C. R. Acad. Sci. Paris, Ser. I 350 (2012) 711–715

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

A Cramér–Rao inequality for non-differentiable models

Une inégalité de type Cramér–Rao pour des modèles non différentiables

Thibault Espinasse, Paul Rochet

Institut de mathématique de Toulouse, université Paul-Sabatier, 118, route de Narbonne, 31068 Toulouse cedex 9, France

A R T I C L E   I N F O

Article history:
Received 13 April 2012
Accepted after revision 20 August 2012
Available online 6 September 2012

Presented by Paul Deheuvels

A B S T R A C T

We compute a variance lower bound for unbiased estimators in statistical models. The construction of the bound is related to the original Cramér–Rao bound, although it does not require the differentiability of the model. Moreover, we show our efficiency bound to be always greater than the Cramér–Rao bound in smooth models, thus providing a sharper result.

R É S U M É

Nous obtenons une minoration pour la variance d'un estimateur sans biais dans un modèle statistique. La construction de la borne est liée à celle de la borne de Cramér–Rao, mais elle ne nécessite pas d'hypothèse de différentiabilité sur le modèle. De plus, nous montrons que la borne est toujours supérieure ou égale à la borne de Cramér–Rao dans les modèles différentiables, et fournit ainsi un résultat plus fort.

Version française abrégée

Les différentes théories de l'efficacité développées dans la littérature ont pour but d'établir un critère objectif pour juger de l'optimalité d'un estimateur dans une certaine classe. L'exemple le plus célèbre est sans aucun doute l'inégalité de Cramér–Rao, qui dans sa forme la plus simple, affirme que la variance d'un estimateur sans biais dans un modèle paramétrique est supérieure à l'inverse de l'information de Fisher. Ce résultat datant de 1945 [6] est à la base de nombreuses théories sur l'efficacité qui s'appuient sur les travaux de Le Cam [5] et Hájek [4], et qui ont permis d'étendre l'inégalité de Cramér–Rao à des modèles plus généraux et sous des hypothèses de régularité différentes (voir [1] et [7]).

Alors que l'inégalité de Cramér–Rao impose que le modèle soit différentiable, il s'avère que la construction d'une borne inférieure sur la variance d'un estimateur sans biais n'est pas soumise à des conditions de différentiabilité du modèle ou du paramètre à estimer. En fait, la preuve initiale de l'inégalité de Cramér–Rao, qui dérive d'une simple application de l'inégalité de Cauchy–Schwarz, peut facilement se généraliser à des modèles non différentiables, sous la seule hypothèse que l'estimateur est sans biais pour les lois considérées. Précisément, on montre dans le théorème 1 que si un estimateur T d'un paramètre $\psi(.)$ est sans biais pour deux lois de probabilité ν et μ (μ étant la loi des observations), alors

$$\mathrm{var}(T) \geqslant \frac{(\psi(\mu) - \psi(\nu))^2}{(\mathcal{Q}(\nu|\mu) + 1)^n - 1},$$

E-mail addresses: espinasse@math.univ-toulouse.fr (T. Espinasse), paul.rochet@gmail.com (P. Rochet).

1631-073X/$ – see front matter
http://dx.doi.org/10.1016/j.crma.2012.08.005

où $\mathcal{Q}(\nu|\mu)$ désigne la divergence quadratique de ν par rapport μ et n est le nombre d'observations. Ainsi, si l'on suppose l'estimateur T sans biais dans un modèle donné, chaque élément de ce modèle (exceptée la vraie mesure μ) fournit un minorant de la variance de T. Cette propriété donne lieu à une construction légèrement différente d'une borne d'efficacité qui peut s'appliquer à tout type de modèle, paramétrique ou semi paramétrique, sans aucune hypothèse de différentiabilité.

