Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 336 (2003) 857–862
COMPTES RENDUS MATHEMATIQUE

Statistics/Probability Theory

Dual representation of ϕ-divergences and applications

Amor Keziou

LSTA, boîte courrier 158, 8A, Université Paris-6, 175, rue du Chevaleret, 75013 Paris, France

Received 20 February 2003; accepted 15 April 2003

Presented by Paul Deheuvels

Abstract

In this Note, we give a “dual” representation of divergences. We make use of this representation to define and study some new estimates of the law and of the divergences for discrete and continuous parametric models. ***To cite this article: A. Keziou, C. R. Acad. Sci. Paris, Ser. I 336 (2003).***

Résumé

Représentation duale des ϕ-divergences et applications. Dans cette Note, nous donnons une représentation « duale » des divergences. Nous utilisons cette représentation pour définir et étudier de nouveaux estimateurs de la loi et des divergences pour des modèles paramétriques discrets et continus. ***Pour citer cet article : A. Keziou, C. R. Acad. Sci. Paris, Ser. I 336 (2003).***

Version française abrégée

Soit $(\mathcal{X}, \mathcal{B})$ un espace mesurable. Soit φ une fonction convexe définie sur $[0, +\infty)$ dans $[0, +\infty]$ satisfant $\varphi(1) = 0$ et $\varphi(0) = \lim_{x\to 0^+} \varphi(x)$. Soit P une loi de probabilité définie sur $(\mathcal{X}, \mathcal{B})$. Notons M^1 l'espace des lois de probabilité définies sur $(\mathcal{X}, \mathcal{B})$ et notons $M^1(P)$ le sous-espace des lois de probabilité absolument continues par rapport à P (a.c. p.r.à. P). Pour toute loi de probabilité $Q \in M^1(P)$, la ϕ-divergence entre Q et P est définie par

$$\phi(Q, P) := \int \varphi\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}P. \tag{1}$$

Lorsque Q n'est pas a.c. p.r.à. P, on pose $\phi(Q, P) := +\infty$. Cette définition a été introduite par Rüschendorf [8], et elle est la version modifiée de la définition originale introduite par Csiszar [3], qui nécessite une mesure dominante commune σ-finie λ pour la loi P et les lois Q. Comme nous allons considérer tout l'espace $M^1(P)$, il convient d'utiliser la définition (1). Notons que les deux définitions coïncident sur le sous-espace des lois de probabilité a.c. p.r.à. P et dominées par la mesure σ-finie λ.

Les divergences de Kullback–Leibler (KL), Kullback–Leibler modifiée (KL_m) et Hellinger (H) sont obtenues respectivement pour $\varphi(x) = -\log(x) + x - 1$, $\varphi(x) = x\log(x) - x + 1$ et $\varphi(x) = 2(\sqrt{x} - 1)^2$. Ces divergences

E-mail address: keziou@ccr.jussieu.fr (A. Keziou).

1631-073X/03/$ – see front matter
doi:10.1016/S1631-073X(03)00215-2

font partie de la classe des divergences de puissance introduite par Cressie et Read (cf. [2] et Liese et Vajda [5], Chapitre 2).

Soit $\{P_\theta,\ \theta \in \Theta\}$ un modèle paramétrique identifiable avec Θ un sous-ensemble de $\mathbb{R}^d$. On considère le problème d'estimation de la vraie valeur inconnue θ_0 du paramètre θ et l'estimation des divergences $\phi(P_\alpha, P_{\theta_0})$ à partir d'un échantillon $X_1, \ldots, X_n$ de loi P_{θ_0}. On suppose que le support S des lois P_θ ne dépend pas de θ. Si S est discret fini, on a $\phi(P_\theta, P_{\theta_0}) = \sum_{i\in S} \varphi\big(\frac{P_\theta(i)}{P_{\theta_0}(i)}\big) P_{\theta_0}(i)$; Pour ces modèles, Lindsay [6] et Morales, Pardo et Vajda [7] ont introduit « les estimateurs de minimum des ϕ-divergences » (EMϕ's) définis par

$$\hat{\theta}_n := \arg\inf_{\theta\in\Theta} \phi(P_\theta, P_n), \tag{2}$$

où $\phi(P_\theta, P_n)$ est l'estimateur « plug-in » de $\phi(P_\theta, P_{\theta_0})$

$$\phi(P_\theta, P_n) = \sum_{i\in S} \varphi\left(\frac{P_\theta(i)}{P_n(i)}\right) P_n(i), \tag{3}$$

et P_n est la mesure empirique construite à partir de l'échantillon $X_1, \ldots, X_n$. L'estimateur du maximum de vraisemblance (EMV) est obtenu pour $\varphi(x) = -\log(x) + x - 1$.

