C. R. Acad. Sci. Paris, Ser. I 352 (2014) 749–754

Contents lists available at ScienceDirect

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

www.sciencedirect.com

Statistics

On Bayesian estimation via divergences

Sur l'estimation bayésienne via les divergences

Mohamed Cherfi

Laboratoire de mathématiques et applications, Département de mathématiques, Faculté des sciences, Université Hassiba-Benbouali de Chlef, 02000 Chlef, Algeria

A R T I C L E   I N F O

Article history:
Received 7 February 2012
Accepted after revision 28 June 2014
Available online 4 August 2014

Presented by the Editorial Board

A B S T R A C T

In this note, we introduce a new methodology for Bayesian inference through the use of ϕ-divergences and of the duality technique. The asymptotic laws of the estimates are established.

R É S U M É

Dans cette Note, nous introduisons une nouvelle méthodologie d'inférence bayésienne en utilisant les ϕ-divergences et la technique de dualité. Nous obtenons les lois asymptotiques des estimateurs.

1. Introduction

Bayesian techniques are particularly attractive since they can incorporate information other than the data into the model in the form of prior distributions. Another feature that makes them increasingly attractive is that they can handle models that are difficult to estimate with classical methods by use of simulation techniques, see for instance [24].

The aim of this note is to discuss the use of divergences as a basis for Bayesian inference. The use of divergence measures in a Bayesian context has been considered in [10] and [22]. Ragusa [23] used Bayesian ϕ-divergences in a Generalized Empirical Likelihood framework.

The misspecification of prior distributions, the presence of large outliers with respect to the specified model, may lead to unreliable posterior distributions for parameters in Bayesian inference. In order to estimate model parameters and circumvent possible difficulties encountered with the likelihood function, we follow up common robustification ideas, see for instance [11,12], and propose to replace the likelihood in the formula of the posterior distribution by the dual form of the divergence between a postulated parametric model and the empirical distribution. A major advantage of the method is that it does not require additional accessories such as kernel density estimation or other forms of nonparametric smoothing to produce nonparametric density estimates of the true underlying density function in contrast with the method proposed by Hooker and Vidyashankar [13], which is based on the concept of a minimum disparity procedure introduced by Lindsay

E-mail address: mohamed.cherfi@gmail.com.

http://dx.doi.org/10.1016/j.crma.2014.06.013

[20]. The plug-in of the empirical distribution function is sufficient for the purpose of estimating the divergence in the case of i.i.d. data.

The proposed estimators are based on integration rather than on optimization. This is particularly an issue when the parameter space is "large", since the search has to be done over a large-dimensional space. Other reasons, which are commonly put forward to use the proposed approach, are their computational attractiveness through the use of Markov chain Monte Carlo (MCMC), see [25], and the fact they can easily handle a large number of parameters.

The outline of the note is as follows. Together with a brief review of definitions and properties of divergences, Section 2 discusses the procedure to obtain the estimates. In Section 3, we give the limit laws of the proposed estimators. Some final remarks conclude the note.

2. Estimation

2.1. Background on dual divergences inference

Keziou [15] and Broniatowski and Keziou [5] introduced the class of dual divergences estimators for general parametric models. In the following, we shortly recall their context and definition.

Recall that the ϕ-divergence between a bounded signed measure Q and a probability measure (p.m.) P on $\mathscr{D}$, when Q is absolutely continuous with respect to P, is defined by

$$D_\phi(Q,P) := \int_{\mathscr{D}} \phi\left(\frac{\mathrm{d}Q}{\mathrm{d}P}(x)\right)\mathrm{d}P(x),$$

where ϕ is a convex function from $]-\infty,\infty[$ to $[0,\infty]$ with $\phi(1)=0$.

Different choices for ϕ have been proposed in the literature. For a good overview, see [21]. A well-known class of divergences is the class of the so-called "power divergences" introduced by Cressie and Read [9] (see also [18], Chapter 2); it contains the most known and used divergences. They are defined through the class of convex functions

$$x \in]0,+\infty[\mapsto \phi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma-1)} \tag{1}$$

if $\gamma \in \mathbb{R}\setminus\{0,1\}$, $\phi_0(x) := -\log x + x - 1$ and $\phi_1(x) := x\log x - x + 1$.