Cette borne d'efficacité satisfait également deux propriétés intéressantes dans le cadre des modèles différentiables. Premièrement, on montre qu'il existe toujours un élément ν du modèle pour lequel la borne obtenue est supérieure ou égale à la borne de Cramér–Rao. En effet, dans un modèle $\mathcal{P}$ différentiable,

$$\sup_{\nu\in\mathcal{P}} \frac{(\psi(\mu)-\psi(\nu))^2}{(\mathcal{Q}(\nu|\mu)+1)^n-1} \geqslant \mathcal{B}_\psi,$$

où $\mathcal{B}_\psi$ désigne la borne de Cramér–Rao classique (Proposition 1). Cette construction fournit ainsi un résultat au moins aussi fort, l'ingalité de Cramér–Rao devenant ainsi une conséquence directe du théorème 1. On montre également la convergence, quand le nombre d'observations tend vers l'infini, du maximum vers la borne de Cramér–Rao dans les cas réguliers. Nous donnons une brève description de ces résultats, qui sont présentés plus en détail dans [3].

1. Construction of the efficiency bound

Let $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ be an open subset of $\mathbb{R}^p$ endowed with its Borel field, we denote by $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$. We consider the classical statistical framework where we observe an i.i.d. sample $X_1,\dots,X_n$ drawn from an unknown measure μ and we want to estimate a parameter $\psi(\mu)$.

The construction of an efficiency bound relies only on two aspects which are the set $\mathcal{P}$ of possible values for the measure μ (the model) and the parameter to estimate $\psi:\mathcal{P}\to\mathcal{H}$. In this paper, we assume for simplicity that $\mathcal{H}\subset\mathbb{R}$, although the computation of the efficiency bound can be easily generalized to parameters in $\mathbb{R}^q$, as discussed in [3].

Let μ and ν be two probability measures on $(\mathcal{X},\mathcal{B})$, we define the quadratic divergence (or Q-divergence) of ν with respect to μ as

$$\mathcal{Q}(\nu|\mu)=\int_{\mathcal{X}}\left(1-\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right)^2\mathrm{d}\mu \quad \text{if } \nu\ll\mu, \qquad \mathcal{Q}(\nu|\mu)=+\infty \quad \text{otherwise}.$$

The Q-divergence is Csiszár's f-divergence associated to the convex function $f:x\mapsto(1-x)^2$ (see [2]). Let A be a subset of $\mathcal{P}(\mathcal{X})$, we define $\mathcal{Q}(A|\mu)=\inf_{\nu\in A}\mathcal{Q}(\nu|\mu)$. Any measure $\mu^*\in A$ such that $\mathcal{Q}(\mu^*|\mu)=\mathcal{Q}(A|\mu)$ is called *Q-projection* of μ onto A.

In the next theorem we show that to each element of a model, can be associated a variance lower bound for an unbiased estimator of a parameter $\psi(\mu)\in\mathbb{R}$. We use the convention $1/\infty=0$.

Theorem 1. *Let $\mathcal{P}$ be a model and $\psi:\mathcal{P}\to\mathbb{R}$ a parameter. If $T=T(X_1,\dots,X_n)$ is an unbiased estimator of ψ in the model $\mathcal{P}$, then $\forall\nu\in\mathcal{P}^*=\mathcal{P}\setminus\{\mu\}$:*

$$\operatorname{var}(T)\geqslant\frac{(\psi(\mu)-\psi(\nu))^2}{(\mathcal{Q}(\nu|\mu)+1)^n-1}.$$

We emphasize that this result does not require smoothness conditions on the model nor on the parameter. Besides, it is not needed that ν be absolutely continuous w.r.t. μ, although in this case the efficiency bound is null and provides no information.