Les estimateurs (2) de θ_0 et les estimateurs (3) des ϕ-divergences ne sont pas définis si le support S n'est pas discret ; dans Broniatowski [1], une nouvelle méthode d'estimation est proposée dans le cas continu pour estimer la divergence de Kullback–Leibler ; il utilise la représentation duale bien connue de la divergence de Kullback–Leibler comme la transformée de Fenchel–Legendre de la fonction génératrice des moments. En étendant [1], nous donnons dans cette Note une nouvelle représentation générale pour l'ensemble des ϕ-divergences. Nous obtenons cette représentation par application du lemme de dualité (cf. Dembo et Zeitouni [4], Lemme 4.5.8). Cette représentation permet de définir les estimateurs de minimum des ϕ-divergences lorsque le support S n'est pas nécessairement discret. On présente le comportement asymptotique de ces estimateurs.

1. Introduction and notations

Let $(\mathcal{X}, \mathcal{B})$ be a measurable space. Let φ be a non-negative convex function defined on $[0, +\infty)$ in $[0, +\infty]$ and satisfying $\varphi(1) = 0$ and $\varphi(0) = \lim_{x\to 0^+} \varphi(x)$. Let P be a probability mesures (p.m.) defined on $(\mathcal{X}, \mathcal{B})$. Denote by M^1 the whole space of p.m.'s defined on $(\mathcal{X}, \mathcal{B})$ and denote by $M^1(P)$ the subspace of p.m.'s absolutely continuous (a.c.) w.r.t. P. For all p.m. $Q \in M^1(P)$. The ϕ-divergence between Q and P is defined by

$$\phi(Q, P) := \int \varphi\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}P. \tag{4}$$

When Q is not a.c. w.r.t. P, we set $\phi(Q, P) := +\infty$. This definition has been introduced by Rüschendorf [8]. It is the modified version of the orginal definition introduced by Csiszar [3]; his definition requires a common σ-finite dominating measure λ for P and for the p.m.'s Q. Since we will consider the whole space $M^1(P)$, it is more convenient to use definition (4). Note that both definitions coincide on the subspace of p.m.'s a.c. w.r.t. P and dominated by λ.

The Kullback–Leibler divergence (*KL*), modified Kullback–Leibler divergence (KL_m) and Hellinger divergence (H) are obtained respectively for $\varphi(x) = x\log(x) - x + 1$, $\varphi(x) = -\log(x) + x - 1$ and $\varphi(x) = 2(\sqrt{x} - 1)^2$. All these examples of divergences are pecular cases of the so-called "power divergences", introduced by Cressie and Read (cf. [2] and [5], Chapter 2).

Let $\{P_\theta,\ \theta \in \Theta\}$ be a parametric identifiable model defined on $(\mathcal{X}, \mathcal{B})$ with Θ a subset of $\mathbb{R}^d$. Let $X_1, \ldots, X_n$ be an i.i.d. sample with common unknown distribution P_{θ_0}. We consider the estimation problem of θ_0 the true unknown value of the parameter and the estimation problem of the divergences $\phi(P_\alpha, P_{\theta_0})$. If all p.m.'s P_θ have the same discrete finite support S, we have $\phi(P_\theta, P_{\theta_0}) = \sum_{i\in S} \varphi\big(\frac{P_\theta(i)}{P_{\theta_0}(i)}\big) P_{\theta_0}(i)$; For such models, Lindsay [6] and Morales, Pardo and Vajda [7] introduced the so-called "Minimum ϕ-divergences estimates" (MϕE's) defined by

