gamma_S(\theta_0) \succcurlyeq 0\,. \tag{6}$$

Proof. The proof is standard and follows, e.g., the exposition of Bobrovsky et al. [1987, Lemma 4] or Letac [2008]. The first matrix is seen to be a positive semi-definite matrix by a rewriting as an integral of such matrices, namely, thanks to Lemma 2 as far as the cross-products $\nabla\gamma_S(\theta_0)$ and $\nabla\gamma_S(\theta_0)^{\mathrm{T}}$ are concerned,

$$\begin{bmatrix} \mathrm{Var}_{\theta_0}(S) & \nabla\gamma_S(\theta_0)^{\mathrm{T}} \\ \nabla\gamma_S(\theta_0) & \mathcal{I}(\theta_0) \end{bmatrix} = \int \begin{bmatrix} \xi_{\theta_0}\big(S - \mathbb{E}_{\theta_0}[S]\big) \\ 2\dot{\xi}_{\theta_0} \end{bmatrix} \otimes \begin{bmatrix} \xi_{\theta_0}\big(S - \mathbb{E}_{\theta_0}[S]\big) \\ 2\dot{\xi}_{\theta_0} \end{bmatrix} \mathrm{d}\mu\,.$$

The second part of the corollary simply relies on the fact that the Schur complement of $\mathcal{I}(\theta_0)$ in the above matrix is positive semi-definite because the matrix itself is. (A proof of this well-known fact is recalled in Section A.3 in appendix, see Lemma 9.) □

3.3. A multidimensional Bayesian Cramér-Rao bound

We adapt here the exposition of Letac [2008] to our setting of assumed $\mathbb{L}_2(\mu)$–differentiability. We consider a model satisfying Assumption $\mathbb{L}_2$.Diff.θ_0 at all points $\theta_0 \in \Theta$.

Assumption ($\mathbb{L}_2$.Diff)**.** *The model $\big(\mathbb{P}_\theta\big)_{\theta\in\Theta}$, where $\Theta \subseteq \mathbb{R}^p$ is an open set, is dominated by μ and is differentiable in $\mathbb{L}_2(\mu)$ at all points of Θ.*

We also fix a statistic $S : \mathcal{X} \to R^s$ and now introduce a continuous density function $q : \Theta \to \mathbb{R}_+$ satisfying the following assumption.

Assumption (Int.q)**.** *The following integrals with respect to q are finite,*

$$\int_\Theta \mathbb{E}_\theta\big[\|S\|^2\big]\, q(\theta)\, \mathrm{d}\theta < \infty \qquad \textit{and} \qquad \int_\Theta \mathrm{Tr}\big(\mathcal{I}(\theta)\big)\, q(\theta)\, \mathrm{d}\theta < +\infty\,.$$

The so-called Bayesian Cramér-Rao bound will be discussed and compared to the van Trees inequality later on in this article (in Section 4.5).

Corollary 2 (Bayesian Cramér-Rao bound)**.** *Under Assumptions $\mathbb{L}_2$.Diff and Int.q, provided that $q : \Theta \to \mathbb{R}_+$ is continuous and that $\|S\|$ is locally bounded in $\mathbb{L}_2(\mathbb{P}_\theta)$ around each $\theta_0 \in \Theta$, we have*

$$\begin{bmatrix} \displaystyle\int_\Theta \mathrm{Var}_\theta(S)\, q(\theta)\, \mathrm{d}\theta & \displaystyle\left(\int_\Theta \nabla\gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta\right)^{\mathrm{T}} \\ \displaystyle\int_\Theta \nabla\gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta & \displaystyle\int_\Theta \mathcal{I}(\theta)\, q(\theta)\, \mathrm{d}\theta \end{bmatrix} \succcurlyeq 0\,,$$

where we used the notation of Lemma 2. In particular, whenever $\displaystyle\int_\Theta \mathcal{I}(\theta)\, q(\theta)\, \mathrm{d}\theta$ is definite,

$$\int_\Theta \mathrm{Var}_\theta(S)\, q(\theta)\, \mathrm{d}\theta - \left(\int_\Theta \nabla\gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta\right)^{\mathrm{T}} \left(\int_\Theta \mathcal{I}(\theta)\, q(\theta)\, \mathrm{d}\theta\right)^{-1} \left(\int_\Theta \nabla\gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta\right) \succcurlyeq 0\,.$$

Proof. The second part of the corollary follows again via the consideration of Schur complements, so that we may focus on its first part only. The hypotheses of Lemma 2 and Corollary 1 are satisfied at all $\theta_0 \in \Theta$:

$$\begin{bmatrix} \mathrm{Var}_{\theta_0}(S) & \nabla\gamma_S(\theta_0)^{\mathrm{T}} \\ \nabla\gamma_S(\theta_0) & \mathcal{I}(\theta_0) \end{bmatrix} \succcurlyeq 0\,, \tag{7}$$

where

$$\nabla\gamma_S(\theta_0) = 2\int_{\mathcal{X}} \xi_{\theta_0}\dot{\xi}_{\theta_0} \otimes \big(S - \mathbb{E}_{\theta_0}[S]\big)\, \mathrm{d}\mu\,.$$

Now,

$$(x, \theta) \longmapsto 2\xi_\theta(x)\dot{\xi}_\theta(x) \otimes \big(S(x) - \mathbb{E}_\theta[S]\big)\, q(\theta)$$

belongs to $\mathbb{L}_1(\mu \otimes \lambda)$, as can be seen by a Cauchy-Schwarz inequality together with Assumption Int.q. Therefore, $q\nabla\gamma_S$ is integrable with respect to the Lebesgue measure on on Θ. The same can be said for the functions $\theta \mapsto \mathrm{Var}_\theta(S)\, q(\theta)$ and $\theta \mapsto \mathcal{I}(\theta)\, q(\theta)$. We can therefore integrate the matrix bound (7) with respect to q over Θ. We get that

$$\begin{bmatrix} \displaystyle\int_\Theta \mathrm{Var}_\theta(S)\, q(\theta)\, \mathrm{d}\theta & \displaystyle\left(\int_\Theta \nabla\gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta\right)^{\mathrm{T}} \\ \displaystyle\int_\Theta \nabla\gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta & \displaystyle\int_\Theta \mathcal{I}(\theta)\, q(\theta)\, \mathrm{d}\theta \end{bmatrix} \succcurlyeq 0\,, \tag{8}$$

which concludes the proof. □

4. The van Trees inequality (under somewhat minimal assumptions)

In this section we state the sharpest version of the van Trees inequality we could obtain. Its direct proof will be deferred after a proof based on the Cramér-Rao inequality, that will require stronger assumptions.

The first series of assumptions is only intended to ensure that all quantities at hand in the statement of the inequality exist. On top of the above-stated Assumptions $\mathbb{L}_2$.Diff and Int.q, we will also consider the following ones.

Assumption (AC.q)**.** *The density function $q : \Theta \to \mathbb{R}$ is absolutely continuous, with* $\mathrm{Tr}(\mathcal{I}_q) < \infty$.

Assumption (Int.ψ)**.** *The absolutely continuous function $\psi : \Theta \to \mathbb{R}^s$ is such that the following integrals are finite,*

$$\int_\Theta \big\| \nabla\psi(\theta) \big\|_1 \, q(\theta) \, \mathrm{d}\theta < +\infty \qquad \textit{and} \qquad \int_\Theta \|\psi(\theta)\|^2 \, q(\theta) \, \mathrm{d}\theta < +\infty .$$

On top of these somewhat necessary[1] assumptions, a set of conditions allowing to integrate by parts is convenient. Several such sets are suitable; the first one (IP.border) is the one we worked out, a second one (IP.Stokes, discussed below) is the one used by Gill and Levit [1995] to apply Stokes' theorem. Note that Θ is not required to be bounded under Assumption IP.border.

Assumption (IP.border)**.** *The functions $\theta \mapsto q(\theta)$ and $\theta \mapsto q(\theta)\psi(\theta)$ tend respectively to* 0 *and* [0] *as θ approaches any point of the border of Θ with finite norm.*

Based on these assumptions, we can now state our main result.

Theorem 1 (The van Trees inequality)**.** *Under Assumptions* $\mathbb{L}_2$.Diff, AC.q, Int.q, *and* Int.ψ, *as well* IP.border*, the following matrix is well-defined and is positive semi-definite,*

$$\begin{bmatrix} \displaystyle\int_\Theta \mathbb{E}_\theta\Big[\big(S - \psi(\theta)\big) \otimes \big(S - \psi(\theta)\big)\Big] \, q(\theta) \, \mathrm{d}\theta & \displaystyle\left(\int_\Theta \nabla\psi(\theta) \, q(\theta) \, \mathrm{d}\theta\right)^{\mathrm{T}} \\ \displaystyle\int_\Theta \nabla\psi(\theta) \, q(\theta) \, \mathrm{d}\theta & \displaystyle\mathcal{I}_q + \int_\Theta \mathcal{I}(\theta) \, q(\theta) \, \mathrm{d}\theta \end{bmatrix} \succcurlyeq 0 . \tag{9}$$

Thus, whenever $\mathcal{I}_q + \displaystyle\int_\Theta \mathcal{I}(\theta) \, q(\theta) \, \mathrm{d}\theta$ is definite,

$$\int_\Theta \mathbb{E}_\theta\Big[\big(S - \psi(\theta)\big) \otimes \big(S - \psi(\theta)\big)\Big] \, q(\theta) \, \mathrm{d}\theta \\ - \left(\int_\Theta \nabla\psi(\theta) \, q(\theta) \, \mathrm{d}\theta\right)^{\mathrm{T}} \left(\mathcal{I}_q + \int_\Theta \mathcal{I}(\theta) \, q(\theta) \, \mathrm{d}\theta\right)^{-1} \left(\int_\Theta \nabla\psi(\theta) \, q(\theta) \, \mathrm{d}\theta\right) \succcurlyeq 0 .$$

Note that the above theorem is stronger and more general than the version obtained by Lenstra [2005], which is stated in the same context of models that are differentiable in $\mathbb{L}_2$.

[1] Section 4.3 explains why Assumption $\mathbb{L}_2$.Diff can be slightly weakened: it suffices that the model be coordinate-wise differentiable in $\mathbb{L}_2(\mu)$ at almost all points of $\mathrm{Supp}(q)$.

4.1. A self-improvement of the inequality

We explain why Theorem 1 self-improves to the following corollary. We realized that such a self-improvement was possible when we worked out the proof of the van Trees inequality based on the Cramér-Rao inequality; see Sections 5 and 7.1, and in particular, Footnote 3.

Corollary 3 (A strengthened van Trees inequality)**.** *Under Assumptions* $\mathbb{L}_2$.Diff*,* AC.q*,* Int.q*, and* Int.ψ*, as well* IP.border*, the following matrix is well-defined and is positive semi-definite,*

$$\begin{bmatrix} \displaystyle\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)\otimes\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta \\ \displaystyle\quad -\int_\Theta\big(\mathbb{E}_\theta[S]-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\otimes\int_\Theta\big(\mathbb{E}_\theta[S]-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta & \displaystyle\left(\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\right)^{\mathrm{T}} \\ \displaystyle\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta & \displaystyle\mathcal{I}_q+\int_\Theta \mathcal{I}(\theta)\,q(\theta)\,\mathrm{d}\theta \end{bmatrix} \succcurlyeq 0\,.$$

Thus, whenever $\mathcal{I}_q+\displaystyle\int_\Theta \mathcal{I}(\theta)\,q(\theta)\,\mathrm{d}\theta$ *is definite,*

$$\begin{aligned}&\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)\otimes\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta \\ &\qquad -\int_\Theta\big(\mathbb{E}_\theta[S]-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\otimes\int_\Theta\big(\mathbb{E}_\theta[S]-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta \\ &\qquad -\left(\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\right)^{\mathrm{T}}\left(\mathcal{I}_q+\int_\Theta \mathcal{I}(\theta)\,q(\theta)\,\mathrm{d}\theta\right)^{-1}\left(\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\right)\succcurlyeq 0\,.\end{aligned}$$

Proof. We note that if ψ satisfies the assumptions of Theorem 1, then so does $\psi+c$ for all constants $c\in\mathbb{R}^s$. In addition, as $\nabla(\psi+c)=\nabla\psi$, only the upper left submatrix in (9) is modified. What would the best choice be for c?

We denote by $\mathbb{M}$ the probability distribution over $\mathcal{X}\times\Theta$ with density $(x,\theta)\mapsto\xi_\theta(x)\,q(\theta)$ with respect to $\mu\otimes\lambda$. We also consider the function $J:(x,\theta)\in\mathcal{X}\times\Theta\mapsto S(x)-\psi(\theta)$. Assumption Int.$q$ shows that $J\in\mathbb{L}_2(\mathbb{M})$ and thus, via Jensen's inequality, that $J\in\mathbb{L}_1(\mathbb{M})$. We choose as c the expectation of J under $\mathbb{M}$, which can be rewritten, thanks to Fubini's theorem, as

$$c=\mathbb{E}_{\mathbb{M}}[J]=\int_\Theta\big(\mathbb{E}_\theta[S]-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\,.$$

We apply Theorem 1 with the above c and get an inequality (9) where the upper right submatrix is replaced by

$$\begin{aligned}&\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)-c\big)\otimes\big(S-\psi(\theta)-c\big)\Big]\,q(\theta)\,\mathrm{d}\theta \\ =\ &\mathbb{E}_{\mathbb{M}}\Big[\big(J-\mathbb{E}_{\mathbb{M}}[J]\big)\otimes\big(J-\mathbb{E}_{\mathbb{M}}[J]\big)\Big]\ =\ \mathbb{E}_{\mathbb{M}}\big[J\otimes J\big]-\mathbb{E}_{\mathbb{M}}[J]\otimes\mathbb{E}_{\mathbb{M}}[J] \\ =\ &\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)\otimes\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta-\int_\Theta\big(\mathbb{E}_\theta[S]-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\otimes\int_\Theta\big(\mathbb{E}_\theta[S]-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\,,\end{aligned}$$

where the equality in the middle of the second line corresponds to a bias-variance decomposition. This concludes the proof. □

4.2. Two cases of interest: $\Theta=\mathbb{R}$ and bounded intervals $\Theta=(a,b)$

For simplicity we restrict our attention to the case of the estimation of a scalar parameter, i.e., to the cases where $\Theta\subseteq\mathbb{R}$ and ψ is the identity function. We provide natural situations when the assumptions of Theorem 1 and Corollary 3 are satisfied.

Corollary 4. *Fix some model $\big(\mathbb{P}_\theta\big)_{\theta\in\mathbb{R}}$ that is dominated by μ and is differentiable in $\mathbb{L}_2(\mu)$ at all points of $\mathbb{R}$. Consider an absolutely continuous density function $q : \mathbb{R} \to \mathbb{R}_+$ such that $\int_{\mathbb{R}} \theta^2 q(\theta)\,\mathrm{d}\theta < \infty$. Then $\mathcal{I}_q > 0$ and for all statistics $T : \mathcal{X} \to \mathbb{R}$,*

$$\int_{\mathbb{R}} \mathbb{E}_\theta\big[(T-\theta)^2\big]\, q(\theta)\,\mathrm{d}\theta \geqslant \left(\int_{\mathbb{R}} \big(\mathbb{E}_\theta[T]-\theta\big)\, q(\theta)\,\mathrm{d}\theta\right)^2 + \frac{1}{\mathcal{I}_q + \displaystyle\int_{\mathbb{R}} \mathcal{I}(\theta)\, q(\theta)\,\mathrm{d}\theta}\,.$$

The proof of this corollary is straightforward; we only write it to show how certain requirements are immediate in the present case.