Let $\mathcal{H}_\psi^n(.|\mu)$ denote the functional defined on $\mathcal{P}^*$ by

$$\mathcal{H}_\psi^n(\nu|\mu)=n\frac{(\psi(\mu)-\psi(\nu))^2}{(\mathcal{Q}(\nu|\mu)+1)^n-1}.$$

The quantity $\mathcal{H}_\psi^n(\nu|\mu)$ provides a lower bound for n times the variance of unbiased estimators (the factor n is used to ease comparison with the Cramér–Rao bound in parametric models, as we shall see in the next section). Because $\mathcal{H}_\psi^n(\nu|\mu)$ is null if ν is not absolutely continuous w.r.t. μ or if $\mathrm{d}\nu/\mathrm{d}\mu$ is not square μ-integrable, sufficient is to consider the values of $\mathcal{H}_\psi^n(\nu|\mu)$ for density measures $\nu=f\mu$ with f in $\mathcal{F}=\{f: f\mu\in\mathcal{P},\ \int f^2\,\mathrm{d}\mu<\infty\}$. The main advantage is that $\mathcal{F}$ being a subspace of $\mathbb{L}^2(\mu)$, it can be endowed with its natural Hilbert space topology.

Since the result is all the more informative that $\mathcal{H}_\psi^n(\nu|\mu)$ is large, we define the efficiency bound as the supremum over the whole model

$$\mathcal{B}_\psi^n(\mathcal{P}):=\sup_{\nu\in\mathcal{P}^*}\mathcal{H}_\psi^n(\nu|\mu)=\sup_{f\in\mathcal{F}\setminus\{1\}}\mathcal{H}_\psi^n(f\mu|\mu).$$

Example. Consider the Gaussian model $\{\mu_\theta\}_{\theta\in\mathbb{Z}}$, $\mu_\theta \sim \mathcal{N}(\theta, 1)$ in which we want to estimate the mean $\psi : \mu_\theta \mapsto \theta$, the true value of the parameter being $\theta_0 = 0$. Because it is supposed known that the parameter is an integer ($\theta \in \mathbb{Z}$), the model is discrete and therefore, non-differentiable. In particular, the CR bound cannot be computed. On the other hand, we have no difficulty computing our efficiency bound. Direct calculation gives $\mathcal{Q}(\mu_\theta|\mu) = e^{\theta^2} - 1$, which yields

$$\mathcal{H}_\psi^n(\mu_\theta|\mu) = n\frac{\theta^2}{e^{n\theta^2} - 1},$$

for $\theta \in \mathbb{Z}^*$. The supremum being reached for $\theta = \pm 1$, we obtain $\mathcal{B}_\psi^n = n(e^n - 1)^{-1}$. This result tells us that no unbiased estimator of the mean in this model can converge at a rate faster than e^n.

Moreover, it turns out that knowing the parameter to be an integer gives us sufficient information to build an unbiased estimator with variance converging to zero at an exponential rate. Consider for instance the estimator

$$T(X_1, \ldots, X_n) = \arg\min_{t\in\mathbb{Z}} \left| \frac{1}{n}\sum_{i=1}^n X_i - t \right|,$$

which is unbiased for all $\theta \in \mathbb{Z}$ by symmetry, and can be shown to have an exponentially decreasing variance using classical bounds on Gaussian distributions.

2. Comparison with the Cramér–Rao bound

In this section, we consider smooth parametric models, for which the Cramér–Rao bound is defined. Let $\mathcal{P} = \{\mu_\theta\}_{\theta\in\mathbb{R}}$ be such that $\mu_{\theta_0} = \mu$. We say that $\{\mu_\theta\}_\theta$ is *differentiable in* $\mathbb{L}^2(\mu)$ at θ_0 if μ_θ has density f_θ w.r.t. μ for θ in a neighborhood of θ_0 and the map $\theta \mapsto f_\theta$ is Fréchet-differentiable at θ_0 for the $\mathbb{L}^2(\mu)$ topology. Its differential $g \in \mathbb{L}^2(\mu)$ is called the *score* of the model and $\mathcal{I} = \int g^2\, d\mu$ is the *Fisher Information*.

Let $\psi : \{\mu_\theta\}_\theta \to \mathbb{R}$ be a parameter such that the map $\theta \mapsto \psi(\mu_\theta)$ is differentiable at θ_0 (we note $\dot\psi(\theta_0)$ its derivative), the Cramér–Rao inequality states that if $T = T(X_1, \ldots, X_n)$ is an unbiased estimator of ψ, then

$$n\,\mathrm{var}(T) \geqslant \frac{\dot\psi(\theta_0)^2}{\mathcal{I}}.$$

The differentiability in $\mathbb{L}^2(\mu)$ is a rather strong condition, but it is appropriate to build a variance lower bound for unbiased estimators. The more usual differentiability in quadratic mean (see [1] and [7]) is a weaker condition but it provides efficiency results for regular estimators.