$$\hat{\theta}_n := \arg\inf_{\theta\in\Theta} \phi(P_\theta, P_n), \tag{5}$$

where $\phi(P_\theta, P_n)$ is the "plug-in" estimate of $\phi(P_\theta, P_{\theta_0})$

$$\phi(P_\theta, P_n) = \sum_{i\in S} \varphi\left(\frac{P_\theta(i)}{P_n(i)}\right) P_n(i), \tag{6}$$

and P_n is the empirical measure associated to the sample $X_1, \ldots, X_n$. The maximum likelihood estimate (MLE) is obtained for $\varphi(x) = -\log(x) + x - 1$.

The estimates (5) and (6) are not defined when the support S is continuous; in Broniatowski [1], a new estimation procedure is proposed in order to estimate the *KL*-divergence between some set of p.m's Ω and some continuous p.m. P, without making use of any partitioning nor smoothing, but merely making use of the well known dual representation of the *KL*-divergence as the Fenchel–Legendre transform of the moment generating function. Extending the paper by Broniatowski [1], we expose a general representation for ϕ-divergences. This is obtained through the duality lemma, whose proof can be found for example in (Dembo and Zeitouni [4], Lemma 4.5.8, Chapter 4). We make use of this representation to define some new estimates of the parameter θ_0 which we will call "minimum dual ϕ-divergences estimates" (MDϕE's) where the p.m's P_θ do not necessarily have discrete finite supports. Also the same representation will be used in order to estimate $\phi(P_{\alpha_0}, P_{\theta_0})$ which leads to various parametric tests.

2. Results

Let M be the space of all finite signed measures defined on $(\mathcal{X}, \mathcal{B})$. We also consider a class $\mathcal{F}$ of measurable real valued functions f defined on $\mathcal{X}$, and we assume that $\mathcal{F}$ contains $\mathcal{M}_b$, the set of all bounded measurable functions defined on $\mathcal{X}$. We will denote by $\langle\mathcal{F}\rangle$ the linear span of $\mathcal{F}$, φ' the derivative function of φ and $\overleftarrow{\varphi'}$ the inverse function of φ'. We will sometimes write Pf for $\int f\,\mathrm{d}P$ for any measure P and any function f. Define $M_\mathcal{F} := \{Q \in M \mid \int |f|\,\mathrm{d}|Q| < \infty,\ \forall f \in \mathcal{F}\}$. We extend the definition in (4) on the whole space $M_\mathcal{F}$ by stating $\varphi(x) = +\infty$ for negative values of x. We equip the linear space $M_\mathcal{F}$ with the $\tau_\mathcal{F}$-topology, which is the coarsest topology for which all mappings $Q \in M \to \int f\,\mathrm{d}Q \in \mathbb{R}$ are continuous for all f in $\langle\mathcal{F}\rangle$.

Proposition 2.1. *$M_\mathcal{F}$ equiped with the $\tau_\mathcal{F}$-topology is a locally convex Hausdorf topological linear space and the topological dual space of $M_\mathcal{F}$ is the set of all mappings $Q \to \int f\,\mathrm{d}Q$ when f belongs to $\langle\mathcal{F}\rangle$. Further the divergence functions $Q \to \phi(Q, P)$ from $(M_\mathcal{F}, \tau_\mathcal{F})$ onto $(-\infty, +\infty]$ are lower semi-continuous* (*l.s.c.*).

According to this proposition, the Fenchel–Legendre transform of $Q \to \phi(Q, P)$ is defined for any f in $\langle\mathcal{F}\rangle$ by $T(f, P) := \sup_{Q\in M_\mathcal{F}}\{\int f\,\mathrm{d}Q - \phi(Q, P)\}$. The conditions in duality lemma hold for the functions $Q \to \phi(Q, P)$ and the topological dual space of $M_\mathcal{F}$ is one to one with $\langle\mathcal{F}\rangle$. Hence by application of the duality lemma, we state

Proposition 2.2. *For any measure Q in $M_\mathcal{F}$ and for any p.m. P, it holds*

$$\phi(Q, P) = \sup_{f\in\langle\mathcal{F}\rangle}\left\{\int f\,\mathrm{d}Q - T(f, P)\right\}.$$