Let $X_1,\dots,X_n$ be an i.i.d. sample and $\mathbb{P}_{\theta_0}$ the true p.m. underlying the data. Consider the problem of estimating the population parameters of interest θ_0, when the underlying identifiable model is given by $\{\mathbb{P}_\theta : \theta \in \Theta\}$ with Θ a subset of $\mathbb{R}^d$. Here the attention is restricted to the case where the probability measures $\mathbb{P}_\theta$ are absolutely continuous with respect to the same σ-finite measure λ; the correspondent densities are denoted p_θ.

Let ϕ be a function of class $\mathcal{C}^2$, strictly convex satisfying

$$\int \left|\phi'\left(\frac{p_\theta(x)}{p_\alpha(x)}\right)\right| p_\theta(x)\,\mathrm{d}x < \infty. \tag{2}$$

By Lemma 3.2 in [4], if the function ϕ satisfies the following condition: there exists $0<\eta<1$ such that for all c in $[1-\eta,1+\eta]$, we can find numbers c_1, c_2, c_3 such that

$$\phi(cx) \le c_1\phi(x) + c_2|x| + c_3, \quad \text{for all real } x, \tag{3}$$

then the assumption (2) is satisfied whenever $D_\phi(\mathbb{P}_\theta,\mathbb{P}_\alpha)$ is finite. From now on, $\mathcal{U}$ will be the set of θ and α such that $D_\phi(\mathbb{P}_\theta,\mathbb{P}_\alpha)<\infty$. Note that all the real convex functions ϕ_γ pertaining to the class of power divergences defined in (1) satisfy condition (3).

Under (2), using Fenchel's duality technique, the divergence $D_\phi(\mathbb{P}_\theta,\mathbb{P}_{\theta_0})$ can be represented as resulting from an optimization procedure; this elegant result was proven in [15,19] and [5]. Broniatowski and Keziou [4] called it the dual form of a divergence, due to its connection with convex analysis.

Under the above conditions, the ϕ-divergence:

$$D_\phi(\mathbb{P}_\theta,\mathbb{P}_{\theta_0}) = \int \phi\left(\frac{p_\theta(x)}{p_{\theta_0}(x)}\right)p_{\theta_0}(x)\,\mathrm{d}x,$$

can be represented as the following form:

$$D_\phi(\mathbb{P}_\theta,\mathbb{P}_{\theta_0}) = \sup_{\alpha\in\mathcal{U}} \int h(\theta,\alpha)\,\mathrm{d}\mathbb{P}_{\theta_0}, \tag{4}$$

where $h(\theta,\alpha): x \mapsto h(\theta,\alpha,x)$, $\forall x \in \mathbb{R}$ and

$$h(\theta,\alpha,x) := \int \phi'\left(\frac{p_\theta}{p_\alpha}\right)p_\theta - \left[\frac{p_\theta(x)}{p_\alpha(x)}\phi'\left(\frac{p_\theta(x)}{p_\alpha(x)}\right) - \phi\left(\frac{p_\theta(x)}{p_\alpha(x)}\right)\right]. \tag{5}$$

Since the supremum in (4) is unique and is attained in $\alpha = \theta_0$, independently upon the value of θ, by replacing the hypothetical probability measure $\mathbb{P}_{\theta_0}$ by the empirical measure $\mathbb{P}_n$ define the class of estimators of θ_0 by

$$\widehat{\alpha}_\phi(\theta) := \arg\sup_{\alpha\in\mathcal{U}} \int h(\theta,\alpha)\,\mathrm{d}\mathbb{P}_n, \quad \theta\in\Theta, \tag{6}$$

where $h(\theta,\alpha)$ is the function defined in (5). This class is called "dual ϕ-divergence estimators" (DϕDE's), see for instance [15] and [5].

Formula (6) defines a family of M-estimators indexed by the function ϕ specifying the divergence and by some instrumental value of the parameter θ, called here escort parameter, see also [6].

Application of dual representation of ϕ-divergences has been considered by many authors—we cite, among others, Keziou and Leoni-Aubin [16] for semi-parametric two-sample density ratio models, robust tests based on saddlepoint approximations in [29,28] have proved that this class contains robust and efficient estimators and proposed robust test statistics based on divergences estimators. Bootstrapped ϕ-divergences estimates are considered in [1]; the extension of dual ϕ-divergences estimators to right censored data are introduced in [7]; the application to tail index estimation is presented in [2], for estimation and tests in copula models we refer to [3] and the references therein.