Proposition 1. *Let $\{\mu_\theta\}_{\theta\in\Theta}$ be a differentiable path with $\mu_{\theta_0} = \mu$. Let $\psi : \{\mu_\theta\}_\theta \to \mathbb{R}$ be a map such that $\theta \mapsto \psi(\mu_\theta)$ is differentiable at θ_0. Then, for all $n \in \mathbb{N}$,*

$$\mathcal{B}_\psi\big(\{\mu_\theta\}_\theta\big) = \lim_{\theta\to\theta_0} \mathcal{H}_\psi^n(\mu_\theta|\mu).$$

In particular, $\mathcal{B}_\psi^n(\{\mu_\theta\}_\theta) \geqslant \mathcal{B}_\psi(\{\mu_\theta\}_\theta)$.

The efficiency bound $\mathcal{B}_\psi^n$ improves on the Cramér–Rao bound since it is defined as the supremum of $\theta \mapsto \mathcal{H}_\psi^n(\mu_\theta|\mu)$ on the model, while $\mathcal{B}_\psi$ is the limit at $\theta \to \theta_0$. As a result, in differentiable models, the functional $\mathcal{H}_\psi^n(.|\mu)$ can be extended by continuity at μ taking the value $\mathcal{H}_\psi^n(\mu|\mu) = \mathcal{B}_\psi$. In some situations, the two bounds are identical (i.e. the maximum of $\mathcal{H}_\psi^n(\nu|\mu)$ is reached as $\nu \to \mu$), which occurs for example if there exists an unbiased efficient estimator. On the other hand, it is not rare to have the strict inequality $\mathcal{B}_\psi^n > \mathcal{B}_\psi$, as we show in the following example.

Example. Consider the Gaussian model $\{\mu_\theta\}_{\theta\in\mathbb{R}}$, where $\mu_\theta \sim \mathcal{N}(\theta, 1)$ and let $\psi : \mu_\theta \mapsto e^\theta$. We take $\mu \sim \mathcal{N}(0, 1)$ as the distribution of the observations. In this model, the Cramér–Rao bound is $\mathcal{B}_\psi = 1$. On the other hand, we have $\mathcal{Q}(\mu_\theta|\mu) = e^{\theta^2} - 1$, yielding

$$\mathcal{H}_\psi^n(\mu_\theta|\mu) = n\frac{(1 - e^\theta)^2}{e^{n\theta^2} - 1},$$

for $\theta \in (-1; +\infty)$. The supremum is reached for $\theta_n = \frac{1}{n}$, which gives $\mathcal{B}_\psi^n = n(e^{1/n} - 1)$. Thus, we observe a strict inequality $\mathcal{B}_\psi^n > \mathcal{B}_\psi$ for all $n \in \mathbb{N}$. In this case, it is interesting to notice that $\mathcal{B}_\psi^n$ is the actual variance of the optimal unbiased estimator of e^θ in this model.

3. Application to semiparametric models

Extending the Cramér–Rao inequality to semiparametric models can be made using a more general definition of the Fisher Information, calculated by studying differentiable submodels. Based on the idea that, the larger the model, the less information we have, a natural definition of the Fisher Information in large models is to consider the infimum of the Fisher Informations calculated in differentiable submodels (see for instance [8]). A differentiable submodel $\{\mu_\theta\}_{\theta\in\Theta}$ for which the infimum is reached is called *least favorable path*, and the functional $\mathcal{H}_\psi^n(.|\mu)$ turns out to be an efficient tool to construct one. To see it, consider the level sets $\mathcal{F}_\theta = \{f \in \mathbb{L}^2(\mu): \ \psi(f\mu) = \theta\}$ for all values θ taken by the parameter ψ. Setting $\theta_0 = \psi(\mu)$, the expression of the efficiency bound can be written as

$$\mathcal{B}_\psi^n = \sup_{\theta\neq\theta_0} \sup_{f\in\mathcal{F}_\theta} \mathcal{H}_\psi^n(f\mu|\mu) = \sup_{\theta\neq\theta_0} n\frac{(\theta-\theta_0)^2}{(\mathcal{Q}(\mathcal{F}_\theta|\mu)+1)^n - 1}. \tag{1}$$