From this proposition, using directional derivatives, we calculate $T(f, P)$, and we obtain

Theorem 2.1. *Assume that the function φ is srtictly convex and is $\mathcal{C}^2$ on $(0,+\infty)$. Let Q and P be two p.m's with Q a.c. w.r.t. P and $\phi(Q,P) \leq \infty$. Let $\mathcal{F}$ be a class of functions such that $\varphi'(\mathrm{d}Q/\mathrm{d}P)$ belongs to $\mathcal{F}$, for all f in $\mathcal{F}$, $\int |f|\,\mathrm{d}Q$ is finite and $\overleftarrow{\varphi'}(f(x))$ is defined for all $x \in \mathcal{X}$. Then, the divergence $\phi(Q,P)$ admits the "dual representation"*

$$\phi(Q,P) = \sup_{f\in\mathcal{F}} \left\{ \int f\,\mathrm{d}Q - \int f\overleftarrow{\varphi'}(f) - \varphi\big(\overleftarrow{\varphi'}(f)\big)\,\mathrm{d}P \right\}. \tag{7}$$

The supremum in (7) *is unique* (P*-a.s.*) *and is reached at* $f = \varphi'(\mathrm{d}Q/\mathrm{d}P)$ (P*-a.s.*).

2.1. Definition of estimates through dual representation

We assume that the function φ is strictly convex and is $\mathcal{C}^2$ on $(0,+\infty)$. We assume that for any $\theta \in \Theta$, P_θ has density p_θ with respect to some dominating σ-finite measure λ. We assume also that for any α in Θ, the following condition holds (C.0): $\int |\varphi'(p_\alpha/p_\theta)|\,\mathrm{d}P_\alpha(x) < \infty$ for any $\theta \in \Theta$. This condition is fulfilled if $\phi(P_\alpha, P_\theta) := \int \varphi(p_\alpha/p_\theta)\,\mathrm{d}P_\theta < \infty$ for any $\theta \in \Theta$ and φ fulfills the condition of Lemma 8.7 in Liese and Vajda [5] (see [5], Lemma 8.9). Consider the class of functions $\mathcal{F}$ defined by $\mathcal{F} := \{x \to \varphi'(p_\alpha(x)/p_\theta(x)),\ \theta \in \Theta\}$. By Theorem 2.1, we obtain

$$\begin{aligned}
&\phi(P_\alpha, P_{\theta_0}) = \sup_{f\in\mathcal{F}} \left\{ \int f\,\mathrm{d}P_\alpha - \int f\overleftarrow{\varphi'}(f) - \varphi\big(\overleftarrow{\varphi'}(f)\big)\,\mathrm{d}P_{\theta_0} \right\}, \quad \text{i.e.,}\\
&\phi(P_\alpha, P_{\theta_0}) = \sup_{\theta\in\Theta} P_{\theta_0} m(\theta,\alpha), \quad \text{with } m(\theta,\alpha): x \to m(\theta,\alpha,x) \quad \text{and}\\
&m(\theta,\alpha,x) := \int \varphi'\left(\frac{p_\alpha}{p_\theta}\right)\mathrm{d}P_\alpha - \left\{ \varphi'\left(\frac{p_\alpha}{p_\theta}(x)\right)\frac{p_\alpha}{p_\theta}(x) - \varphi\left(\frac{p_\alpha}{p_\theta}(x)\right) \right\}.
\end{aligned} \tag{8}$$

Remark 1. The function $\theta \to P_{\theta_0} m(\theta,\alpha)$ has a unique maximizer $\theta = \theta_0$. See Theorem 2.1.