Proof. First, we note that all quantities appearing in the stated inequality are well-defined[2], even though some may be equal to $+\infty$. Also, $\mathcal{I}_q > 0$ as q cannot be constant over $\mathbb{R}$. Therefore, the stated inequality is trivial unless all the inequalities

$$\mathcal{I}_q < +\infty\,, \qquad \int_{\mathbb{R}} \mathcal{I}(\theta)\, q(\theta)\,\mathrm{d}\theta < +\infty\,, \qquad \text{and} \qquad \int_{\mathbb{R}} \mathbb{E}_\theta\big[(T-\theta)^2\big]\, q(\theta)\,\mathrm{d}\theta < +\infty$$

hold, which we can therefore safely assume for the rest of this proof. Since by the assumption on q we have that $\theta \mapsto \theta^2\, q(\theta)$ is Lebesgue-integrable over $\mathbb{R}$, we get that

$$\int_{\mathbb{R}} \mathbb{E}_\theta\big[T^2\big]\, q(\theta)\,\mathrm{d}\theta < +\infty\,.$$

To apply Corollary 3, which leads to the stated inequality, it only remains to see that Assumptions Int.ψ and IP.border are satisfied. The latter is void as the border is $\{-\infty, +\infty\}$. That Int.ψ is satisfied is because the target function ψ is the identity function over $\mathbb{R}$ (the first integral therein equals 1 and the second one is finite by the assumption on q). □

A similar proof leads to the following corollary in the case of a bounded interval.

Corollary 5. *Consider a parameter set $\Theta = (a,b)$ formed by a bounded open interval. Fix some model $\big(\mathbb{P}_\theta\big)_{\theta\in(a,b)}$ that is dominated by μ and is differentiable in $\mathbb{L}_2(\mu)$ at all points of (a,b). Consider an absolutely continuous density function $q : (a,b) \to \mathbb{R}_+$ such that $q(\theta) \to 0$ as $\theta \to a$ or $\theta \to b$. Then $\mathcal{I}_q > 0$ and for all statistics $T : \mathcal{X} \to \mathbb{R}$,*

$$\int_a^b \mathbb{E}_\theta\big[(T-\theta)^2\big]\, q(\theta)\,\mathrm{d}\theta \geqslant \left(\int_a^b \big(\mathbb{E}_\theta[T]-\theta\big)\, q(\theta)\,\mathrm{d}\theta\right)^2 + \frac{1}{\mathcal{I}_q + \displaystyle\int_a^b \mathcal{I}(\theta)\, q(\theta)\,\mathrm{d}\theta}\,.$$

4.3. Comparison to the classical version by Gill and Levit [1995]

The comparison focuses on the differences in the needed assumptions to get the multivariate matrix version (9) of the van Trees inequality. (How other multivariate matrix versions can be obtained is discussed in Sections 4.4 and A.4.) Before performing the mentioned comparison, we recall, for the sake of self-completeness, the setting and assumptions needed by Gill and Levit [1995].

[2]This is why no extra integrability assumption on $(x,\theta) \mapsto T^2(x)\, q(\theta)$ is needed here, unlike in the case of multi-dimensional statistics where the covariance terms did not necessarily exist.

The van Trees inequality of Gill and Levit [1995]. It is stated in a setting where the following (pointwise) assumptions are made on the model $(\mathbb{P}_\theta)_{\theta\in\Theta}$, replacing Assumption $\mathbb{L}_2$.Diff.

Assumption (AC.f_θ). *The density functions $f_\theta : \mathcal{X} \to \mathbb{R}_+$ are such that $(x, \theta) \in \mathcal{X} \times \Theta \mapsto f_\theta(x)$ is measurable and for μ–almost all x, the function $\theta \in \Theta \mapsto f_\theta(x)$ is absolutely continuous, with vector of partial derivatives at θ denoted by $\nabla f_\theta(x)$. Furthermore, $\nabla f_\theta \, \mathbb{I}_{\{f_\theta>0\}} \big/ \sqrt{f_\theta}$ belongs to $\mathbb{L}_2(\mu)$ for almost all $\theta \in \Theta$.*

Assumption (Int.∇f_θ). *For almost all $\theta \in \Theta$, the following integral is defined and is null,*

$$\int_{\mathcal{X}} \nabla f_\theta(x) \, \mathrm{d}\mu(x) = [0] \, .$$

Also, Assumption IP.border is to be replaced by the following (more restrictive) conditions.

Assumption (IP.Stokes). *The set $\Theta \subseteq \mathbb{R}^p$ is a compact set whose boundary $\partial\Theta$ is piecewise C^1–smooth. In addition, q is null on $\partial\Theta$.*

We now cite the main result of Gill and Levit [1995]. Assumptions AC.q, Int.q, and Int.ψ were not explicitly mentioned therein but were used implicitly (e.g., to ensure that the quantities at hand in the van Trees inequality indeed exist).

Theorem 2. *Consider a compact set $\Theta \subset \mathbb{R}^p$ and a function $q : \Theta \to \mathbb{R}$ satisfying Assumption* IP.Stokes*, as well as a statistical model $(\mathbb{P}_\theta)_{\theta\in\Theta}$ satisfying Assumptions* AC.f_θ *and* Int.∇f_θ*. Under Assumptions* AC.q, Int.q, *and* Int.ψ*, the van Trees inequality* (9) *holds.*

Overview of the comparison. Theorems 1 and 2 have some assumptions in common but they differ by resorting, respectively, to Assumptions $\mathbb{L}_2$.Diff and IP.border versus Assumptions AC.f_θ, Int.∇f_θ, and IP.Stokes.

The comparison is best understood after reading the proof of Theorem 1, which is provided in Section 7.2. The structures of the proof therein and of the one by Gill and Levit [1995] are similar: they both establish an information equality like the one reported in Lemma 4, from which (9) follows easily. The question is thus to get the information equality and differences arise in the process of doing so. The proof of the information equality at hand is decomposed into two main parts: the first part (see Section 7.2.1) reduces the problem to properly handling given integrals over Θ, which is performed in a second part (see Section 7.2.2) by integrating by parts. We group the differences according to which of these two parts they correspond.

Differences that are not differences. In our setting, Θ is an open set while in the setting of Gill and Levit [1995], this set is closed. However, IP.Stokes indicates that the border $\partial\Theta$ has a null probability mass under the distribution defined by q.

Differences when performing the integration by parts. We compare here IP.border and IP.Stokes: the latter is (much) more stringent than the former. (Under IP.Stokes, ψ is an absolutely continuous function defined on the compact set Θ, thus is bounded, so that $q\psi$ is null on the border of Θ.) Actually, the results stated in Lemma 7 and 8 are obtained, in the setting of Gill and Levit [1995] (see also Letac, 2008), by a straightforward application of Stokes' theorem (or Green's identity). The compactness of Θ and the regularity assumptions on its border are key to apply it legitimately. We showed on the contrary how more ad hoc arguments, based however, among others, on the use of Assumption $\mathbb{L}_2$.Diff (see Lemma 6 and its consequences), could avoid resorting to Stokes' theorem. More precisely, Assumption $\mathbb{L}_2$.Diff is used in a subtle way around (29). It is unclear whether our proof based on milder assumptions can be adapted to the setting of Gill and Levit [1995] since a typical issue in the setting with pointwise assumptions is to legitimately differentiate under the integral signs. This, probably, is not a mere detail but a sign that the Hájek–Le Cam setting is more convenient to work in. The next paragraph shows however that, surprisingly enough, under assumption Int.q, (a version of) the required assumption $\mathbb{L}_2$.Diff is a consequence of the pointwise assumption AC.f_θ.

Differences for the reduction to handling integrals over Θ. We compare here $\mathbb{L}_2$.Diff versus AC.f_θ and Int.∇f_θ. To do so, we consider the following relaxation of Assumption $\mathbb{L}_2$.diff, which is itself based on a relaxation of the notion of differentiability in $\mathbb{L}_2$. To state it we consider the canonical basis $(e_1, \ldots, e_p)$ of $\mathbb{R}^p$, that is, $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ where only the ith coordinate of e_i equals 1. Then, we restrict our attention to paths $\theta \to \theta_0$ along one of the p canonical directions.

Definition 3 (Coordinate-wise differentiability in $\mathbb{L}_2$)**.** *The dominated statistical model $(\mathbb{P}_\theta)_{\theta\in\Theta}$ is coordinate-wise differentiable in $\mathbb{L}_2(\mu)$ at $\theta_0 \in \Theta$ if for all $i \in \{1, \ldots, p\}$, there exists a scalar function $\dot{\xi}_{\theta_0,i} \in \mathbb{L}_2(\mu)$ such that*

$$\left\| \xi_{\theta_0+te_i} - \xi_{\theta_0} - t\,\dot{\xi}_{\theta_0,i} \right\|_\mu = o(t) \qquad \text{as} \quad t \to 0.$$

The vector-valued function $\dot{\xi}_{\theta_0} = \big(\dot{\xi}_{\theta_0,i}\big)_{i\in\{1,\ldots,p\}}$ is called the coordinate-wise derivative of the model at θ_0 and the definition of the Fisher information $\mathcal{I}(\theta_0)$ is as in (1).

We can now state the milder $\mathbb{L}_2$–type assumption needed on the model.

Assumption ($\mathbb{L}_2$.Diff.weak)**.** *The model $(\mathbb{P}_\theta)_{\theta\in\Theta}$, where $\Theta \subseteq \mathbb{R}^p$ is an open set, is dominated by μ and is coordinate-wise differentiable in $\mathbb{L}_2(\mu)$ at almost all points of $\Theta \cap \mathrm{Supp}(q)$.*

Now, we mention two facts, which we prove later on, in Section 7.3. (Actually, the same adaptation indicated in Fact 1 for the nullity stated in (4) shows that Assumption $\mathbb{L}_2$.Diff.weak entails Assumption Int.∇f_θ.)

Fact 1. *Theorem 1 holds with Assumption* $\mathbb{L}_2$.Diff *replaced by Assumption* $\mathbb{L}_2$.Diff.weak.

Fact 2. *Assumptions* Int.q *and* AC.f_θ *entail Assumption* $\mathbb{L}_2$.Diff.weak.

The combination of these two facts shows that the pointwise assumptions on the model needed by Gill and Levit [1995] are strictly stronger than the ones we require for our Hájek–Le Cam version of the van Trees inequality.

4.4. On the various multivariate formulations of the van Trees inequality

Gill and Levit [1995] mention that van Trees [1968] and Bobrovsky et al. [1987] report several different multivariate formulations of the van Trees inequality. In view of Section 5, which derives the van Trees inequality (9) as a Cramér-Rao bound for a well-chosen location model, we may call the version (9) the canonical form of the van Trees inequality. It is at least as canonical as (5) is in terms of the Cramér-Rao bound.

However, variations over the canonical form, as the ones presented by [Gill and Levit, 1995, Section 4], can be obtained in a straightforward manner. We detail this after the proof of Theorem 1, in Section A.4.

4.5. Comparison between the Bayesian Cramér-Rao bound and the van Trees inequality

We first recall the two bounds and put them into a common umbrella. Under the setting and assumptions of Theorem 1 and Corollary 3 and the additional assumption that $\int_\Theta \mathcal{I}(\theta)\, q(\theta)\, \mathrm{d}\theta$ is definite, the function $\gamma_S : \theta \longmapsto \gamma_S(\theta) = \mathbb{E}_\theta[S]$ is well-defined and differentiable at least on $\mathrm{Supp}(q)$ and the following two inequalities hold: the van Trees inequality of Corollary 3,

$$\begin{aligned}\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)\otimes\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta \qquad\qquad &\qquad\qquad\qquad\qquad\qquad\qquad \text{(vT)}\\ -\int_\Theta \big(\gamma_S(\theta)-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta \otimes \int_\Theta \big(\gamma_S(\theta)-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta &\\ -\left(\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\right)^{\mathrm{T}} \left(\mathcal{I}_q + \int_\Theta \mathcal{I}(\theta)\,q(\theta)\,\mathrm{d}\theta\right)^{-1} \left(\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\right) \succcurlyeq 0\,, &\end{aligned}$$

and the Bayesian Cramér-Rao bound of Corollary 2,

$$\begin{aligned}\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)\otimes\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta \qquad\qquad &\qquad\qquad\qquad\qquad\qquad\qquad \text{(BCR)}\\ -\int_\Theta \big(\gamma_S(\theta)-\psi(\theta)\big)\otimes\big(\gamma_S(\theta)-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta &\\ -\left(\int_\Theta \nabla\gamma_S(\theta)\,q(\theta)\,\mathrm{d}\theta\right)^{\mathrm{T}} \left(\int_\Theta \mathcal{I}(\theta)\,q(\theta)\,\mathrm{d}\theta\right)^{-1} \left(\int_\Theta \nabla\gamma_S(\theta)\,q(\theta)\,\mathrm{d}\theta\right) \succcurlyeq 0\,, &\end{aligned}$$

where we used a bias-variance decomposition to get the same left-most term as in (vT).

Part of the following discussion can be found in Letac [2008].

(BCR) **is better than** (vT) **when** S **is unbiased for the estimation of** $\psi(\theta)$**.** In the case when $\gamma_S = \psi$, the middle terms disappear in both equations and the right-most terms can be easily compared, the one in (BCR) being the largest. In the case of biased estimators, when $\gamma_S \neq \psi$, which of two bounds is better than the other strongly depends on S, ψ, q, and the model.

For general estimators S**, only** (vT) **is useful.** The third term in (BCR) is not intrinsic enough in the case of a biased estimator and it is difficult to rely on it to issue efficiency statements for possibly biased estimators. (The second terms in (BCR) and (vT) are not intrinsic either but they can be omitted without much loss.) This is in strong contract with the van Trees inequality (vT), which is stated in terms of the goal $\psi(\theta)$ only. We illustrate in Section 6 how efficiency results can be obtained for general, not necessarily unbiased, estimators.

(BCR) **is better than a version of** (vT) **when** q **is the uniform density over a bounded domain.** We mostly discuss this case for the records. Reading in details the direct proof of the van Trees inequality, one can see, based on inequality (24) of Lemma 5 that (vT) holds with the factors

$$\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta$$

in its right-most term replaced by

$$-\int_\Theta \nabla q(\theta)\otimes\psi(\theta)\,\mathrm{d}\theta + \int_\Theta \nabla q(\theta)\otimes\gamma_S(\theta)\,\mathrm{d}\theta + \int_\Theta \nabla\gamma_S(\theta)\,q(\theta)\,\mathrm{d}\theta\,,$$

even when Assumption IP.border is not satisfied. The alternative expression reduces to

$$\int_\Theta \nabla\gamma_S(\theta)\,q(\theta)\,\mathrm{d}\theta$$

in the case when Θ is bounded and q is the uniform density over Θ. Note that in this case we also have that $\mathcal{I}_q$ is a null matrix, $\mathcal{I}_q = [0]$. Our claim follows however, because, by Jensen's inequality,

$$\int_\Theta \big(\gamma_S(\theta)-\psi(\theta)\big)\otimes\big(\gamma_S(\theta)-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta - \int_\Theta \big(\gamma_S(\theta)-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta \otimes \int_\Theta \big(\gamma_S(\theta)-\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta \succcurlyeq 0\,,$$

and hence, (BCR) is the sharpest inequality.