In these settings, we see that computing the efficiency bound reduces to maximizing a function of θ. The idea is that if we choose the least favorable density in each set $\mathcal{F}_\theta$, that is, a function f_θ maximizing $f \mapsto \mathcal{H}_\psi^n(f\mu|\mu)$, the resulting submodel would have to be a least favorable path. Since by construction, the term $\psi(f\mu) - \psi(\mu)$ is constant when f ranges over $\mathcal{F}_\theta$, a density maximizing $\mathcal{H}_\psi^n(.|\mu)$ on $\mathcal{F}_\theta$ is in fact a minimizer of $f \mapsto \mathcal{Q}(f\mu|\mu)$, which explains the term $\mathcal{Q}(\mathcal{F}_\theta|\mu)$ in (1).

In the sequel, we call *Q-projection path* a submodel $\{\mu_\theta\}_{\theta\in\Theta}$ satisfying $\mathcal{Q}(\mu_\theta|\mu) = \mathcal{Q}(\mathcal{F}_\theta|\mu)$ and $\psi(\mu_\theta) = \theta$ for all $\theta \in \Theta$. Remark that a Q-projection path does not necessarily exist, for instance if the infimum of $\mathcal{Q}(.|\mu)$ on $\mathcal{F}_\theta$ is not reachable for some values of θ. However, one does exist as soon as the map $f \mapsto \psi(f\mu)$ is continuous on $\mathcal{F}$ and if $\mathcal{Q}(\mathcal{F}_\theta|\mu)$ is finite for all $\theta \in \Theta$. By making this assumption, we avoid considering trivial cases, the efficiency bound being infinite if $f \mapsto \psi(f\mu)$ is not continuous as f tends to 1. Remark also that if the sets $\mathcal{F}_\theta$ are convex, a Q-projection path $\{\mu_\theta\}_{\theta\in\Theta}$ is unique (see [2]), μ_θ being defined as the quadratic projection of μ on $\mathcal{P}_\theta = \{\nu \in \mathcal{P}: \ \psi(\nu) = \theta\}$.

Eq. (1) tells us that a Q-projection path $\{\mu_\theta\}_\theta$ satisfies $\mathcal{B}_\psi^n(\mathcal{P}) = \mathcal{B}_\psi^n(\{\mu_\theta\}_\theta)$ for all $n \in \mathbb{N}$, which conveys that it contains the whole information of the model. In particular, it is a least favorable path if and only if it is differentiable at $\mu = \mu_{\theta_0}$.

4. Asymptotic properties

We are now interested in the asymptotic analysis of the efficiency bound. Writing the first order expansion

$$\big(\mathcal{Q}(\nu|\mu)+1\big)^n - 1 = n\mathcal{Q}(\nu|\mu) + \frac{n(n-1)}{2}\mathcal{Q}(\nu|\mu)^2 + \cdots,$$

it appears that the sequence $\{\mathcal{H}_\psi^n(.|\mu)\}_{n\in\mathbb{N}}$ is decreasing and converges pointwise toward 0 as $n \to \infty$. So, the non-negative sequence $\{\mathcal{B}_\psi^n\}_{n\in\mathbb{N}}$ is also decreasing and therefore, it converges (or is infinite). We now aim to prove that, in regular situations, the efficiency bound converges toward the Cramér–Rao bound. For this, we need the following lemma:

Lemma 1. *Assume that $\mathcal{B}_\psi^{n_0} < \infty$ for some $n_0 \in \mathbb{N}$. Then, for all $\varepsilon > 0$, $\mathcal{H}_\psi^n(.|\mu)$ converges uniformly towards 0 on the set $\{\nu \in \mathcal{P}: \ \mathcal{Q}(\nu|\mu) > \varepsilon\}$ as $n \to \infty$.*

The proof of the lemma is given in [3]. The condition that $\mathcal{B}_\psi^{n_0}$ is finite for some integer n_0 is necessary to ensure the existence of an unbiased estimator with finite variance asymptotically. It may occur that this condition is not fulfilled while the Cramér–Rao bound exists and is finite.