For all $\alpha \in \Theta$, define the what we call "dual ϕ-divergences estimates" (DϕE's) of θ_0 by

$$\hat{\theta}_n(\alpha) := \arg\sup_{\theta\in\Theta} P_n m(\theta,\alpha). \tag{9}$$

The divergence $\phi(P_\alpha, P_{\theta_0})$ between P_α and P_{θ_0} can be estimated by

$$\hat{\phi}_n(P_\alpha, P_{\theta_0}) := P_n m\big(\hat{\theta}_n(\alpha),\alpha\big) = \sup_{\theta\in\Theta} P_n m(\theta,\alpha). \tag{10}$$

Further we have $\inf_{\alpha\in\Theta} \phi(P_\alpha, P_{\theta_0}) = \phi(P_{\theta_0}, P_{\theta_0}) = 0$. The infimum in the above display is unique when φ is strictly convex on a neighborhood of 1, and it is achieved at $\alpha = \theta_0$. It follows that a natural definition of estimates of θ_0, which we call "minimum dual ϕ-divergences estimates" (MDϕE's), is

$$\hat{\alpha}_n := \arg\inf_{\alpha\in\Theta} \hat{\phi}_n(P_\alpha, P_{\theta_0}) = \arg\inf_{\alpha\in\Theta}\sup_{\theta\in\Theta} P_n m(\theta,\alpha). \tag{11}$$

Remark 2. The maximum likelihood estimate (*MLE*) belongs to these class of esitmates. Indeed it is obtained when $\varphi(x) = -\log(x) + x - 1$, that is as the dual modified *KL*-divergence estimate or as the minimum dual modified *KL*-divergence estimate, i.e., $MLE = DKL_m E = MDKL_m E$. Indeed we have for $\varphi(x) = -\log(x) + x - 1$, $P_n m(\theta,\alpha) = -\int \log(p_\alpha/p_\theta)\,\mathrm{d}P_n$, hence from definitions (9) and (11), we get $\hat{\theta}_n(\alpha) = \hat{\theta}_n := \arg\sup_{\theta\in\Theta} -\int \log(P_\alpha/P_\theta)\,\mathrm{d}P_n = \arg\inf_{\alpha\in\Theta} -\int \log(P_\alpha/P_{\hat{\theta}_n})\,\mathrm{d}P_n := \hat{\alpha}_n = MLE$.

2.2. The asymptotic behaviour of the estimates $\hat{\theta}_n(\alpha)$ and $\hat{\phi}_n(P_\alpha, P_{\theta_0})$ for fixed α in Θ

In this section we state the asymptotic normality of the estimates $\hat{\theta}_n(\alpha)$ and evaluate their limiting variance. The hypotheses handled here are similar to those used in ([9], Chapter 5) in the study of M-estimates. Notice that indeed for fixed α, $\hat{\theta}_n(\alpha)$ are M-estimates. We state also the asymptotic behaviour of the estimates $\hat{\phi}_n(P_\alpha, P_{\theta_0})$. Denote by $m'(\theta,\alpha)$ the d-dimensional vector with entries $\frac{\partial}{\partial\theta_i}m(\theta,\alpha)$ and by $m''(\theta,\alpha)$ the $d\times d$-matrix with entries $\frac{\partial^2}{\partial\theta_i\partial\theta_j}m(\theta,\alpha)$. In the sequel we will assume that condition (C.0) holds, $\phi(P_\alpha, P_{\theta_0}) < \infty$ and that the estimates $\hat{\theta}_n(\alpha)$ exist. Define the function $x \to g(\theta,\alpha,x) := \varphi'(p_\alpha(x)/p_\theta(x))p_\alpha(x)$, and denote by $\|\cdot\|$ the Euclidian norm and by I_{θ_0} the information matrix, i.e., $I_{\theta_0} = \int \dot{p}_{\theta_0}\dot{p}_{\theta_0}^t/p_{\theta_0}\,\mathrm{d}\lambda$ where $\dot{p}_{\theta_0}$ is the gradient of p_{θ_0}. We will consider the following conditions

(C.1) $\hat{\theta}_n(\alpha)$ converges in probability to θ_0.

(C.2) The function φ is $\mathcal{C}^3$ and there exists a neighborhood $V(\theta_0)$ of θ_0 such that for all θ in $V(\theta_0)$, the gradient $\dot{p}_\theta$ and the Hessian matrix $\ddot{p}_\theta$ of p_θ exist (λ-a.e.), the partial derivatives of order 1 of p_θ and the partial derivatives of ordre 1 and 2 of $\theta \to g(\theta,\alpha,x)$ are dominated (λ-a.e.) by some λ-integrable functions.