5. The van Trees inequality as a Cramér-Rao bound for a location model

We explain in this section that the van Trees inequality is exactly a Cramér-Rao bound for a well-chosen location model given by mixture distributions. This observation however requires somewhat stronger assumptions than the ones stated in the previous section to obtain a direct proof of the van Trees inequality. Even worse, the set of assumptions we could work out involve some pointwise regularity on the mappings $\theta \mapsto f_\theta(x)$, for μ–almost all x, and can therefore be considered of a totally different nature (more related to the classical regularity assumptions and less related to Le Cam's viewpoint on statistics).

We keep Assumptions $\mathbb{L}_2$.Diff, AC.q, and Int.q as they are. We strengthen Assumptions Int.ψ and IP.border into assumptions referred to as Int$^+$.ψ and IP$^+$.border, and also consider the assumption referred to as AC.f_θ in Section 4.3. We do not claim that this set of assumptions is minimal to interpret the van Trees inequality as a Cramér-Rao bound: we merely tried to work out a convenient and realistic enough such set.

Assumption (IP$^+$.border)**.** *There exists $\delta > 0$ such that, denoting by $\Theta^{(-\delta)}$ the open set of the elements in Θ that are at least δ far from its border, then* $\mathrm{Supp}(q)$ *is included in $\Theta^{(-\delta)}$; that is, for all $\theta \in \Theta$ with $q(\theta) > 0$, for all $\alpha \in \mathbb{R}^d$ with $\|\alpha\| \leqslant \delta$, one has $\theta + \alpha \in \Theta$.*

Assumption (Int$^+$.ψ)**.** *There exists an open neighborhood U of* $[0]$ *which is contained in the $\delta/2$–open ball around* $[0]$ *such that*

$$\int_{\Theta^{(-\delta)}} \left(\sup_{\alpha \in U} \big\| \nabla\psi(\theta+\alpha) \big\| \right) q(\theta)\,\mathrm{d}\theta < +\infty \qquad \textit{and} \qquad \sup_{\alpha \in U} \int_{\Theta^{(-\delta)}} \big\| \psi(\theta+\alpha) \big\|^2 q(\theta)\,\mathrm{d}\theta < +\infty\,.$$

Assumption IP$^+$.border is key for defining the location model below. This assumption implies Assumption IP.border, which only asserts that $q(\theta)$ and $q(\theta)\psi(\theta)$ should tend to 0 as θ approaches a point of the border of Θ with finite norm. Indeed, under Assumption IP$^+$.border, we have that q is null on a δ–open ball around such a border point with finite norm (and thus so is $q\psi$). Of course, Assumption IP$^+$.border does not exclude the case when the $\mathrm{Supp}(q) = \Theta = \mathbb{R}^p$ as it only puts more severe restrictions around border points with finite norms.

5.1. Construction of the location model and statement of the result

The family of (mixture) distributions at hand is indexed by $\alpha \in \mathbb{R}^p$ with $\|\alpha\| < \delta/2$. The distribution $\mathbb{M}_\alpha$ is over $\mathcal{X} \times \Theta^{(-\delta/2)}$ and is defined as the probability distribution with density

$$m_\alpha : \ (x,\theta) \in \mathcal{X} \times \Theta^{(-\delta/2)} \ \longmapsto \ m_\alpha(x,\theta) = f_{\theta-\alpha}(x)\, q(\theta-\alpha)$$

with respect to the product measure $\mu \otimes \lambda$. (Because of Assumption IP$^+$.border, this indeed defines the density of a probability distribution.) The model $(\mathbb{M}_\alpha)_{\alpha:\|\alpha\|<\delta/2}$ is thus a location model.

Given a statistic $S : \mathcal{X} \to \mathbb{R}^s$ for the $(\mathbb{P}_\theta)_{\theta\in\Theta}$ model and an absolutely continuous target function $\psi : \Theta \to \mathbb{R}^s$, we construct the statistic $J : (x,\theta) \mapsto S(x) - \psi(\theta)$ for the $(\mathbb{M}_\alpha)_{\alpha:\|\alpha\|<\delta/2}$ model. (Note that this statistic J was already introduced in the proof of Corollary 3, which also considered a distribution $\mathbb{M}$ corresponding to the above-defined distribution $\mathbb{M}_{[0]}$.)

Proposition 1. *Under Assumptions* $\mathbb{L}_2$.Diff, AC.q, Int.q, Int$^+$.ψ, IP$^+$.border, *and* AC.f_θ*, the Cramér-Rao bound* (6) *holds for the model $(\mathbb{M}_\alpha)_{\alpha:\|\alpha\|<\delta/2}$ and the statistic J at $\alpha_0 = [0]$ and is given by the van Trees inequality* (9).

5.2. Discussion and comparison to the hypotheses of Theorem 1

As explained above, the set of conditions for Proposition 1 is strictly stronger than the one for Theorem 1. We discuss here how severe or how mild the three additional (sets of) requirements are. We recall that we anyway only need them for the interpretation stated in Proposition 1; the van Trees inequality (Theorem 1) holds under the weaker sets of assumptions.

Int.ψ $\rightarrow$ Int$^+$.ψ: Whenever ψ is a Lipschitz function, these two sets of assumptions are actually equivalent. Indeed, the integrals involving $\nabla\psi$ are bounded by the Lipschitz constant L of ψ. As for the two other integrals to be compared, we note that in this case,

$$\left\| \psi(\theta+\alpha) \right\|^2 \leqslant \Big(\left\| \psi(\theta) \right\| + L\|\alpha\| \Big)^2 \leqslant 2 \left\| \psi(\theta) \right\|^2 + 2L^2(\delta/2)^2 .$$

Since a main case of interest is the identity function, $\psi(\theta) = \theta$, the stronger local requirements in Assumption Int$^+$.ψ can be considered not too dramatic.

IP.border $\rightarrow$ IP$^+$.border: The stronger condition IP$^+$.border on the support of q, which, as already indicated above, supersedes IP.border, could probably be circumvented by regularization (by replacing any q satisfying IP.border with some q_δ satisfying IP$^+$.border and letting $\delta \rightarrow 0$).

On the addition of AC.f_θ: The main and most stringent new constraint is given by the pointwise regularity of the density functions stated in Assumption AC.f_θ, which is most unappreciated in an à la Le Cam viewpoint. The latter would only consider assumptions on the density functions that involve integrated values, and avoid any pointwise restrictions.

6. Application to local asymptotic minimax (LAM) lower bounds

Gill and Levit [1995] provide several applications of the van Trees inequality; of course, all of them hold for our version of the inequality in view of the discussion in Section 4.3. We recall below one such application, a derivation of an efficiency bound, namely, an asymptotic Cramér-Rao bound. We provide yet another application to efficiency bounds: a simple derivation of local asymptotic minimax (LAM) lower bounds for the square risk in parametric models (simple enough to be used as a classroom material). More general (for other risk functions) and stronger such bounds are provided by van der Vaart [1998, Section 8.7] but their proof requires sophisticated arguments, which is in contrast with our proof below.

The setting is that of a sequence $(X_n)_{n\geqslant 1}$ of independent random variables taking values in $\mathcal{X}$, and thus of a sequence of statistical models $\bigl((\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta}\bigr)_{n\geqslant 1}$. In the sequel, we index by ${}^{\otimes n}$ all quantities (expectations, Fisher information, etc.) relative to the product model $(\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta}$.

We assume that the base statistical model $(\mathbb{P}_\theta)_{\theta\in\Theta}$ satisfies Assumption $\mathbb{L}_2$.Diff. Direct calculations then show that each product model $(\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta}$ also satisfies Assumption $\mathbb{L}_2$.Diff, with derivative at $\theta_0 \in \Theta$ equal to

$$(x_1,\ldots,x_n) \in \mathcal{X}^n \longmapsto \sum_{k=1}^{n} \left(\dot{\xi}_{\theta_0}(x_k) \prod_{k'\neq k} \xi_{\theta_0}(x_{k'}) \right);$$

In particular, the Fisher information of the product model at θ_0 equals

$$\mathcal{I}^{\otimes n}(\theta_0) = n\,\mathcal{I}(\theta_0)\,. \tag{10}$$

We consider an absolutely continuous function $\psi : \Theta \to \mathbb{R}^s$. We will assume that there exists some $\theta_0 \in \Theta$ such that the following regularity conditions are met. This point θ_0 will be the point of interest in the LAM bound.

Assumption (Reg.θ_0). *The functions $\theta \mapsto \mathcal{I}(\theta)$ and $\theta \mapsto \nabla\psi(\theta)$ are continuous at θ_0; moreover, the matrix $\mathcal{I}(\theta_0)$ is definite.*

6.1. The asymptotic Cramér-Rao bound of Gill and Levit [1995]

The bound of [Gill and Levit, 1995, Section 3] is adapted in our context as follows; we state it merely for the sake of completeness. (We omit its proof as it follows exactly the exposition in the mentioned reference; it bears some resemblance with the proof of Theorem 4.) It imposes a restriction on the sequences of statistics that are studied.

Definition 4. *A sequence $(S_n)_{n\geqslant 1}$ of statistics $S_n : \mathcal{X}^n \to \mathbb{R}^s$ is Hájek regular at θ_0 for the estimation of ψ if there exists a probability distribution $\mathcal{L}_{\psi,\theta_0}$ over $\mathbb{R}^s$ such for all $h \in \mathbb{R}^p$, we have the following convergences in distribution:*

$$\sqrt{n}\Bigl(S_n - \psi\Bigl(\theta_0 + \tfrac{h}{\sqrt{n}}\Bigr)\Bigr) \rightsquigarrow \mathcal{L}_{\psi,\theta_0} \qquad \text{under the sequence } \ \mathbb{P}_{\theta_0+h/\sqrt{n}}\,.$$

Theorem 3. *Consider a model satisfying Assumption* $\mathbb{L}_2$.Diff *and some absolutely continuous function $\psi : \Theta \to \mathbb{R}^s$ such that Assumption* Reg.θ_0 *is met for some $\theta_0 \in \Theta$. Then, for all sequences $(S_n)_{n\geqslant 1}$ of statistics $S_n : \mathcal{X}^n \to \mathbb{R}^s$ that are Hájek regular at θ_0 for the estimation of ψ with a limit distribution $\mathcal{L}_{\psi,\theta_0}$ admitting a second-order moment, the variance of this limit distribution satisfies*

$$\mathrm{Var}\bigl(\mathcal{L}_{\psi,\theta_0}\bigr) - \nabla\psi(\theta_0)^{\mathrm{T}}\,\mathcal{I}(\theta_0)^{-1}\,\nabla\psi(\theta_0) \succcurlyeq 0\,.$$

6.2. Our local asymptotic minimax (LAM) lower bound

We propose the following lower bound, which does not impose any restriction on the sequence of statistics at hand. In its statement, we define, for all normal distributions $\mathcal{N}(\mu,\Gamma)$ over $\mathbb{R}^s$,

$$\int_{\mathbb{R}^s} \ell \, \mathrm{d}\mathcal{N}(\mu,\Gamma) = \int_{\mathbb{R}^s} \ell(v)\, q_{\mathcal{N}(\mu,\Gamma)}(v)\, \mathrm{d}v\,,$$

where $q_{\mathcal{N}(\mu,\Gamma)}(v)$ denotes the density of the $\mathcal{N}(\mu,\Gamma)$ distribution.

Theorem 4. *Consider a model satisfying Assumption* $\mathbb{L}_2$.Diff *and some absolutely continuous function* $\psi : \Theta \to \mathbb{R}^s$ *such that Assumption* Reg.θ_0 *is met for some* $\theta_0 \in \Theta$*. Then, for all sequences* $(S_n)_{n \geqslant 1}$ *of statistics* $S_n : \mathcal{X}^n \to \mathbb{R}^s$ *and for all positive quadratic forms* $\ell : \mathbb{R}^s \to \mathbb{R}_+$,

$$\liminf_{c\to+\infty} \liminf_{n\to+\infty} \sup_{\|h\|<1} \mathbb{E}^{\otimes n}_{\theta_0 + \frac{ch}{\sqrt{n}}} \left[\ell\left(\sqrt{n}\left(S_n - \psi\left(\theta_0 + \tfrac{ch}{\sqrt{n}}\right)\right)\right)\right] \geqslant \int_{\mathbb{R}^s} \ell \, \mathrm{d}\mathcal{N}\Big([0],\ \nabla\psi(\theta_0)^{\mathrm{T}} \mathcal{I}(\theta_0)^{-1}\, \nabla\psi(\theta_0)\Big)\,.$$

Proof. We consider an auxiliary random variable $H \in \mathbb{R}^p$ the support of whose distribution is given by the (open) unit ball $\mathcal{B}_p\big([0],1\big)$ of $\mathbb{R}^p$; we choose it such that its distribution is given by a bounded and absolutely continuous density function $q : \mathbb{R}^p \to \mathbb{R}$, with $q = 0$ outside $\mathcal{B}_p\big([0],1\big)$. We also assume that $\mathcal{I}_q$ is definite and that $\mathrm{Tr}\big(\mathcal{I}_q\big) < +\infty$.

We fix some $c > 0$ and some $\varepsilon > 0$. There exists $n(c,\varepsilon)$ such that for all $n \geqslant n(c,\varepsilon)$, first, the open ball $\mathcal{B}_p\big(\theta_0,\, c/\sqrt{n}\big)$ centered at θ_0 and with radius $c/\sqrt{n}$ is contained in Θ, and second, for all $\theta \in \mathcal{B}_p\big(\theta_0,\, c/\sqrt{n}\big)$,

$$\max\Big\{ \big|\psi(\theta) - \psi(\theta_0)\big|,\ \big\|\nabla\psi(\theta) - \nabla\psi(\theta_0)\big\|,\ \big|\mathrm{Tr}(\mathcal{I}(\theta)) - \mathrm{Tr}(\mathcal{I}(\theta_0))\big| \Big\} \leqslant \varepsilon\,. \tag{11}$$

The above inequalities stem from the continuity of ψ, $\nabla\psi$, and $\mathcal{I}$ at θ_0.

We denote by q_n the density of the distribution of the random variable $H_n = \theta_0 + cH/\sqrt{n}$. This density q_n satisfies Assumption AC.q, and $\mathcal{I}_{q_n} = (n/c^2)\,\mathcal{I}_q$ is definite.

We consider some vector $U \in \mathbb{R}^s$ and will apply Theorem 1 to the model $\big(\mathbb{P}_\theta^{\otimes n}\big)_{\theta\in\Theta}$, the density q_n, the real-valued statistic $U^{\mathrm{T}} S_n$, and the real-valued target function $U^{\mathrm{T}}\psi$. Assumptions Int.ψ and IP.border are satisfied, because of (11) and because of the support of q_n. It only remains to see whether Assumption Int.q holds. Its second part does, again by (11) and in view of the support of q_n. However, its first part may well not be satisfied. In the case where it is satisfied, we are all set to apply Theorem 1 and we get, as $\mathcal{I}_{q_n} = (n/c^2)\,\mathcal{I}_q$ is definite and in view of (10),

$$\int_{\mathcal{B}_p([0],1)} \mathbb{E}^{\otimes n}_{\theta_0 + ch/\sqrt{n}} \left[\left(U^{\mathrm{T}} \left(S_n - \psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\right)\right)^2 \right] q(h)\,\mathrm{d}h \tag{12}$$

$$\geqslant \left(\int_{\mathcal{B}_p([0],1)} \nabla\psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big) U\, q(h)\,\mathrm{d}h \right)^{\mathrm{T}} \left(\frac{n}{c^2}\,\mathcal{I}_q + n \int_{\mathcal{B}_p([0],1)} \mathcal{I}\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\, q(h)\,\mathrm{d}h \right)^{-1} \times \left(\int_{\mathcal{B}_p([0],1)} \nabla\psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big) U\, q(h)\,\mathrm{d}h \right).$$

(Because of our conventions in terms of gradients of vector-valued functions, we note that the gradient of $U^{\mathrm{T}}\psi$ is actually given by $\nabla\psi\, U$.) In the case where the first part of Assumption Int.q is not satisfied, we have, because its second part is satisfied on the contrary, that

$$\int_{\mathcal{B}_p([0],1)} \mathbb{E}^{\otimes n}_{\theta_0+ch/\sqrt{n}}\Big[\big(U^{\mathrm{T}} S_n\big)^2\Big]\, q(h)\,\mathrm{d}h = +\infty\,,$$

$$\text{and thus that} \qquad \int_{\mathcal{B}_p([0],1)} \mathbb{E}^{\otimes n}_{\theta_0+ch/\sqrt{n}}\left[\left(U^{\mathrm{T}}\left(S_n - \psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\right)\right)^2\right] q(h)\,\mathrm{d}h = +\infty\,;$$

therefore, inequality (12) is satisfied as well, in a trivial way, as its left-hand side equals $+\infty$ while its right-hand side is a finite nonnegative number.