Theorem 2. *Assume that $\mathcal{B}_\psi^{n_0}(\mathcal{P}) < \infty$ for some $n_0 \in \mathbb{N}$. If there exists a Q-projection path $\{\mu_\theta\}_\theta \subset \mathcal{P}$ differentiable at μ, then*

$$\lim_{n\to\infty} \mathcal{B}_\psi^n(\mathcal{P}) = \mathcal{B}_\psi(\mathcal{P}).$$

This result is not surprising as we know by the previous lemma that the efficiency bound only depends asymptotically on the behavior of the model in the neighborhood of μ. We emphasize that, in a parametric model $\{\mu_\theta\}_\theta$, the efficiency bound has a positive limit $\mathcal{B}_\psi^\infty$ in non-trivial cases as soon as the map $\theta \mapsto \sqrt{\mathcal{Q}(\mu_\theta|\mu)}$ is differentiable at θ_0, while the construction of the Cramér–Rao bound requires the much stronger condition of differentiability in $\mathbb{L}^2(\mu)$. Thus, the asymptotic efficiency bound $\mathcal{B}_\psi^\infty$ is computable in a larger class of models, while providing at least as good an asymptotic analysis as the Cramér–Rao inequality in smooth models.

5. Proofs

Proof of Theorem 1. If ν is not absolutely continuous w.r.t. μ, the inequality is trivially verified. Otherwise, remark that the unbiasedness condition guarantees the following equality

$$\psi(\mu) - \psi(\nu) = \mathbb{E}\left(\left(T - \psi(\mu)\right)\left(1 - \frac{\mathrm{d}\nu^{\otimes n}}{\mathrm{d}\mu^{\otimes n}}\right)\right),$$

where the expectation is meant under the true distribution of the observations, $\mu^{\otimes n}$. Cauchy–Schwarz inequality yields $(\psi(\mu) - \psi(\nu))^2 \leqslant \mathrm{var}(T)\mathcal{Q}(\nu^{\otimes n}|\mu^{\otimes n})$ which, combined with the equality $\mathcal{Q}(\nu^{\otimes n}|\mu^{\otimes n}) = (\mathcal{Q}(\nu|\mu) + 1)^n - 1$, gives

$$\mathrm{var}(T) \geqslant \frac{(\psi(\mu) - \psi(\nu))^2}{(\mathcal{Q}(\nu|\mu) + 1)^n - 1}. \qquad \square$$

Proof of Proposition 1. First remark that if $\{\mu_\theta\}_{\theta\in\Theta}$ is differentiable in $\mathbb{L}^2(\mu)$ at $\mu = \mu_{\theta_0}$ with score g, the limit as $\theta \to \theta_0$ of $\mathcal{Q}(\mu_\theta|\mu)/(\theta - \theta_0)^2$ exists and is equal to the Fisher information $\int g^2\,\mathrm{d}\mu$. In particular, we have for a fixed $n \in \mathbb{N}$,

$$\left(\mathcal{Q}(\mu_\theta|\mu) + 1\right)^n - 1 = n\mathcal{Q}(\mu_\theta|\mu) + o\left(|\theta - \theta_0|^2\right).$$

Hence,

$$\lim_{\theta\to\theta_0} \mathcal{H}_\psi^n(\mu_\theta|\mu) = \lim_{\theta\to\theta_0} \frac{(\psi(\mu) - \psi(\mu_\theta))^2}{(\theta - \theta_0)^2} \frac{(\theta - \theta_0)^2}{\mathcal{Q}(\mu_\theta|\mu)} = \mathcal{B}_\psi\left(\{\mu_\theta\}_\theta\right). \qquad \square$$