(C.3) The function $\theta \to m(\theta,\alpha,x)$ is $\mathcal{C}^3$ on a neighborhood $V(\theta_0)$ of θ_0 for all x and all partial derivatives of order 3 of $\theta \to m(\theta,\alpha,x)$ are dominated on $V(\theta_0)$ by some P_{θ_0}-integrable function $x \to H(x)$.

(C.4) $P_{\theta_0}\|m'(\theta_0,\alpha)\|^2 < \infty$ and the matrix $P_{\theta_0}m''(\theta_0,\alpha)$ exists and is invertible.

Theorem 2.2. *Assume that conditions* (C.1)–(C.4) *hold. Then*

(1) (a) $\sqrt{n}(\hat{\theta}_n(\alpha)-\theta_0)$ *converges in distribution to a centered normal variable with covariance matrix*

$$V = \left[-P_{\theta_0}m''(\theta_0,\alpha)\right]^{-1}P_{\theta_0}m'(\theta_0,\alpha)m'(\theta_0,\alpha)^t\left[-P_{\theta_0}m''(\theta_0,\alpha)\right]^{-1}. \tag{12}$$

(b) *If* $\alpha=\theta_0$, *then* $-P_{\theta_0}m''(\theta_0,\alpha) = \frac{1}{\varphi''(1)}P_{\theta_0}m'(\theta_0,\alpha)m'(\theta_0,\alpha)^t$ *and* $V = I_{\theta_0}{}^{-1}$.

(2) *If* $\alpha=\theta_0$, *then the statistics* $\frac{2n}{\varphi'(1)}\hat{\phi}_n(P_\alpha,P_{\theta_0})$ *converge in distribution to a* χ^2 *variable with* d *degrees of freedom.*

(3) *If* $\alpha\neq\theta_0$, *then* $\sqrt{n}(\hat{\phi}_n(P_\alpha,P_{\theta_0})-\phi(P_\alpha,P_{\theta_0}))$ *converges in distribution to a centered normal variable with variance* $\sigma^2 = P_{\theta_0}m(\theta_0,\alpha)^2-(P_{\theta_0}m(\theta_0,\alpha))^2$.

Remark 3. Using Theorem 2.2, the estimates $\hat{\phi}_n(P_{\alpha_0},P_{\theta_0})$ can be used to perform a test of a hypothese H_0: $\theta_0=\alpha_0$ against the alternatives H_1: $\theta_0\neq\alpha_0$ for some knouwn value α_0. Those statistics $\hat{\phi}_n(P_{\alpha_0},P_{\theta_0})$, from Theorem 2.2, are n-consistent estimates of $\phi(P_{\alpha_0},P_{\theta_0})=0$ under H_0 and $\sqrt{n}$-consistent estimates of $\phi(P_{\alpha_0},P_{\theta_0})$ under H_1. Since $\phi(P_{\alpha_0},P_{\theta_0})$ is positive and takes value 0 only when $\theta_0=\alpha_0$, the tests are defined through the critical region $CR_\phi := \{\frac{2n}{\varphi''(1)}\hat{\phi}_n(P_{\alpha_0},P_{\theta_0}) > q_\alpha\}$ where q_α is the α-quantile of the χ^2 distribution with d degrees of freedom. Also these tests are all asymptotically powerful since the estimates $\hat{\phi}_n(P_{\alpha_0},P_{\theta_0})$ are $\sqrt{n}$-consistent under H_1. When $\varphi(x)=-\log(x)+x-1$, we obtain the critical region $CR_{KL_m} := \{2n\sup_{\theta\in\Theta}P_n\log(p_\theta/p_{\alpha_0}) > q_\alpha\}$ which is to say that the test is precisely the likelihood ratio test. Note that, in the discrete case, the test performed from the statistic $KL_m(P_{\alpha_0},P_n)$ defined in (6) is different from the likelihood ratio test.