Now, recall that any positive quadratic form $\ell : \mathbb{R}^s \to \mathbb{R}_+$ can be decomposed as follows: there exists an orthogonal basis $U_1,\ldots,U_s$ of $\mathbb{R}^s$ and nonnegative real numbers $\lambda_1,\ldots,\lambda_s \geqslant 0$ such that for all $v \in \mathbb{R}^s$,

$$\ell(v) = \sum_{k=1}^{s} \lambda_k \big(U_k^{\mathrm{T}}\, v\big)^2 = \sum_{k=1}^{s} \lambda_k\, U_k^{\mathrm{T}}\, v\, v^{\mathrm{T}}\, U_k\,.$$

Note that in particular, for all $s \times s$ symmetric positive matrices Γ,

$$\int_{\mathbb{R}^s} v\, v^{\mathrm{T}}\, \mathrm{d}\mathcal{N}\big([0],\Gamma\big)(v) = \Gamma \qquad \text{so that} \qquad \int_{\mathbb{R}^s} \ell(v)\, \mathrm{d}\mathcal{N}\big([0],\Gamma\big)(v) = \sum_{k=1}^{s} \lambda_k\, U_k^{\mathrm{T}}\, \Gamma\, U_k\,.$$

Linear combinations according to the $\lambda_k \geqslant 0$ of versions of (12) for the U_k thus show that for all positive quadratic forms $\ell : \mathbb{R}^s \to \mathbb{R}_+$,

$$\int_{\mathcal{B}_p([0],1)} \mathbb{E}^{\otimes n}_{\theta_0+\frac{ch}{\sqrt{n}}}\left[\ell\left(S_n - \psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\right)\right] q(h)\,\mathrm{d}h \geqslant \int_{\mathbb{R}^s} \ell\, \mathrm{d}\mathcal{N}\big([0],\Gamma_{c,n}/n\big)\,,$$

or, given that for all $a \in R$ and all $v \in \mathbb{R}^s$, one has $\ell(av) = a^2\,\ell(v)$,

$$\int_{\mathcal{B}_p([0],1)} \mathbb{E}^{\otimes n}_{\theta_0+\frac{ch}{\sqrt{n}}}\left[\ell\left(\sqrt{n}\left(S_n - \psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\right)\right)\right] q(h)\,\mathrm{d}h \geqslant \int_{\mathbb{R}^s} \ell\, \mathrm{d}\mathcal{N}\big([0],\Gamma_{c,n}\big)\,, \qquad (13)$$

where

$$\Gamma_{c,n} = \left(\int_{\mathcal{B}_p([0],1)} \nabla\psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\, q(h)\,\mathrm{d}h\right)^{\mathrm{T}} \left(\frac{1}{c^2}\,\mathcal{I}_q + \int_{\mathcal{B}_p([0],1)} \mathcal{I}\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\, q(h)\,\mathrm{d}h\right)^{-1} \times \left(\int_{\mathcal{B}_p([0],1)} \nabla\psi\Big(\theta_0 + \tfrac{ch}{\sqrt{n}}\Big)\, q(h)\,\mathrm{d}h\right).$$

But Assumption Reg.θ_0 was precisely stated for the following convergence to take place,

$$\lim_{c\to+\infty}\ \lim_{n\to+\infty}\ \Gamma_{c,n} \ = \ \nabla\psi(\theta_0)^{\mathrm{T}}\, \mathcal{I}(\theta_0)^{-1}\, \nabla\psi(\theta_0)\,.$$

This convergence, the linearity of the right-hand side of (13) in $\Gamma_{c,n}$, and upper bounding the left-hand side integral of (13) by the supremum over $\mathcal{B}_p([0],1)$ of its integrand conclude the proof. □

Remark 1. *As a corollary, a similar result can be obtained for semi-parametric models $\mathcal{P}$ (details omitted; see van der Vaart, 1998, Chapter 25 for a description of the setting and the definitions of the objects mentioned in the next sentences). More precisely, the goal is to estimate functions ψ of the law $P \in \mathcal{P}$ of the model. We assume that ψ is differentiable at some P_0 relatively to a tangent cone $\mathcal{T}_{P_0}$ with efficient influence function $\widetilde{\psi}_{P_0}$. Now, at least when $\mathcal{T}_{P_0}$ is a linear vector space, the local asymptotic minimax risk at P_0, as measured by a positive quadratic form ℓ, is lower bounded by the integral of ℓ against the centered normal distribution with variance equal to the variance of the efficient influence function $\widetilde{\psi}_{P_0}$ under P_0.*

7. Proofs of the van Trees inequality

We provide two proofs, one for Proposition 1, which relies on stronger assumptions and resorts to the Cramér-Rao bound, and one for Theorem 1, which is more direct. The second one can best be appreciated with the first one in mind.

7.1. Proof of Proposition 1

We first show that the hypotheses of Lemma 2 and Corollary 1 are satisfied; we then instantiate their results and explain why they can be identified with the van Trees inequality.

Hypothesis 1: Local boundedness of J. The statistic J is locally bounded in $\mathbb{L}_2(\mathbb{M}_\alpha)$ around $[0] \in \mathbb{R}^p$; this is because for all $\alpha \in \mathbb{R}^p$ with $\|\alpha\| < \delta/2$,

$$\begin{aligned} \mathbb{E}_{\mathbb{M}_\alpha}\big[\|J\|^2\big] &= \int_{\mathcal{X}\times\Theta^{(-\delta/2)}} \big\| S(x) - \psi(\theta) \big\|^2 f_{\theta-\alpha}(x)\, q(\theta-\alpha)\, \mathrm{d}\mu(x)\mathrm{d}\theta \\ &\leqslant \int_{\mathcal{X}\times\Theta^{(-\delta/2)}} 2\Big(\big\| S(x) \big\|^2 + \big\| \psi(\theta) \big\|^2 \Big) f_{\theta-\alpha}(x)\, q(\theta-\alpha)\, \mathrm{d}\mu(x)\mathrm{d}\theta \\ &= 2\int_\Theta \mathbb{E}_\theta\big[\|S\|^2\big]\, q(\theta)\,\mathrm{d}\theta + 2\int_{\Theta^{(-\delta)}} \big\| \psi(\theta+\alpha) \big\|^2 q(\theta)\,\mathrm{d}\theta\,, \end{aligned} \tag{14}$$

from which Assumptions Int.q and Int$^+$.ψ yield the claim. The last equality is by Fubini's theorem for nonnegative integrands and by noting that, after a change of variable from $\theta-\alpha$ to θ, we are left with integrating over $-\alpha + \Theta^{(-\delta/2)}$, which is the same as integrating over the larger set Θ or the smaller set $\Theta^{(-\delta)}$ in view of the terms $q(\theta)$.

Hypothesis 2: Differentiability of the model $(\mathbb{M}_\alpha)_{\alpha:\|\alpha\|<\delta/2}$ — first part. This model is a location model. Its differentiability follows from an application of Proposition 2 in appendix. Assumptions AC.f_θ and AC.q indeed guarantee that the hypotheses (i)–(iii) of the mentioned proposition are satisfied, where α, $[0]$, (x,θ), and ν play respectively the role of β, β_0, y, and ν. Thus, the candidate for the $\mathbb{L}_2(\mu)$–derivative of the model at $\alpha \in \mathbb{R}^p$, which we denote by ζ_α, is given by

$$\begin{aligned} \zeta_\alpha : (x,\theta) \longmapsto\; & \frac{1}{2}\frac{\nabla m_\alpha(x,\theta)}{\sqrt{m_\alpha(x,\theta)}}\mathbb{I}_{\{m_\alpha(x,\theta)>0\}} \\ &= -\frac{1}{2}\left(\frac{f_{\theta-\alpha}(x)\,\nabla q(\theta-\alpha)}{\sqrt{f_{\theta-\alpha}(x)\,q(\theta-\alpha)}} + \frac{q(\theta-\alpha)\,\nabla f_{\theta-\alpha}(x)}{\sqrt{f_{\theta-\alpha}(x)\,q(\theta-\alpha)}} \right) \mathbb{I}_{\{f_{\theta-\alpha}(x)\,q(\theta-\alpha)>0\}} \\ &= -\left(\frac{1}{2}\frac{\nabla q(\theta-\alpha)}{\sqrt{q(\theta-\alpha)}}\,\mathbb{I}_{\{q(\theta-\alpha)>0\}} \underbrace{\sqrt{f_{\theta-\alpha}(x)}}_{=\,\xi_{\theta-\alpha}(x)} + \sqrt{q(\theta-\alpha)}\; \underbrace{\frac{1}{2}\frac{\nabla f_{\theta-\alpha}(x)}{\sqrt{f_{\theta-\alpha}(x)}}\,\mathbb{I}_{\{f_{\theta-\alpha}(x)>0\}}}_{=\,\dot{\xi}_{\theta-\alpha}(x)} \right). \end{aligned}$$

In the last equality a $\dot{\xi}_{\theta-\alpha}(x)$ term is identified: this follows from the lemma below. We still have to check the hypothesis (iv) of Proposition 2, which we do in the next paragraph.

Lemma 3. *Under Assumptions* $\mathbb{L}_2$.Diff *and* AC.f_θ*, for all* $\theta \in \Theta$*, we have,* μ*–almost everywhere,*

$$\dot{\xi}_\theta = \frac{1}{2}\frac{\nabla f_\theta}{\sqrt{f_\theta}}\,\mathbb{I}_{\{f_\theta>0\}}\,.$$

Proof. The proof is adapted from Pollard [2001; 2005] and is provided mostly for the sake of completeness. Recall that by Fatou's lemma, the $\mathbb{L}^2(\mu)$ and μ–almost-everywhere limits of a sequence of functions, when they both exist, coincide. We fix $\theta \in \Theta$ and will consider all directions $h \in \mathbb{R}^d$ around

θ; the quantities that follow will all be defined at least for $t > 0$ small enough. First, by definition of $\mathbb{L}_2(\mu)$–differentiability, as $t \to 0$ with $t > 0$,

$$\frac{1}{t}\big(\xi_{\theta+th} - \xi_\theta\big) \longrightarrow h^{\mathrm{T}}\,\dot{\xi}_\theta \qquad \text{in } \mathbb{L}^2(\mu). \tag{15}$$

Second, on $\{f_\theta > 0\}$, we have, by Assumption AC.f_θ, as $t \to 0$ with $t > 0$,

$$\frac{1}{t}\big(\xi_{\theta+th} - \xi_\theta\big)\mathbb{I}_{\{f_\theta>0\}} = \frac{1}{t}\Big(\sqrt{f_{\theta+th}} - \sqrt{f_\theta}\Big)\mathbb{I}_{\{f_\theta>0\}} \longrightarrow \frac{1}{2}h^{\mathrm{T}}\,\frac{\nabla f_\theta}{\sqrt{f_\theta}}\,\mathbb{I}_{\{f_\theta>0\}} \qquad \mu\text{–almost everywhere.}$$

The equality stated in the lemma thus holds on $\{f_\theta > 0\}$. To conclude the proof, it only remains to show that μ–almost everywhere on $\{f_\theta = 0\}$, we have $\dot{\xi}_\theta = [0]$. Indeed, for fixed θ and h, we have by (15) and Fatou's lemma that there exists a sequence $t_n \to 0$ such that

$$\frac{1}{t_n}\xi_{\theta+t_n h}\,\mathbb{I}_{\{f_\theta=0\}} \longrightarrow h^{\mathrm{T}}\,\dot{\xi}_\theta\,\mathbb{I}_{\{f_\theta=0\}} \qquad \mu\text{–almost everywhere.}$$

In particular, since the left-hand side elements are nonnegative,

$$h^{\mathrm{T}}\,\dot{\xi}_\theta\,\mathbb{I}_{\{f_\theta=0\}} \geqslant 0\,;$$

since this is true for all directions $h \in \mathbb{R}^d$, this implies that μ–almost everywhere, $\dot{\xi}_\theta\,\mathbb{I}_{\{f_\theta=0\}} = [0]$, which concludes the proof. □

Hypothesis 2, continued: Differentiability of the model $(\mathbb{M}_\alpha)_{\alpha:\|\alpha\|<\delta/2}$ — second part. We now check the hypothesis (iv) of Proposition 2. Since the derivatives ζ_α all have the same (possibly infinite) square integral with respect to $\mu \otimes \lambda$, we only need to show that this common square integral is finite. (We use again here the fact that integrating over $-\alpha + \Theta^{(-\delta/2)}$ is the same as integrating over Θ in view of the terms involving q.)

For the sake of concise notation (and for later purposes), we introduce the function $\Delta : \mathcal{X}\times\Theta \to \mathbb{R}^p$ defined by

$$\Delta(x,\theta) = \frac{1}{2}\frac{\nabla q(\theta)}{\sqrt{q(\theta)}}\,\mathbb{I}_{\{q(\theta)>0\}}\,\xi_\theta(x) + \sqrt{q(\theta)}\;\dot{\xi}_\theta(x) \tag{16}$$

for all $x \in \mathcal{X}$ and $\theta \in \Theta$, so that the derivative $\zeta_{[0]}$ of the model at $\alpha = [0]$ is given by $-\Delta$. That $\Delta \in \mathbb{L}_2(\mu\otimes\lambda)$ follows from the fact that each of its summands is in $\mathbb{L}_2(\mu\otimes\lambda)$ by Assumption Int.q (and the use of Fubini's theorem for nonnegative integrands).

We thus have shown that under the stated assumptions, the $(\mathbb{M}_\alpha)_{\alpha:\|\alpha\|<\delta/2}$ model is differentiable in $\mathbb{L}_2(\mu\otimes\lambda)$ at all points $\alpha \in \mathbb{R}^p$ with $\|\alpha\| < \delta/2$, with derivative at $[0]$ given by $-\Delta$. We denote by $\mathcal{I}_{\mathbb{M}}([0])$ its Fisher information at this point.