2.2. Estimation

Let us now turn to the estimation using divergences in our setting. For parameter θ, let us consider a prior density π on Θ, and let ρ be a suitable function. Then Hanousek [11] considered the following Bayes-type or B-estimator of θ_0, corresponding to the prior density π and generated by the function ρ,

$$\widehat{\theta}_n^* = \frac{\int_\Theta \theta \exp\{-\sum_{i=1}^n \rho(X_i,\theta)\}\pi(\theta)\,\mathrm{d}\theta}{\int_\Theta \exp\{-\sum_{i=1}^n \rho(X_i,\theta)\}\pi(\theta)\,\mathrm{d}\theta}$$

if both integrals exist. This type of estimator is close in spirit to the class of Laplace-type estimators introduced in [8].

The posterior M-estimator corresponding to the prior density π and generated by ρ is defined as:

$$\widehat{\theta}_n^+ = \arg\max_{\theta\in\Theta}\left(-\sum_{i=1}^n \rho(X_i,\theta) + \ln\pi(\theta)\right).$$

Hanousek [11] showed that $\widehat{\theta}_n^*$ is asymptotically equivalent to the M-estimator generated by ρ for a large class of priors and under some conditions on ρ and $\mathbb{P}_{\theta_0}$. The asymptotic equivalence provides the access to the study of asymptotics for B-estimators via the M-estimators.

In the context of the Bayesian methods examined in this note, instead of a likelihood function, our work will use a criterion function $\mathbb{P}_n h(\theta,\alpha) := \int h(\theta,\alpha)\,\mathrm{d}\mathbb{P}_n$. Therefore, the inference is based on the ϕ-posterior

$$p_{\phi,n}(\alpha|X_1,\cdots,X_n) = \frac{\exp\{n\mathbb{P}_n h(\theta,\alpha)\}\pi(\alpha)}{\int_\mathcal{U} \exp\{n\mathbb{P}_n h(\theta,\alpha)\}\pi(\alpha)\,\mathrm{d}\alpha}.$$

A risk function is the expected loss or error in which the researcher incurs when choosing a certain value for the parameter estimate. Let $\mathcal{L}_n$ be a loss function. The risk function takes the form

$$\mathcal{R}_n(\widetilde{\alpha}) = \int_\mathcal{U} \mathcal{L}_n(\alpha-\widetilde{\alpha})\,p_{\phi,n}(\alpha|X_1,\cdots,X_n)\,\mathrm{d}\alpha,$$

where $p_{\phi,n}(\alpha|X_1,\cdots,X_n)$ is the ϕ-posterior density, $\widetilde{\alpha}$ is the selected value, and α is all other possible value that we are integrating over. The loss function can penalize the selection of α asymmetrically, and is a function of the selected value and the rest of the possible values of the parameters in $\mathcal{U}$.

The dual ϕ-divergence Bayes type estimator minimizes the expected loss for different forms of the loss function

$$\widehat{\alpha}_\phi^*(\theta) = \arg\inf_{\widetilde{\alpha}\in\mathcal{U}} \mathcal{R}_n(\widetilde{\alpha}).$$

Choosing different loss functions will change the objective function such that the estimators bear different interpretations, other familiar forms obtained for different loss functions are modes, medians and quantiles. For instance, when the loss is squared error ($\mathcal{L}_n(u) = |\sqrt{n}u|^2$), for fixed θ, the dual ϕ-divergence Bayes type estimator is defined as

$$\widehat{\alpha}_\phi^*(\theta) = \int_\mathcal{U} \alpha\,p_{\phi,n}(\alpha|X_1,\cdots,X_n)\,\mathrm{d}\alpha := \frac{\int_\mathcal{U} \alpha\exp\{n\mathbb{P}_n h(\theta,\alpha)\}\pi(\alpha)\,\mathrm{d}\alpha}{\int_\mathcal{U} \exp\{n\mathbb{P}_n h(\theta,\alpha)\}\pi(\alpha)\,\mathrm{d}\alpha}, \tag{7}$$

if both integrals exist.

The posterior dual ϕ-divergences estimator is defined as

$$\widehat{\alpha}_{\phi}^{+}(\theta)=\arg\sup_{\alpha\in\mathcal{U}}\big(\mathbb{P}_n h(\theta,\alpha)+\ln\pi(\alpha)\big).$$

It is obvious that posterior dual ϕ-divergences estimates naturally inherit the properties of dual ϕ-divergences estimates and hence we focus on dual ϕ-divergences Bayes-type estimators only.