2.3. The asymptotic behaviour of the estimates $\hat{\alpha}_n$ and $\hat{\theta}_n(\hat{\alpha}_n)$

In this section we state the limiting distributions of the estimates $\hat{\theta}_n(\hat{\alpha}_n)$ and of the MDϕE's $\hat{\alpha}_n$ of θ_0 defined in (11), we show that the MDϕE's are all asymptotically efficient. We assume that condition (C.0) is fulfilled, there exists a neighborhood $V(\theta_0)$ of θ_0 such that $\phi(P_\alpha,P_{\theta_0})<\infty$ for all $\alpha\in V(\theta_0)$ and that both estimates $\hat{\theta}_n(\hat{\alpha}_n)$ and $\hat{\alpha}_n$ exist.

We will make use of the following conditions

(C.5) Both estimates $\hat{\alpha}_n$ and $\hat{\theta}_n(\hat{\alpha}_n)$ converge in probability to θ_0.

(C.6) The function φ is $\mathcal{C}^3$ and there exists a neighborhood $V(\theta_0,\theta_0)$ of (θ_0,θ_0) such that for all (θ,α) in $V(\theta_0,\theta_0)$, the gradient $\dot{p}_\theta$ and the Hessian matrix $\ddot{p}_\theta$ exist (λ-a.e.), the partial derivatives of order 1 of p_θ and the partial derivatives of ordre 1 and 2 of $(\theta,\alpha)\to g(\theta,\alpha,x)$ are dominated (λ-a.e.) by some λ-integrable functions.

(C.7) The function $(\theta,\alpha)\to m(\theta,\alpha,x)$ is $\mathcal{C}^3$ on some neighborhood $V(\theta_0,\theta_0)$ of (θ_0,θ_0) for all x and the partial derivatives of order 3 of $(\theta,\alpha)\to m(\theta,\alpha,x)$ are all dominated on $V(\theta_0,\theta_0)$ by some P_{θ_0}-integrable function $x\to H(x)$.

(C.8) $P_{\theta_0}\|\frac{\partial}{\partial\theta}m(\theta_0,\theta_0)\|^2<\infty$, $P_{\theta_0}\|\frac{\partial}{\partial\alpha}m(\theta_0,\theta_0)\|^2<\infty$ and the Information matrix I_{θ_0} exists and is invertible.

Theorem 2.3. *Assume that conditions* (C.5)–(C.8) *hold. Then both* $\sqrt{n}(\hat{\alpha}_n-\theta_0)$ *and* $\sqrt{n}(\hat{\theta}_n(\hat{\alpha}_n)-\theta_0)$ *converge in distribution to a centered normal variable with covariance matrix* $V={I_{\theta_0}}^{-1}$ *and the estimates* $\hat{\alpha}_n$ *and* $\hat{\theta}_n(\hat{\alpha}_n)$ *are asymptotically uncorrelated.*

Remark 4. Using theorem 5.7 in [9], we can give sufficient conditions for (C.1) and (C.5).

Acknowledgement

We wish to thank Professor Michel Broniatowski for his discussions and suggestions leading to improvement of this Note.

References

[1] M. Broniatowski, Estimation through Kullback–Leibler divergence, Math. Methods Statist. (2003) (submitted).
[2] N. Cressie, T. Read, Multinomial goodness-of-fit tests, J. Roy. Statist. Soc. Ser. B 46 (3) (1984) 440–464.
[3] I. Csiszar, On topological properties of f-divergences, Studia Sci. Math. Hungar. 2 (1967) 329–339.
[4] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Jones & Bartlett, 1998.
[5] F. Liese, I. Vajda, Convex Statistical Distances, Teubner, Leipzig, 1987.
[6] B.G. Lindsay, Efficiency versus robustness: the case for minimum Hellinger distance and related methods, Ann. Statist. 22 (1994) 1081–1114.
[7] D. Morales, L. Pardo, I. Vajda, Asymptotic divergence of estimates of discrete distributions, J. Statist. Plann. Inference 48 (3) (1995) 347–369.
[8] L. Rüschendorf, On the minimum discrimination information theorem, Statist. Decisions, Suppl. 1 (1984) 263–283.
[9] A. van der Vaart, Asymptotic Statistics, in: Cambridge Series in Statistical and Probabilitic Mathematics, 1998.