Exploiting the Cramér–Rao bound. The hypotheses of Lemma 2 and Corollary 1 are then satisfied and the following facts hold. The mapping

$$\Gamma : \alpha \in U \;\longmapsto\; \Gamma(\alpha) = \mathbb{E}_{\mathbb{M}_\alpha}\big[J\big]$$

is differentiable at $[0]$ and the following matrix is positive semi-definite,

$$\begin{bmatrix} \mathrm{Var}_{\mathbb{M}_{[0]}}(J) & \nabla\Gamma\big([0]\big)^{\mathrm{T}} \\ \nabla\Gamma\big([0]\big) & \mathcal{I}_{\mathbb{M}}\big([0]\big) \end{bmatrix} \succcurlyeq 0\,.$$

Thus, it is also the case for the following matrix, where we denote by Id_p and Id_s the $p\times p$ and $s\times s$ identity matrices and where the empty cells are for null submatrices:

$$\begin{bmatrix} \mathrm{Id}_s & \\ & -\mathrm{Id}_p \end{bmatrix} \begin{bmatrix} \mathrm{Var}_{\mathbb{M}_{[0]}}(J) & \nabla\Gamma\big([0]\big)^{\mathrm{T}} \\ \nabla\Gamma\big([0]\big) & \mathcal{I}_{\mathbb{M}}\big([0]\big) \end{bmatrix} \begin{bmatrix} \mathrm{Id}_s & \\ & -\mathrm{Id}_p \end{bmatrix} = \begin{bmatrix} \mathrm{Var}_{\mathbb{M}_{[0]}}(J) & -\nabla\Gamma\big([0]\big)^{\mathrm{T}} \\ -\nabla\Gamma\big([0]\big) & \mathcal{I}_{\mathbb{M}}\big([0]\big) \end{bmatrix} \succcurlyeq 0\,. \tag{17}$$

We now identify or bound the three quantities at hand: we will show that

$$\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)\otimes\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta \;-\; \mathrm{Var}_{\mathbb{M}_{[0]}}(J) \succcurlyeq 0\,, \tag{18}$$

$$\int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta \;=\; -\nabla\Gamma\big([0]\big)\,, \tag{19}$$

$$\mathcal{I}_{\mathbb{M}}\big([0]\big) \;=\; \mathcal{I}_q + \int_\Theta \mathcal{I}(\theta)\,q(\theta)\,\mathrm{d}\theta\,, \tag{20}$$

from which it will follow from (17) that the van Trees inequality (9) is satisfied.

Identifying the three quantities at hand. The quantities in the difference (18) exist by (14). Actually, the proof of Corollary 3 already[3] showed that the desired inequality (18) holds true.

As for the gradient (19), we have by arguments similar to the ones used to establish (14) (and in particular, the use of Fubini's theorem based on some integrability following from the second part of $\mathrm{Int}^+.\psi$ and the Cauchy-Schwarz inequality), that for all $\alpha\in U$,

$$\Gamma(\alpha)=\mathbb{E}_{\mathbb{M}_\alpha}\big[J\big] = \int_{\mathcal{X}\times\Theta^{(-\delta/2)}}\big(S(x)-\psi(\theta)\big)\,f_{\theta-\alpha}(x)\,q(\theta-\alpha)\,\mathrm{d}\mu(x)\mathrm{d}\theta$$
$$= \int_\Theta \mathbb{E}_\theta[S]\,q(\theta)\,\mathrm{d}\theta - \int_{\Theta^{(-\delta)}}\psi(\theta+\alpha)\,q(\theta)\,\mathrm{d}\theta\,.$$

Provided that we can differentiate under the integral sign at $\alpha=[0]$, we get

$$\nabla\Gamma\big([0]\big) = -\int_{\Theta^{(-\delta)}}\nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\,;$$

this is the desired equality (19) as integrating over $\Theta^{(-\delta)}$ or Θ is the same here, because of the $q(\theta)$ term. The stated differentiability under the integral sign follows, e.g., from the absolute continuity of ψ and from the first part of Assumption $\mathrm{Int}^+.\psi$.

The derivative of the model at $[0]$ being $-\Delta$, we have by definition

$$\mathcal{I}_{\mathbb{M}}\big([0]\big) = 4\int_{\mathcal{X}\times\Theta}\Delta\otimes\Delta\,\mathrm{d}\mu\mathrm{d}\lambda\,.$$

Now, we already mentioned above that by Assumption $\mathrm{Int}.q$ the two summands in the defining equation (16) of 2Δ are square integrable. Their cross product is therefore integrable. At (x,θ) it equals

$$\nabla q(\theta)\mathbb{I}_{\{q(\theta)>0\}}\,\xi_\theta(x)\dot{\xi}_\theta(x) = \nabla q(\theta)\,\xi_\theta(x)\dot{\xi}_\theta(x)\,.$$

This is because $\nabla q=[0]$ at all points of $\{q=0\}$ where it is defined (where q is differentiable), since 0 is a global minimum of q when it is achieved. Because of (4), the integral of this cross product with respect to $\mathrm{d}\mu\mathrm{d}\lambda$ is null. Therefore (all exchanges of orders in integration below being valid thanks to Fubini's theorem),

$$\begin{aligned}
&4\int_{\mathcal{X}\times\Theta}\big(\Delta(x,\theta)\otimes\Delta(x,\theta)\big)\,\mathrm{d}\mu(x)\mathrm{d}\theta\\
&=\int_{\mathcal{X}\times\Theta}\frac{\nabla q(\theta)\nabla q(\theta)^{\mathrm{T}}}{q(\theta)}\,\mathbb{I}_{\{q(\theta)>0\}}\,f_\theta(x)\,\mathrm{d}\mu(x)\mathrm{d}\theta + \int_{\mathcal{X}\times\Theta}4\,\dot{\xi}_\theta(x)\dot{\xi}_\theta(x)^{\mathrm{T}}\,q(\theta)\,\mathrm{d}\theta\mathrm{d}\mu(x)\\
&=\mathcal{I}_q+\int_\Theta\mathcal{I}(\theta)\,q(\theta)\,\mathrm{d}\theta\,,
\end{aligned}\tag{21}$$

which is (20).

[3] For the records: It is indeed the fact that we got $\mathrm{Var}_{\mathbb{M}_{[0]}}(J)$ in (17) that led us from Theorem 1 to the self-improvement stated as Corollary 3.

7.2. Proof of Theorem 1

The proof consists of proving directly what Lemma 2 and Corollary 1 yield in the location model considered in the previous subsection. More precisely, we will prove in a direct way the following result (but will not prove that $-\Delta$ is the derivative of some model at some point).

Lemma 4 (The van Trees version of the information equality)**.** *If the assumptions of Theorem 1 are satisfied for some statistic $S : \mathcal{X} \to \mathbb{R}^s$ and some absolutely continuous function $\psi : \Theta \to \mathbb{R}^s$, then the following integrals are defined and are equal,*

$$2 \int_{\mathcal{X}\times\Theta} \Big(\Delta(x,\theta) \otimes \big(S(x) - \psi(\theta)\big) \Big) \xi_\theta(x) \sqrt{q(\theta)}\, \mathrm{d}\mu(x)\mathrm{d}\theta = \int_\Theta \nabla\psi(\theta)\, q(\theta)\, \mathrm{d}\theta\,.$$

We now explain why Lemma 4 entails Theorem 1. We use on top of it (21), which follows from Assumptions AC.q, Int.q, and $\mathbb{L}_2$.Diff, and the fact that the following integrals exist and are equal,

$$\int_{\mathcal{X}\times\Theta} \big(S(x) - \psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)} \otimes \big(S(x) - \psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)}\mathrm{d}\mu(x)\mathrm{d}\theta$$
$$= \int_\Theta \mathbb{E}_\theta\Big[\big(S - \psi(\theta)\big) \otimes \big(S - \psi(\theta)\big) \Big]\, q(\theta)\, \mathrm{d}\theta\,,$$

which stems from the first part of Int.q and the second part of Int.ψ thanks to Fubini's theorem. All in all, we get that the following integrals are well-defined and that the matrix of interest in Theorem 1 is well-defined and is positive semi-definite,

$$\left[\begin{array}{cc} \displaystyle\int_\Theta \mathbb{E}_\theta\Big[\big(S - \psi(\theta)\big) \otimes \big(S - \psi(\theta)\big)\Big]\, q(\theta)\, \mathrm{d}\theta & \displaystyle\left(\int_\Theta \nabla\psi(\theta)\, q(\theta)\, \mathrm{d}\theta\right)^{\mathrm{T}} \\ \displaystyle\int_\Theta \nabla\psi(\theta)\, q(\theta)\, \mathrm{d}\theta & \displaystyle\mathcal{I}_q + \int_\Theta \mathcal{I}(\theta)\, q(\theta)\, \mathrm{d}\theta \end{array} \right]$$
$$= \int_{\mathcal{X}\times\Theta} \left[\begin{array}{c} \big(S(x) - \psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)} \\ 2\Delta(x,\theta) \end{array} \right] \otimes \left[\begin{array}{c} \big(S(x) - \psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)} \\ 2\Delta(x,\theta) \end{array} \right] \mathrm{d}\mu(x)\mathrm{d}\theta \;\succcurlyeq 0\,. \qquad (22)$$

The second part of Theorem 1 follows by considering Schur complements (see Lemma 9 in Section A.3 of the appendix). Thus, it only remains to prove the information equality stated as Lemma 4. The proof is split into three parts: the first one only relies on the assumptions necessary for the quantities at hand in the van Trees inequality to exist. It leads to some information equality, with a not so natural integral over Θ on the right-hand side. The second part of the proof is a matter of analysis and uses the additional hypothesis IP.border to integrate by parts the integral, so as to get the desired and more intrinsic integral at hand in the van Trees inequality. However, these first and second part only deal with bounded statistics. The third part of the proof provides the extension to general statistics.

Remark 2. *A careful reading of the proof shows that it is the separation of the variables in $J(x,\theta) = S(x) - \psi(\theta)$ that is crucial to relax the hypotheses needed to derive the van Trees inequality.*

7.2.1. Some information equality (to be further improved)

Lemma 5. *We consider a model, some statistic $S : \mathcal{X} \to \mathbb{R}^s$, and some absolutely continuous function $\psi : \Theta \to \mathbb{R}^s$ such that Assumptions* $\mathbb{L}_2$.Diff, AC.q, Int.q, *and* Int.ψ *are satisfied.*

Then, there exists a function $\Psi_S \in \mathbb{L}_1(\mu \otimes \lambda)$ such that Ψ_S is a non-decreasing function of $\|S\|$ and for all $x \in \mathcal{X}$ and $\theta \in \Theta$,

$$\left\| \Big(\Delta(x,\theta) \otimes \big(S(x) - \psi(\theta)\big) \Big) \xi_\theta(x) \sqrt{q(\theta)} \right\|_1 \leqslant \Psi_S(x,\theta)\,. \qquad (23)$$

Moreover, if S is bounded, then the function $\gamma_S : \theta \in \Theta \mapsto \mathbb{E}_\theta[S]$ is well-defined and differentiable and the following integrals are defined and are equal,

$$2\int_{\mathcal{X}\times\Theta} \Big(\Delta(x,\theta) \otimes \big(S(x) - \psi(\theta)\big)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta \\ = -\int_\Theta \nabla q(\theta) \otimes \psi(\theta)\,\mathrm{d}\theta + \int_\Theta \nabla q(\theta)\otimes\gamma_S(\theta)\,\mathrm{d}\theta + \int_\Theta \nabla\gamma_S(\theta)\,q(\theta)\,\mathrm{d}\theta\,. \quad (24)$$

What we mean here by the monotonicity property of Ψ_S in terms of $\|S\|$ is that if two statistics S and S' are such that $\|S\| \leqslant \|S'\|$, then $\Psi_S \leqslant \Psi_{S'}$.

Proof. We already indicated above that $\Delta \in \mathbb{L}_2(\mu\otimes\lambda)$ thanks to Assumption Int.q; similarly, Assumptions Int.q and Int.ψ respectively show that the functions

$$(x,\theta) \longmapsto S(x)\,\xi_\theta(x)\sqrt{q(\theta)} \qquad \text{and} \qquad (x,\theta) \longmapsto \psi(\theta)\,\xi_\theta(x)\sqrt{q(\theta)}$$

belong to $\mathbb{L}_2(\mu\otimes\lambda)$. Thus, by the Cauchy-Schwarz inequality, the integrand can be decomposed as the sum of four integrable functions,

$$\begin{aligned} &2\Big(\Delta(x,\theta)\otimes\big(S(x)-\psi(\theta)\big)\Big)\xi_\theta(x)\sqrt{q(\theta)} \\ &= f_\theta(x)\,\mathbb{I}_{\{q(\theta)>0\}}\nabla q(\theta)\otimes S(x) + 2\,\xi_\theta(x)q(\theta)\,\dot\xi_\theta(x)\otimes S(x) \\ &\quad - f_\theta(x)\,\mathbb{I}_{\{q(\theta)>0\}}\nabla q(\theta)\otimes\psi(\theta) - 2\,\xi_\theta(x)q(\theta)\,\dot\xi_\theta(x)\otimes\psi(\theta)\,. \end{aligned} \quad (25)$$

More precisely, given that $\|u\otimes v\|_1 \leqslant \sqrt{ps}\,\|u\|\|v\|$ for $u\in\mathbb{R}^p$ and $v\in\mathbb{R}^s$, we have the stated domination (23) by the integrable function

$$\Psi_S : (x,\theta)\in\mathcal{X}\times\Theta \longmapsto \sqrt{ps}\Big(\xi_\theta(x)\,\mathbb{I}_{\{q(\theta)>0\}}\,\big\|\nabla q(\theta)\big\| + 2\,\big\|\dot\xi_\theta(x)\big\|\,q(\theta)\Big)\;\xi_\theta(x)\Big(\big\|S(x)\big\| + \big\|\psi(\theta)\big\|\Big).$$

We note that Ψ_S indeed satisfies the stated monotonicity in $\|S\|$.

Actually, the indicator functions can be safely omitted in the display (25), since, as proved right before (21), we have $\mathbb{I}_{\{q(\theta)>0\}}\nabla q(\theta) = \nabla q(\theta)$ for almost all $\theta\in\Theta$.

As for the second part of the lemma, since each of the four integrands in (25) is integrable, we may, by Fubini's theorem, integrate first over x and then over θ. Using that $x\mapsto f_\theta(x)$ integrates to 1 and that by (4) the integral of $x\mapsto\xi_\theta(x)\dot\xi_\theta(x)$ is null, we finally get

$$\begin{aligned} &2\int_{\mathcal{X}\times\Theta}\Big(\Delta(x,\theta)\otimes\big(S(x)-\psi(\theta)\big)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta \\ &= \int_\Theta \nabla q(\theta)\otimes\Big(\int_{\mathcal{X}} S(x)\,f_\theta(x)\,\mathrm{d}\mu(x)\Big)\,\mathrm{d}\theta + 2\int_\Theta\Big(\int_{\mathcal{X}}\xi_\theta(x)\dot\xi_\theta(x)\otimes S(x)\mathrm{d}\mu(x)\Big)\,q(\theta)\,\mathrm{d}\theta \\ &\quad -\int_\Theta \nabla q(\theta)\otimes\psi(\theta)\,\mathrm{d}\theta\,. \end{aligned}$$

The boundedness of S (i.e., $\|S\|\leqslant K$ a.s. for some K) and Assumption $\mathbb{L}_2$.Diff show that the hypotheses of Lemma 2 are satisfied at all $\theta\in\Theta$. Therefore, the function $\gamma_S : \theta\in\Theta\mapsto\mathbb{E}_\theta[S]$ is well-defined and differentiable at each $\theta\in\Theta$, with gradient

$$\nabla\gamma_S(\theta) = 2\int_{\mathcal{X}}\xi_\theta\dot\xi_\theta\otimes S\,\mathrm{d}\mu\,.$$

We get

$$\int_\Theta \nabla q(\theta) \otimes \left(\int_{\mathcal{X}} S(x)\, f_\theta(x)\, \mathrm{d}\mu(x) \right) \mathrm{d}\theta + 2 \int_\Theta \left(\int_{\mathcal{X}} \xi_\theta(x) \dot{\xi}_\theta(x) \otimes S(x) \mathrm{d}\mu(x) \right) q(\theta)\, \mathrm{d}\theta$$
$$= \int_\Theta \nabla q(\theta) \otimes \gamma_S(\theta)\, \mathrm{d}\theta + \int_\Theta \nabla \gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta\,,$$

which concludes the proof. □

7.2.2. Some analysis to rewrite the integrals at hand

The rewriting of the integrals at hand in the left-hand side of (24) will rely on the following ad hoc version of Stokes' theorem in our context.