Proof of Theorem 2. The theorem is true if $\mathcal{B}_\psi^\infty = 0$. Now, assume that $\mathcal{B}_\psi^\infty > 0$. We know that $\mathcal{B}_\psi^n(\mathcal{P}) = \mathcal{B}_\psi^n(\{\mu_\theta\}_\theta)$ for all $n \in \mathbb{N}$. Let $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of measures in $\{\mu_\theta\}_\theta$, suitably chosen so that $\lim_{n\to\infty} \mathcal{H}_\psi^n(\mu_n|\mu) = \mathcal{B}_\psi^\infty$. We want to prove that $\mathcal{Q}(\mu_n|\mu)$ tends to 0 as $n \to \infty$. By contradiction, if there exists $\varepsilon > 0$ and an increasing sequence of integers $\{n_k\}_{k\in\mathbb{N}}$ such that $\forall k \in \mathbb{N},\ \mathcal{Q}(\mu_{n_k}|\mu) > \varepsilon$, then:

$$\mathcal{H}_\psi^{n_k}(\mu_{n_k}|\mu) \leqslant \sup_{\mathcal{Q}(\nu|\mu)>\varepsilon} \mathcal{H}_\psi^{n_k}(\nu|\mu) \longrightarrow 0, \quad \text{as } k \to \infty,$$

by Lemma 1, which conflicts with the fact that $\lim_{k\to\infty} \mathcal{H}_\psi^{n_k}(\mu_{n_k}|\mu) = \mathcal{B}_\psi^\infty > 0$. So, we conclude that $\lim_{n\to\infty} \mathcal{Q}(\mu_n|\mu) = 0$. Since $\mathcal{H}_\psi^n(.|\mu)$ is pointwise decreasing as $n \to \infty$, we get that for all $n_0 \in \mathbb{N}$,

$$\mathcal{B}_\psi^\infty = \lim_{n\to\infty} \mathcal{H}_\psi^n(\mu_n|\mu) \leqslant \lim_{\theta\to\theta_0} \mathcal{H}_\psi^{n_0}(\mu_\theta|\mu) = \mathcal{B}_\psi\left(\{\mu_\theta\}_\theta\right).$$

So, $\{\mu_\theta\}_\theta$ is a least favorable path of the model and therefore satisfies $\mathcal{B}_\psi(\mathcal{P}) = \mathcal{B}_\psi(\{\mu_\theta\}_\theta)$, yielding $\mathcal{B}_\psi^\infty(\mathcal{P}) \leqslant \mathcal{B}_\psi(\mathcal{P})$. The reverse inequality being an obvious consequence of Proposition 1, we conclude that $\mathcal{B}_\psi^\infty(\mathcal{P}) = \mathcal{B}_\psi(\mathcal{P})$. $\square$

References

[1] P.J. Bickel, C. Klaassen, Y. Ritov, J. Wellner, Efficient and Adaptive Estimation for Semiparametric Models, Springer-Verlag, New York, 1998.
[2] I. Csiszár, On topology properties of f-divergences, Studia Sci. Math. Hungar. 2 (1967) 329–339.
[3] T. Espinasse, P. Rochet, A Cramér–Rao inequality for non-differentiable models, arXiv:1204.2763, 2012.
[4] J. Hájek, A characterization of limiting distributions of regular estimates, Z. Wahr. Verw. Geb. 14 (1969/1970) 323–330.
[5] L. Le Cam, Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses, Univ. California Publ. Statist. 3 (1960) 37–98.
[6] C. Radhakrishna Rao, Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. 37 (1945) 81–91.
[7] A. van der Vaart, Asymptotic Statistics, Cambridge University Press, Cambridge, 1998.
[8] A. van der Vaart, Semiparametric statistics, in: Saint-Flour, 1999, Lectures on Probability Theory and Statistics 1781 (2002) 331–457.