Lemma 6. *We consider an open domain $\Theta \subseteq \mathbb{R}^p$ and an absolutely continuous function $\varphi : \Theta \to \mathbb{R}$ such that*

$$\int_\Theta |\varphi|\, \mathrm{d}\lambda < +\infty \qquad \textit{and} \qquad \int_\Theta \|\nabla \varphi\|_1\, \mathrm{d}\lambda < +\infty \tag{26}$$

and that $\varphi(\theta)$ tends to 0 as θ approaches any point of the border of Θ with finite norm. Then

$$\int_\Theta \nabla \varphi\, \mathrm{d}\lambda = [0]\,.$$

Based on this result, we will prove the following two lemmas.

Lemma 7. *If Assumptions* AC.q*,* Int.ψ*, and* IP.border *are satisfied for some absolutely continuous function $\psi : \Theta \to \mathbb{R}^s$, then the following integrals are defined and are equal,*

$$-\int_\Theta \nabla q(\theta) \otimes \psi(\theta)\, \mathrm{d}\theta = \int_\Theta \nabla \psi(\theta)\, q(\theta)\, \mathrm{d}\theta\,.$$

Lemma 8. *If the assumptions of Lemma 5 as well as Assumption* IP.border *are satisfied for some bounded statistic $S : \mathcal{X} \to \mathbb{R}^s$, then the function $\gamma_S : \theta \mapsto \mathbb{E}_\theta[S]$ is well-defined and differentiable over Θ and the following integrals are defined and cancel out to the null matrix,*

$$\int_\Theta \nabla q(\theta) \otimes \gamma_S(\theta)\, \mathrm{d}\theta + \int_\Theta \nabla \gamma_S(\theta)\, q(\theta)\, \mathrm{d}\theta = [0]\,.$$

These lemmas are proved below, in the following order: first, Lemmas 7 and 8, and then, Lemma 6.

Proof of Lemma 7. We prove the matrix equality column-wise. For all $j \in \{1, \ldots, s\}$, we denote by ψ_j the jth component of ψ and will apply Lemma 6 to $\psi_j q$ to get

$$-\int_\Theta \psi_j(\theta) \nabla q(\theta)\, \mathrm{d}\theta = \int_\Theta \nabla \psi_j(\theta)\, q(\theta)\, \mathrm{d}\theta\,.$$

Indeed, $\varphi = \psi_j q$ is absolutely continuous on Θ, as a product of absolutely continuous functions. The behavior on the border is taken care of by Assumption IP.border. Thus, we need only to check hypothesis (26). That $\psi_j q \in \mathbb{L}_1(\lambda)$ follows from the second part of Assumption Int.ψ via the Cauchy-Schwarz inequality. It remains to check the integrability of

$$\nabla(\psi_j q) = q \nabla \psi_j + \psi_j \nabla q = q \nabla \psi_j + \psi_j \nabla q\, \mathbb{I}_{\{q>0\}}\,,$$

where the last equality is by a fact used several times already, e.g., around (25): that $\nabla q = [0]$ almost surely on $\{q = 0\}$. Now, the integrability of $q \nabla \psi_j$ is stated by Assumption Int.ψ. As for the one of $\psi_j \nabla q \mathbb{I}_{\{q>0\}}$, it follows from yet another application of the Cauchy-Schwarz inequality,

$$\int_\Theta \left\| \psi_j \nabla q\, \mathbb{I}_{\{q>0\}} \right\|_1 \mathrm{d}\lambda \leqslant \sqrt{p} \sqrt{\int_\Theta \psi_j^2\, q\, \mathrm{d}\lambda}\, \sqrt{\mathrm{Tr}(\mathcal{I}_q)} < +\infty\,, \tag{27}$$

where the claimed finiteness is by the second part of Assumption Int.ψ and AC.q. □

Proof of Lemma 8. We again prove the matrix equality column-wise. For all $j \in \{1, \ldots, s\}$ and $\theta \in \Theta$, we denote by S_j and $\gamma_{S_j}(\theta)$ the jth components of S and $\gamma_S(\theta)$. We will apply Lemma 6 to $\gamma_{S_j} q$ to get the following equality to the null column vector of $\mathbb{R}^p$,

$$\int_\Theta \gamma_{S_j}(\theta) \nabla q(\theta)\,\mathrm{d}\theta + \int_\Theta \nabla\gamma_{S_j}(\theta)\, q(\theta)\,\mathrm{d}\theta = [0]\,. \tag{28}$$

By assumption, S_j is bounded by some $K > 0$, that is, $|S_j| \leqslant K$ a.s.; in this case, Lemma 1 shows that $\gamma_{S_j} : \theta \in \Theta \mapsto \mathbb{E}_\theta[S_j]$ is well-defined and differentiable over Θ. Since q is absolutely continuous and γ_{S_j} is bounded, $\gamma_{S_j} q$ is also absolutel To be corrected! hus is almost everywhere differentiable, with gradient given by $q \nabla\gamma_{S_j} + \gamma_{S_j} \nabla q$.

The boundedness of γ_{S_j} and Assumption IP.border show that $\gamma_{S_j} q$ also satisfies the required vanishing-at-the-finite-border condition. The integrability of $\gamma_{S_j} q$ follows from the boundedness of γ_{S_j} and the integrability of q (it integrates to 1).

To legitimately apply Lemma 6 and get the desired result (28), it only remains to show that the gradient $q\nabla\gamma_{S_j} + \gamma_{S_j}\nabla q$ is integrable. Since γ_{S_j} is bounded by K, replacing ψ_j by K in (27) shows that $\gamma_{S_j}\nabla q$ is integrable. As for $q\nabla\gamma_{S_j}$, we first note from Lemma 1 that for all $\theta \in \Theta$,

$$\nabla\gamma_{S_j}(\theta) = 2\int_{\mathcal{X}} \xi_\theta \dot{\xi}_\theta S_j \,\mathrm{d}\mu\,, \tag{29}$$

so that, by a Cauchy-Schwarz inequality,

$$q(\theta)\,\big\|\nabla\gamma_{S_j}(\theta)\big\|_1 \leqslant q(\theta) K \int_{\mathcal{X}} 2\xi_\theta \big\|\dot{\xi}_\theta\big\|_1 \,\mathrm{d}\mu \leqslant q(\theta) K\sqrt{p}\,\sqrt{\mathrm{Tr}\big(\mathcal{I}(\theta)\big)}\,,$$

where the latter upper bound is seen to be integrable over Θ by Jensen's inequality for $\sqrt{\cdot}$ and Assumption Int.q. □

Proof of Lemma 6. Recall the notation θ_{-i} and θ_i from Definition 2. For all $i \in \{1, \ldots, p\}$, we denote by

$$\Theta_{-i} = \big\{\theta_{-i} : \exists\theta_i \in \mathbb{R} \text{ such that } (\theta_i, \theta_{-i}) \in \Theta\big\}$$

the projection of Θ ignoring the ith coordinates. With each $\theta_{-i} \in \Theta_{-i}$, we associate the set

$$\Theta(\theta_{-i}) = \big\{\theta_i : (\theta_i, \theta_{-i}) \in \Theta\big\}$$

of ith coordinates that complete θ_{-i} in an element of Θ. (This piece of notation was actually already considered in Definition 2.) Since Θ is an open domain, $\Theta(\theta_{-i})$ is an open subset of $\mathbb{R}$. It can thus be written as an (at most) countable disjoint union of open intervals,

$$\Theta(\theta_{-i}) = \bigsqcup_{n \geqslant 1} \big(a_n(\theta_{-i}),\, b_n(\theta_{-i})\big)$$

where $a_n(\theta_{-i}) \in \mathbb{R} \cup \{-\infty\}$ and $b_n(\theta_{-i}) \in \mathbb{R} \cup \{+\infty\}$; at most one of the $a_n(\theta_{-i})$, respectively, of the $b_n(\theta_{-i})$, may equal $-\infty$, respectively, $+\infty$.

Almost all $\theta_{-i} \in \Theta_{-i}$ are such that for all $n \geqslant 1$, the following facts hold true:

$$\int_{a_n(\theta_{-i})}^{b_n(\theta_{-i})} \left|\frac{\partial\varphi}{\partial\theta_i}(\theta_i, \theta_{-i})\right| \mathrm{d}\theta_i < +\infty\,, \tag{30}$$

$$\int_{a_n(\theta_{-i})}^{b_n(\theta_{-i})} \big|\varphi(\theta_i, \theta_{-i})\big| \mathrm{d}\theta_i < +\infty\,, \tag{31}$$

and for all real numbers $a > a_n(\theta_{-i})$ and $b < b_n(\theta_{-i})$,

$$\int_a^b \frac{\partial \varphi}{\partial \theta_i}(\theta_i, \theta_{-i})\,\mathrm{d}\theta_i = \varphi(b, \theta_{-i}) - \varphi(a, \theta_{-i})\,. \tag{32}$$

This is, respectively, because of Fubini's theorem together with the fact that $\partial\varphi/\partial\theta_i$ and φ are integrable, as assumed in (26), and of the absolute continuity of φ.

We now show that for these $\theta_{-i} \in \Theta_{-i}$, for all $n \geqslant 1$,

$$\int_{a_n(\theta_{-i})}^{b_n(\theta_{-i})} \frac{\partial \varphi}{\partial \theta_i}(\theta_i, \theta_{-i})\,\mathrm{d}\theta_i = 0\,. \tag{33}$$

To do so, we would like to let $a \to a_n(\theta_{-i})$ and $b \to b_n(\theta_{-i})$ in (32). That by assumption $\varphi(\theta)$ tends to 0 as θ approaches any point of the border of Θ with finite norm means exactly that $\varphi(a, \theta_{-i}) \to 0$ and $\varphi(b, \theta_{-i}) \to 0$ except maybe in the cases where $a_n(\theta_{-i}) = -\infty$ or $b_n(\theta_{-i}) = +\infty$. In the latter cases, we proceed as follows. By symmetry we write the argument only for the case where $b_n(\theta_{-i}) = +\infty$. We fix some $c_n(\theta_{-i}) > a_n(\theta_{-i})$. By (32), for all $b \in \big(c(\theta_{-i}),\, +\infty\big)$,

$$\varphi(b, \theta_{-i}) = \varphi\big(c(\theta_{-i}),\, \theta_{-i}\big) + \int_{c(\theta_{-i})}^{b} \frac{\partial \varphi}{\partial \theta_i}(\theta_i, \theta_{-i})\mathrm{d}\theta_i\,.$$

We thus get from (30) that $\varphi(b, \theta_{-i})$ has a limit $\ell(\theta_{-i})$ as $b \to +\infty$. Since by (31),

$$\int_{c(\theta_{-i})}^{+\infty} \big|\varphi(\theta_i, \theta_{-i})\big|\,\mathrm{d}\theta_i < \infty\,,$$

one necessarily has $\ell(\theta_{-i}) = 0$.

The proof can now be concluded. Summing (33) over n, we get

$$\int_{\Theta(\theta_{-i})} \frac{\partial \varphi}{\partial \theta_i}(\theta_i, \theta_{-i})\,\mathrm{d}\theta_i = 0\,.$$

Integrating this equality over $\theta_{-i} \in \Theta_{-i}$ and repeating the argument for all i leads to

$$\int_\Theta \nabla\varphi(\theta)\,\mathrm{d}\theta = [0]\,,$$

as claimed □

7.2.3. From bounded statistics to general statistics

Lemmas 5, 7, and 8 yield Lemma 4 and thus Theorem 1 in the case of bounded statistics, $\|S\| \leqslant K$ a.s. for some $K > 0$. The extension to general statistics is obtained by the dominated convergence theorem.

We indeed denote the integrand in Lemma 5 by

$$\mathrm{VT}_S : \mathcal{X} \times \Theta \longmapsto \Big(\Delta(x,\theta) \otimes \big(S(x) - \psi(\theta)\big)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,.$$

We know that it is dominated by an integrable function denoted by Ψ_S. From a general, not necessarily bounded, statistic S, we form thresholded versions of it at some level $K > 0$,

$$S_K = S\,\mathbb{I}_{\big\{\|S\| \in [-K,K]\big\}}\,.$$

We already proved that Lemma 4 was valid for the bounded S_K statistics, so that for all $K > 0$,

$$2\int_{\mathcal{X}\times\Theta} \mathrm{VT}_{S_K}(x,\theta)\,\mathrm{d}\mu(x)\mathrm{d}\theta = \int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\,.$$

The right-hand side is independent of K, while the integrands of the left-hand side converge in a pointwise manner to $\mathrm{VT}(x,\theta)$ as $K \to +\infty$ and, by Lemma 5, are each dominated by Ψ_{S_K} and thus are all dominated by Ψ_S (given the monotonicity property of $S \mapsto \Psi_S$). Therefore, the dominated convergence theorem may be applied and yields the claimed equality,

$$2 \int_{\mathcal{X}\times\Theta} \mathrm{VT}_S(x,\theta)\,\mathrm{d}\mu(x)\mathrm{d}\theta = \int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\,,$$

which shows that Lemma 4 is also valid in the case of not necessarily bounded statistics.

7.3. Modifications needed under the milder Assumption $\mathbb{L}_2$.Diff.weak

We prove here Facts 1 and 2.

Proof of Fact 1: It suffices to show that Lemmas 5 and 8, as well as (21), are still valid when Assumption $\mathbb{L}_2$.Diff is replaced by $\mathbb{L}_2$.Diff.weak. Note that in the whole paper the gradient ∇ is merely a compact notation for the vector of partial derivatives, as introduced in (2); the full differentiability of the functions $\Theta \to \mathbb{R}$ or $\Theta \to \mathbb{R}^s$ considered is actually never needed, as we proceed component by component (this is best seen, e.g., in the proof of Lemma 6, and is also visible in the definition of absolute continuity). Also, in view of the statement of the van Trees inequality, we can always assume, with no loss of generality, that $\Theta = \mathrm{Supp}(q)$. In particular, the proofs of Lemmas 5 and 8 only rely on the fact that for almost all $\theta \in \mathrm{Supp}(q)$, for all $i \in \{1,\ldots,p\}$,

$$\int_{\mathcal{X}} \xi_\theta \dot\xi_{\theta,i}\,\mathrm{d}\mu = 0 \qquad \text{and} \qquad \frac{\partial\gamma_S}{\partial\theta_i}(\theta) = 2\int_{\mathcal{X}} \xi_\theta\dot\xi_{\theta,i} \otimes S\,\mathrm{d}\mu\,.$$

Both equalities follow from (4) and Lemma 2 applied in the one-dimensional models $(P_{\theta+te_i})_{|t|<\delta_\theta}$ at $t=0$, where δ_θ is sufficiently small; indeed, thanks to Assumption $\mathbb{L}_2$.Diff.weak, these models are almost all differentiable in the sense of Definition 1.

Proof of Fact 2: We exploit the remark following the statement of Proposition 2 thanks to the fact that almost all points of a locally integrable function are Lebesgue points; we will do so by, again, restricting our attention to one-dimensional models along one canonical coordinate. More precisely, under Assumption AC.f_θ, we consider the matrix-valued function

$$\mathcal{J}:\ \theta\in\Theta \longmapsto \int_{\mathcal{X}} \frac{\nabla f_\theta \otimes \nabla f_\theta}{f_\theta}\,\mathbb{I}_{\{f_\theta>0\}}\,\mathrm{d}\mu\,.$$

This is the Fisher information in the setting with pointwise assumptions. Of course, via (the proof of) Lemma 3, we know that at the points θ where the model is $\mathbb{L}_2(\mu)$–differentiable, we have $\mathcal{I}(\theta) = \mathcal{J}(\theta)$. What we mean when we say that Gill and Levit [1995] also consider Assumption Int.q is, as far as its second part is concerned, that

$$\int_\Theta \mathrm{Tr}\big(\mathcal{J}(\theta)\big)\,q(\theta)\,\mathrm{d}\theta < +\infty\,.$$

Since q is continuous, this entails that $\mathcal{J}$ is locally integrable on $\mathrm{Supp}(q)$. We fix some $i \in \{1,\ldots,p\}$. The ith diagonal element of $\mathcal{J}$,

$$\mathcal{J}_{i,i}:\ \theta\in\Theta \longmapsto \int_{\mathcal{X}} \left(\frac{\partial f_\theta}{\partial\theta_i}\right)^2 \frac{\mathbb{I}_{\{f_\theta>0\}}}{f_\theta}\,\mathrm{d}\mu\,,$$

is locally integrable as well. Therefore, by Fubini's theorem, for almost all $\theta_{-i}\in\Theta_{-i}$, the function $\theta_i \in \Theta(\theta_{-i}) \mapsto \mathcal{J}_{i,i}(\theta_i,\theta_{-i})$ is locally integrable as well and therefore, as is well-known in integration theory, almost all points of $\Theta(\theta_{-i})$ are Lebesgue points for this functions. At these points, by definition,

$$\frac{1}{t}\int_0^t \mathcal{J}_{i,i}(\theta_i + t', \theta_{-i})\,\mathrm{d}t' \longrightarrow \mathcal{J}_{i,i}(\theta_i, \theta_{-i}) \qquad \text{and} \qquad \frac{1}{t}\int_0^t \mathcal{J}_{i,i}(\theta_i - t', \theta_{-i})\,\mathrm{d}t' \longrightarrow \mathcal{J}_{i,i}(\theta_i, \theta_{-i})$$

as $t \to 0$ with positive values. This shows, by Proposition 2 and the remark following its statement, that the one-dimensional model $(P_{\theta+te_i})_{|t|<\delta_{(\theta_i,\theta_{-i})}}$, where $\delta_{(\theta_i,\theta_{-i})}$ is small enough, is differentiable in $\mathbb{L}_2(\mu)$ at $t = 0$ (in the sense of Definition 1).

All in all (e.g., by Fubini's theorem), we showed that the model is coordinate-wise differentiable in $\mathbb{L}_2(\mu)$ at almost all $\theta \in \mathrm{Supp}(q)$.

References

P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J.A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, 1993.

B.Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai. Some classes of global Cramér–Rao bounds. *The Annals of Statistics*, 15(4):1421–1438, 1987.

R. Gill and B. Levit. Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1 (1–2):59–79, 1995.

I.A. Ibragimov and R.Z. Has'minskii. *Statistical Estimation: Asymptotic Theory*. Springer, 1981.

P.E. Jupp. A van Trees inequality for estimators on manifolds. *Journal of Multivariate Analysis*, 101:1814–1825, 2010.

A.J. Lenstra. Cramér-Rao revisited. *Bernoulli*, 11(2):263–282, 2005.

G. Letac. Cramér–Rao inequalities and Fisher information. Technical report, 2008.

D. Pollard. Asymptotia (book in progress), chapter on Hellinger differentiability, 2001; 2005. Lecture notes from the Paris 2001 semester, with a final edit in 2005; available at `http://www.stat.yale.edu/~pollard/Courses/607.spring05/handouts/DQM.pdf`.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Unversity Press, 1998.

H.L. van Trees. *Detection, Estimation and Modulation Theory*. Wiley & Sons, 1968.

A. Appendix

This appendix is provided solely for the convenience of the reader. It contains only well-known results and proof techniques.

A.1. Some sufficient conditions for $\mathbb{L}_2(\mu)$–differentiability

We consider a statistical model $(\mathbb{G}_\beta)$ defined on a measurable space $(\mathcal{Y}, \mathcal{G})$ and indexed by $\beta \in V$, where $V \subseteq \mathbb{R}^p$ is some open set containing a point β_0 of interest. We assume that the statistical model is dominated by a measure ν and denote the density functions of the $\mathbb{G}_\beta$ with respect to ν by g_β.

The following result corresponds to Bickel et al. [1993, Proposition 1], see also the adaptation by Pollard [2001; 2005].

Proposition 2. *Suppose that*

(i) *the map $(y, \beta) \in \mathcal{Y} \times V \longmapsto g_\beta(y)$ is product measurable;*

(ii) *for ν–almost all y, the function $\beta \in V \longmapsto g_\beta(y)$ is absolutely continuous on V, with gradient function denoted by $\beta \in V \longmapsto \nabla g_\beta(y)$, which is defined for all β with the convention that it equals $[0]$ where it was not defined;*

(iii) *for ν–almost all y, the function $\beta \in V \longmapsto g_\beta(y)$ is differentiable at β_0;*

(iv) *for each $\beta \in V$, the function*

$$\zeta_\beta : y \in \mathcal{Y} \;\longmapsto\; \frac{1}{2} \frac{\nabla g_\beta(y)}{\sqrt{g_\beta(y)}} \mathbb{I}_{\{g_\beta(y) > 0\}}$$

is $\mathbb{L}_2(\nu)$–integrable, with the convergence, as $\beta \to \beta_0$,

$$\mathbb{E}_\nu\big[\zeta_\beta^2\big] \;\longrightarrow\; \mathbb{E}_\nu\big[\zeta_{\beta_0}^2\big]\,.$$

Then the model $(\mathbb{G}_\beta)_{\beta\in V}$ *is* $\mathbb{L}_2(\nu)$*–differentiable at* β_0*, with derivative* ζ_{β_0}*.*

The proof only uses the continuity stated in the second part of (iv) to show that some sequence of integral averages of $\mathcal{J}(\beta') = \mathbb{E}_\nu\big[\zeta^2_{\beta'}\big]$ over β' in shrinking neighborhoods of β_0 converges to $\mathbb{E}_\nu\big[\zeta^2_{\beta_0}\big]$.

This continuity condition can be relaxed at least in one-dimensional models (i.e., when $p = 1$). We consider the function $\mathcal{J} : \beta \in V \mapsto \mathbb{E}_\nu\big[\zeta^2_\beta\big]$. The second part of (iv) can then be replaced by assuming the convergences, as $t \to 0$ with positive values,

$$\frac{1}{t}\int_0^t \mathcal{J}(\beta_0 + t')\,\mathrm{d}t' \longrightarrow \mathcal{J}(\beta_0) \qquad \text{and} \qquad \frac{1}{t}\int_0^t \mathcal{J}(\beta_0 - t')\,\mathrm{d}t' \longrightarrow \mathcal{J}(\beta_0)\,.$$

These convergences correspond to β_0 being what is called a Lebesgue point of $\mathcal{J}$.

A.2. Proof of Lemma 1

Proof. The application γ_T is well-defined because of the hypothesis on the second moments of T under the $\mathbb{P}_\theta$; so is also the candidate for the gradient, since by the Cauchy-Schwarz inequality,

$$\int_{\mathcal{X}} \Big|\xi_{\theta_0}\dot{\xi}_{\theta_0} T\Big|\,\mathrm{d}\mu \leqslant \sqrt{p\,\mathbb{E}_{\theta_0}\big[T^2\big]}\,\Big\|\dot{\xi}_{\theta_0}\Big\|_\mu\,.$$

Similar arguments show that all integrals considered below exist. Now,

$$\begin{aligned}
&\gamma_T(\theta_0) - \gamma_T(\theta) - 2(\theta_0-\theta)^{\mathrm{T}}\int_{\mathcal{X}} \xi_{\theta_0}\dot{\xi}_{\theta_0} T\,\mathrm{d}\mu \\
&= \int_{\mathcal{X}} \Big(\xi^2_{\theta_0} - \xi^2_\theta - 2(\theta_0-\theta)^{\mathrm{T}}\,\dot{\xi}_{\theta_0}\xi_{\theta_0}\Big)\,T\,\mathrm{d}\mu \\
&= \int_{\mathcal{X}} \Big((\xi_{\theta_0}+\xi_\theta)(\xi_{\theta_0}-\xi_\theta) - (\theta_0-\theta)^{\mathrm{T}}\,\dot{\xi}_{\theta_0}\big((\xi_{\theta_0}+\xi_\theta)+(\xi_{\theta_0}-\xi_\theta)\big)\Big)\,T\,\mathrm{d}\mu \\
&= \int_{\mathcal{X}} (\xi_{\theta_0}+\xi_\theta)\big(\xi_{\theta_0}-\xi_\theta-(\theta_0-\theta)^{\mathrm{T}}\,\dot{\xi}_{\theta_0}\big)\,T\,\mathrm{d}\mu \; - \; (\theta_0-\theta)^{\mathrm{T}}\int_{\mathcal{X}} \dot{\xi}_{\theta_0}(\xi_{\theta_0}-\xi_\theta)\,T\,\mathrm{d}\mu\,. \qquad (34)
\end{aligned}$$

It suffices to show that each of the two summands in the last equality is negligible with respect to $\theta_0 - \theta$. We bound both of them by the Cauchy-Schwarz inequality. First,

$$\left\|\int_{\mathcal{X}} (\xi_{\theta_0}+\xi_\theta)\big(\xi_{\theta_0}-\xi_\theta-(\theta_0-\theta)^{\mathrm{T}}\,\dot{\xi}_{\theta_0}\big)\,T\,\mathrm{d}\mu\right\|_1 \leqslant \sqrt{p}\,\big\|\,(\xi_{\theta_0}+\xi_\theta)\,T\,\big\|_\mu\,\Big\|\xi_{\theta_0}-\xi_\theta-(\theta_0-\theta)^{\mathrm{T}}\,\dot{\xi}_{\theta_0}\Big\|_\mu\,;$$

the second norm in the above bound is $o\big(\|\theta_0-\theta\|\big)$ by the assumption of differentiability in $\mathbb{L}_2(\mu)$ of the model, while the first norm is bounded by

$$\big\|\,\xi_{\theta_0}T\,\big\|_\mu + \big\|\,\xi_\theta T\,\big\|_\mu = \sqrt{\mathbb{E}_{\theta_0}\big[T^2\big]} + \sqrt{\mathbb{E}_\theta\big[T^2\big]} \leqslant 2\sqrt{M_{U,T}}\,.$$

As for the second integral in (34), because of the factor $(\theta_0-\theta)^{\mathrm{T}}$ in front of it, we only need to show that each of its components tends to 0 as $\theta \to \theta_0$. To do so, we split it according to whether $|T|$ is larger or smaller to a given threshold K and resort again to the Cauchy-Schwarz inequality; this yields

$$\begin{aligned}
&\left\|\int_{\mathcal{X}} (\xi_{\theta_0}-\xi_\theta)\dot{\xi}_{\theta_0}\,T\,\mathrm{d}\mu\right\|_1 \\
&\leqslant \int_{\mathcal{X}} \Big\|\,(\xi_{\theta_0}-\xi_\theta)\dot{\xi}_{\theta_0}\,T\Big\|_1 \mathbb{I}_{\{|T|>K\}}\,\mathrm{d}\mu + \int_{\mathcal{X}} \Big\|\,(\xi_{\theta_0}-\xi_\theta)\dot{\xi}_{\theta_0}\,T\Big\|_1 \mathbb{I}_{\{|T|\leqslant K\}}\,\mathrm{d}\mu \\
&\leqslant \sqrt{p}\,\Big\|\,(\xi_{\theta_0}-\xi_\theta)\,T\Big\|_\mu\,\Big\|\dot{\xi}_{\theta_0}\,\mathbb{I}_{\{|T|>K\}}\Big\|_\mu + K\sqrt{p}\,\|\,\xi_{\theta_0}-\xi_\theta\,\|_\mu\,\|\,\dot{\xi}_{\theta_0}\,\|_\mu \\
&\leqslant 2\sqrt{p}\sqrt{M_{U,T}}\,\Big\|\dot{\xi}_{\theta_0}\,\mathbb{I}_{\{|T|>K\}}\Big\|_\mu + K\sqrt{p}\,\|\,\xi_{\theta_0}-\xi_\theta\,\|_\mu\,\|\,\dot{\xi}_{\theta_0}\,\|_\mu\,.
\end{aligned}$$

The differentiability in $\mathbb{L}_2(\mu)$ of the model implies in particular that

$$\left\| \xi_{\theta_0} - \xi_\theta \right\|_\mu = O\big(\|\theta_0 - \theta\|\big)\,,$$

so that

$$\limsup_{\theta \to \theta_0} \left\| \int_{\mathcal{X}} (\xi_{\theta_0} - \xi_\theta)\dot{\xi}_{\theta_0}\, T\, \mathrm{d}\mu \right\| \leqslant 2\sqrt{p}\sqrt{M_{U,T}}\, \left\| \dot{\xi}_{\theta_0}\, \mathbb{I}_{\{|T|>K\}} \right\|_\mu$$

for all $K > 0$. Letting $K \to +\infty$, an argument of dominated convergence shows that

$$\limsup_{\theta \to \theta_0} \left| \int_{\mathcal{X}} (\xi_{\theta_0} - \xi_\theta)\dot{\xi}_{\theta_0}\, T\, \mathrm{d}\mu \right| = 0\,.$$

Substituting this in (34) concludes the proof. □

A.3. Schur complements

Lemma 9. *We consider an $s \times s$ matrix A, a $p \times s$ matrix B, and a $p \times p$ matrix D, where D is invertible. If*

$$M = \begin{bmatrix} A & B^{\mathrm{T}} \\ B & D \end{bmatrix} \succcurlyeq 0\,,$$

then the so-called Schur complement of its D block also satisfies

$$A - B^{\mathrm{T}} D^{-1} B \succcurlyeq 0\,. \tag{35}$$

Proof. We denote by Id_s the $s \times s$ identity matrix. We have the equalities

$$\begin{aligned} &\begin{bmatrix} \mathrm{Id}_s & -\big(D^{-1}B\big)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} A & B^{\mathrm{T}} \\ B & D \end{bmatrix} \begin{bmatrix} \mathrm{Id}_s \\ -D^{-1}B \end{bmatrix} \\ &= \begin{bmatrix} \mathrm{Id}_s & -\big(D^{-1}B\big)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} A - B^{\mathrm{T}} D^{-1} B \\ [0] \end{bmatrix} \\ &= A - B^{\mathrm{T}} D^{-1} B\,. \end{aligned}$$

This indeed entails that $A - B^{\mathrm{T}} D^{-1} B \succcurlyeq 0$ as well. □

Based on this result, we get the following result, used by Gill and Levit [1995] in one of their proofs.

Lemma 10. *We consider three $s \times s$ matrices A, B, and D, where D is invertible, such that*

$$M = \begin{bmatrix} A & B^{\mathrm{T}} \\ B & D \end{bmatrix} \succcurlyeq 0\,.$$

Then

$$\mathrm{Tr}(A) \geqslant \frac{\big(\mathrm{Tr}(B)\big)^2}{\mathrm{Tr}(D)}\,.$$

Proof. We take the trace in (35) and get

$$\mathrm{Tr}(A) \geqslant \mathrm{Tr}\big(B^{\mathrm{T}} D^{-1} B\big)\,.$$

Now, the assumption on the three matrices A, B, D, entails in particular that D is symmetric. Since D is also invertible, we may write it as $D = U\,U^{\mathrm{T}}$ where U is a $s \times s$ invertible matrix. We have the rewriting

$$\mathrm{Tr}\big(B^{\mathrm{T}} D^{-1} B\big) = \mathrm{Tr}\Big(\big(U^{-1}B\big)^{\mathrm{T}}\,\big(U^{-1}B\big)\Big)\,.$$

Now, the Cauchy-Schwarz inequality for the inner product $(M_1, M_2) \mapsto \mathrm{Tr}\big(M_1^{\mathrm{T}} M_2\big)$, with $M_1 = U^{\mathrm{T}}$ and $M_2 = U^{-1}B$, indicates that

$$\big(\mathrm{Tr}(B)\big)^2 = \Big(\mathrm{Tr}\Big(U\,\big(U^{-1}B\big)\Big)\Big)^2 \leqslant \underbrace{\mathrm{Tr}\big(U\,U^{\mathrm{T}}\big)}_{\mathrm{Tr}(D)}\,\mathrm{Tr}\Big(\big(U^{-1}B\big)^{\mathrm{T}}\,\big(U^{-1}B\big)\Big)\,.$$

The proof is concluded by putting all (in)equalities together. □

A.4. Derivation of other multivariate formulations of the van Trees inequality

The other, at first sight more general, multivariate formulations of the inequality that Gill and Levit [1995] follow from two immediate adaptations of what we could call the "van Trees equality." The latter is the key step in proving the van Trees inequality, namely, the equality stated in Lemma 4. We recall it: under suitable assumptions, the following integrals are defined and are equal,

$$2\int_{\mathcal{X}\times\Theta} \Big(\Delta_q(x,\theta)\otimes\big(S(x)-\psi(\theta)\big)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta = \int_\Theta \nabla\psi(\theta)\,q(\theta)\,\mathrm{d}\theta\,, \tag{36}$$

where $S : \mathcal{X} \to \mathbb{R}^s$ is a statistic, $\psi : \Theta \to \mathbb{R}^s$ is an absolutely continuous function, and the pseudo-derivative $\Delta_q : \mathcal{X}\times\Theta \to \mathbb{R}^p$ equals

$$\Delta_q(x,\theta) = \frac{1}{2}\frac{\nabla q(\theta)}{\sqrt{q(\theta)}}\,\mathbb{I}_{\{q(\theta)>0\}}\,\xi_\theta(x) + \sqrt{q(\theta)}\;\dot{\xi}_\theta(x) \tag{37}$$

for all $x \in \mathcal{X}$ and $\theta \in \Theta$. This is a matrix equality. The equality for the elements in the ith row and jth column reads:

$$2\int_{\mathcal{X}\times\Theta}\left(\frac{\partial q}{\partial\theta_i}(\theta)\,\frac{\mathbb{I}_{\{q(\theta)>0\}}}{2\sqrt{q(\theta)}}\,\xi_\theta(x) + \sqrt{q(\theta)}\;\dot{\xi}_{\theta,i}(x)\right)\big(S_j(x)-\psi_j(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta$$
$$= \int_\Theta \frac{\partial\psi_j}{\partial\theta_i}(\theta)\,q(\theta)\,\mathrm{d}\theta\,. \tag{38}$$

Gill and Levit [1995] discuss two degrees of flexibility in (the application of) these equalities, which can be combined. For the sake of clarity we recall them separately, the first one in the form of matrix inequalities while the second one is best stated in terms of scalar inequalities. The latter are derived by Gill and Levit [1995] via the consideration of Schur complements and the application of the Cauchy-Schwarz inequality, at the cost of not being intrinsic.

Different priors for different coordinates. The prior q considered in (37) is the same for all coordinates (i,j) but this obviously does not have to be. It could depend on (i,j) but this would not lead to an elegant inequality; instead we have it depend only on i but allow ourselves to consider several of these modified priors, say, p' of them. That is, for all coordinates $(i,\cdot)$, we introduce some absolutely continuous function $c_{i,\ell} : \Theta \to \mathbb{R}$, where $\ell \in \{1,\ldots,p'\}$ and form the priors $q\,c_{i,\ell}$. Under suitable conditions (38) holds for each coordinate (i,j) with q replaced by $q\,c_{i,\ell}$. Summing these equalities over i, slightly rearranging the integrand, and omitting the indicator functions, we get for all $\ell \in \{1, \ldots, p'\}$ and $j \in \{1, \ldots, s\}$,

$$2 \int_{\mathcal{X}\times\Theta} \underbrace{\sum_{i=1}^{p} \left(\frac{\partial(q\, c_{i,\ell})}{\partial \theta_i}(\theta)\, \frac{1}{2\sqrt{q(\theta)}}\, \xi_\theta(x) + c_{i,\ell}\sqrt{q(\theta)}\; \dot{\xi}_{\theta,i}(x) \right)}_{\widetilde{\Delta}_{C,\ell}} \big(S_j(x) - \psi_j(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)}\, \mathrm{d}\mu(x)\mathrm{d}\theta$$

$$= \int_\Theta \left(\sum_{i=1}^{p} c_{i,\ell} \frac{\partial \psi_j}{\partial \theta_i}(\theta) \right) q(\theta)\, \mathrm{d}\theta\,.$$

Matrix-wise, denoting by $C = (c_{\ell,i})$ the $p' \times p$–matrix valued function at hand and by $\widetilde{\Delta}_C : \mathcal{X}\times\Theta \to \mathbb{R}^{p'}$ the modified pseudo-derivative, we have proved the matrix equality

$$2 \int_{\mathcal{X}\times\Theta} \Big(\widetilde{\Delta}_C(x,\theta) \otimes \big(S(x) - \psi(\theta)\big) \Big) \xi_\theta(x)\sqrt{q(\theta)}\, \mathrm{d}\mu(x)\mathrm{d}\theta = \int_\Theta C(\theta)^{\mathrm{T}}\, \nabla\psi(\theta)\, q(\theta)\, \mathrm{d}\theta\,.$$

Based on it, following the steps detailed after the statement of Lemma 4 (and under suitable assumptions ensuring that all needed quantities exist), we may derive some van Trees inequality:

$$\begin{bmatrix} \displaystyle\int_\Theta \mathbb{E}_\theta\Big[\big(S - \psi(\theta)\big) \otimes \big(S - \psi(\theta)\big) \Big]\, q(\theta)\, \mathrm{d}\theta & \displaystyle\left(\int_\Theta C(\theta)^{\mathrm{T}}\, \nabla\psi(\theta)\, q(\theta)\, \mathrm{d}\theta \right)^{\mathrm{T}} \\ \displaystyle\int_\Theta C(\theta)^{\mathrm{T}}\, \nabla\psi(\theta)\, q(\theta)\, \mathrm{d}\theta & \displaystyle\widetilde{\mathcal{I}}_{C,q} + \int_\Theta C(\theta)\, \mathcal{I}(\theta)\, C(\theta)^{\mathrm{T}}\; q(\theta)\, \mathrm{d}\theta \end{bmatrix}$$

$$= \int_{\mathcal{X}\times\Theta} \begin{bmatrix} \big(S(x) - \psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)} \\ 2\widetilde{\Delta}_C \end{bmatrix} \otimes \begin{bmatrix} \big(S(x) - \psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)} \\ 2\widetilde{\Delta}_C(x,\theta) \end{bmatrix} \mathrm{d}\mu(x)\mathrm{d}\theta \;\succcurlyeq 0\,, \qquad (39)$$

after straightforward calculations, using (4), showing that

$$\int_{\mathcal{X}\times\Theta} \widetilde{\Delta}_C(x,\theta) \otimes \widetilde{\Delta}_C(x,\theta)\, \mathrm{d}\mu(x)\mathrm{d}\theta = \widetilde{\mathcal{I}}_{C,q} + \int_\Theta C(\theta)\, \mathcal{I}(\theta)\, C(\theta)^{\mathrm{T}}\; q(\theta)\, \mathrm{d}\theta\,,$$

where $\widetilde{\mathcal{I}}_{C,q}$ is a $p' \times p'$ matrix, whose component (ℓ,ℓ') equals

$$\widetilde{\mathcal{I}}_{C,q} = \int_{\mathcal{X}\times\Theta} \sum_{i,i'=1}^{p} \frac{\partial(q\, c_{i,\ell})}{\partial \theta_i}(\theta)\, \frac{\partial(q\, c_{i',\ell'})}{\partial \theta_i}(\theta)\, \frac{\mathbb{I}_{\{q(\theta)>0\}}}{q(\theta)}\, \mathrm{d}\theta\,.$$

Scalar inequalities. What we call scalar inequalities is the application of Lemma 10 of Section A.3 to matrix inequalities of the form (22) or (39). For instance, when $p' = s$, the stated lemma and (39) lead to

$$\int_\Theta \mathbb{E}_\theta\Big[\big\| S - \psi(\theta) \big\|^2 \Big]\, q(\theta)\, \mathrm{d}\theta \geqslant \frac{\left(\displaystyle\int_\Theta \mathrm{Tr}\big(C(\theta)^{\mathrm{T}}\, \nabla\psi(\theta) \big)\, q(\theta)\, \mathrm{d}\theta \right)^2}{\mathrm{Tr}\big(\widetilde{\mathcal{I}}_{C,q}\big) + \displaystyle\int_\Theta \mathrm{Tr}\big(C(\theta)\, \mathcal{I}(\theta)\, C(\theta)^{\mathrm{T}} \big)\, q(\theta)\, \mathrm{d}\theta}\,. \qquad (40)$$

Inequalities of the form above are generally less satisfactory than their mother matrix inequalities. This can be seen, in the case when $p = s$ and C is the matrix of identity functions, by noting that the numerator of the right-hand side is given by

$$\int_\Theta \mathrm{Tr}\big(\nabla\psi(\theta)\big)\, q(\theta)\, \mathrm{d}\theta = \int_\Theta \left(\sum_{i=1}^{p} \frac{\partial \psi_i}{\partial \theta_i}(\theta) \right) q(\theta)\, \mathrm{d}\theta\,.$$

This term is not stable by a permutation of the components of ψ, while the left-hand side of (40) is stable by this operation. This is why, following Letac [2008], we only stated matrix inequalities.

Weighted Bayesian quadratic risk. The second variation works best as far as scalar inequalities are concerned, and in the case when $p = s$ or after the first variation, i.e., based on (39), in the case when $p' = s$. For simplicity, we rather illustrate it based on (22) in the case when $p = s$. This variation relies on a function A defined over Θ and with values in the set of $p \times p$ invertible matrices. We associate with it the symmetric matrix $B = A\,A^{\mathrm{T}}$. No regularity assumption is needed on A, only suitable integrability assumptions (for the quantities considered below to be defined) will be required. We start by noting that under suitable integrability assumptions only,

$$\int_{\mathcal{X}\times\Theta} \begin{bmatrix} A(\theta)^{-1}\big(S(x)-\psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)} \\ 2A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta) \end{bmatrix} \otimes \begin{bmatrix} A(\theta)^{-1}\big(S(x)-\psi(\theta)\big)\xi_\theta(x)\sqrt{q(\theta)} \\ 2A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta) \end{bmatrix} \mathrm{d}\mu(x)\mathrm{d}\theta \;\succcurlyeq 0\,.$$

An application of Lemma 10, under an invertibility condition, indicates that

$$\begin{aligned}&\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)^{\mathrm{T}}\,B(\theta)^{-1}\,\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta \\ &\qquad \geqslant \frac{\left(2\displaystyle\int_{\mathcal{X}\times\Theta}\mathrm{Tr}\Big(A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\otimes A(\theta)^{-1}\big(S(x)-\psi(\theta)\big)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta\right)^2}{\displaystyle\int_{\mathcal{X}\times\Theta}\mathrm{Tr}\Big(A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\otimes A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta}\,. \qquad (41)\end{aligned}$$

As

$$\begin{aligned}\mathrm{Tr}\Big(A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\otimes A(\theta)^{-1}\big(S(x)-\psi(\theta)\big)\Big) &= \mathrm{Tr}\Big(\Big(A(\theta)^{-1}\big(S(x)-\psi(\theta)\big)\Big)^{\mathrm{T}}\,A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\Big) \\ &= \mathrm{Tr}\Big(\big(S(x)-\psi(\theta)\big)^{\mathrm{T}}\,\big(A(\theta)^{-1}\big)^{\mathrm{T}}\,A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\Big) \\ &= \mathrm{Tr}\Big(\big(S(x)-\psi(\theta)\big)^{\mathrm{T}}\,\Delta_q(x,\theta)\Big)\,,\end{aligned}$$

we get from (36), still under no regularity condition on A,

$$2\int_{\mathcal{X}\times\Theta}\mathrm{Tr}\Big(A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\otimes A(\theta)^{-1}\big(S(x)-\psi(\theta)\big)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta = \int_\Theta \mathrm{Tr}\big(\nabla\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\,.$$

As for the denominator in (42) we first note that

$$\begin{aligned}\mathrm{Tr}\Big(A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\otimes A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\Big) &= \mathrm{Tr}\Big(\big(A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\big)^{\mathrm{T}}\,A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\Big) \\ &= \mathrm{Tr}\big(\Delta_q(x,\theta)^{\mathrm{T}}\,B(\theta)\,\Delta_q(x,\theta)\big) = \mathrm{Tr}\big(B(\theta)\,\Delta_q(x,\theta)\Delta_q(x,\theta)^{\mathrm{T}}\,\big)\,.\end{aligned}$$

The integration of this term is then handled in the same way (21) was obtained, in particular, thanks to (4); we only write the resulting equality:

$$\begin{aligned}&\int_{\mathcal{X}\times\Theta}\mathrm{Tr}\Big(A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\otimes A(\theta)^{\mathrm{T}}\,\Delta_q(x,\theta)\Big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta \\ &\quad= \int_{\mathcal{X}\times\Theta}\mathrm{Tr}\big(B(\theta)\Delta_q(x,\theta)\Delta_q(x,\theta)^{\mathrm{T}}\,\big)\xi_\theta(x)\sqrt{q(\theta)}\,\mathrm{d}\mu(x)\mathrm{d}\theta \\ &\quad= \int_\Theta \mathrm{Tr}\bigg(B(\theta)\,\frac{\nabla q(\theta)\nabla q(\theta)^{\mathrm{T}}}{q(\theta)}\,\mathbb{I}_{\{q(\theta)>0\}}\bigg)\,\mathrm{d}\theta + \int_\Theta \mathrm{Tr}\big(B(\theta)^{\mathrm{T}}\,\mathcal{I}(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\,.\end{aligned}$$

Putting all (in)equalities together, we get the following weighted inequality:

$$\int_\Theta \mathbb{E}_\theta\Big[\big(S-\psi(\theta)\big)^{\mathrm{T}}\,B(\theta)^{-1}\,\big(S-\psi(\theta)\big)\Big]\,q(\theta)\,\mathrm{d}\theta \geqslant \frac{\left(\displaystyle\int_\Theta \mathrm{Tr}\big(\nabla\psi(\theta)\big)\,q(\theta)\,\mathrm{d}\theta\right)^2}{\displaystyle\int_\Theta \mathrm{Tr}\left(B(\theta)\,\frac{\nabla q(\theta)\nabla q(\theta)^{\mathrm{T}}}{q(\theta)}\,\mathbb{I}_{\{q(\theta)>0\}}\right)\mathrm{d}\theta + \int_\Theta \mathrm{Tr}\big(B(\theta)^{\mathrm{T}}\,\mathcal{I}(\theta)\big)\,q(\theta)\,\mathrm{d}\theta}\,. \tag{42}$$