Remark 1.

(i) The expected a posteriori (EAP) estimator, which is the mean of the posterior distribution, belongs to the class of estimates (7). Indeed, it is obtained when $\phi(x)=-\log x+x-1$, that is as the dual modified KL_m-divergence estimate. Observe that $\phi'(x)=-\frac{1}{x}+1$ and $x\phi'(x)-\phi(x)=\log x$, hence

$$\int h(\theta,\alpha)\,\mathrm{d}\mathbb{P}_n=-\int\log\left(\frac{\mathrm{d}\mathbb{P}_\theta}{\mathrm{d}\mathbb{P}_\alpha}\right)\mathrm{d}\mathbb{P}_n.$$

Keeping in mind definitions (7), we get

$$\widehat{\alpha}_{KL_m}^{*}(\theta):=\frac{\int_{\mathcal{U}}\alpha\prod_{i=1}^{n}p_\alpha(X_i)\pi(\alpha)\,\mathrm{d}\alpha}{\int_{\mathcal{U}}\prod_{i=1}^{n}p_\alpha(X_i)\pi(\alpha)\,\mathrm{d}\alpha},$$

independently upon θ.

(ii) If new data $X_{n+1},\dots,X_N$ are obtained, the posterior for the combined data $X_1,\dots,X_N$ can be obtained by using posterior after n observations, $p_{\phi,n}(\alpha|X_1,\cdots,X_n)$ as a prior for α:

$$p_{\phi,n}(\alpha|X_1,\cdots,X_N)\propto p_{\phi,n}(\alpha|X_1,\cdots,X_n)\times p_{\phi,n}(X_{n+1},\cdots,X_N|\alpha).$$

This leads to easy updating of the posterior distribution by a sequence of incremental changes. The methodology presented in [13] treats a data set as a single observation of a function rather than a set of distinct observations. As such it is difficult to separate the posterior to find the effects of a single observation.

3. Asymptotic properties

In this section we state the asymptotic normality of the estimates based on the ϕ-posterior and evaluate their limiting variance. The hypotheses handled here are similar to those used in [15] and [5] in the frequentist case; these conditions are mild and can be satisfied in most of circumstances. From now on, $\xrightarrow{D}$ denotes the convergence in distribution. The proofs of our asymptotic results rely on the assumptions listed below.

(R.1)

$$\sup_{\alpha\in\mathcal{U}}\left|\mathbb{P}_n h(\theta,\alpha)-\mathbb{P}_{\theta_0}h(\theta,\alpha)\right|\xrightarrow{a.s.}0.$$

(R.2) There exists a neighborhood $N(\theta_0)$ of θ_0 such that the first- and second-order partial derivatives (w.r.t. α) of $\phi'(\frac{p_\theta(x)}{p_\alpha(x)})p_\theta(x)$ are dominated on $N(\theta_0)$ by some integrable functions. The third-order partial derivatives (w.r.t. α) of $h(\theta,\alpha,x)$ are dominated on $N(\theta_0)$ by some $\mathbb{P}_{\theta_0}$-integrable functions.

Let

$$S:=-\mathbb{P}_{\theta_0}\frac{\partial^2}{\partial\alpha^2}h(\theta,\theta_0)\quad\text{and}\quad V:=\mathbb{P}_{\theta_0}\frac{\partial}{\partial\alpha}h(\theta,\theta_0)^\top\frac{\partial}{\partial\alpha}h(\theta,\theta_0).$$

Observe that the matrix S is symmetric and positive since the second derivative ϕ'' is nonnegative by the convexity of ϕ.

(R.3) The matrices S and V are non-singular.

Let

$$U_n(\theta_0):=P_n\frac{\partial}{\partial\alpha}h(\theta,\theta_0).$$

For α in an open neighborhood of θ_0, using (R.2) by a Taylor expansion

$$\mathbb{P}_n h(\theta,\alpha)-\mathbb{P}_n h(\theta,\theta_0)=(\alpha-\theta_0)^\top U_n(\theta_0)-\frac{1}{2}(\alpha-\theta_0)^\top S(\alpha-\theta_0)+R_n(\alpha),$$

where $R_n(\alpha)$ is the rest of the Taylor expansion of $\mathbb{P}_n h(\theta,\alpha)$ in α around θ_0.

(R.4) Given any $\epsilon > 0$, there exists $\delta > 0$ such that, the probability of the event

$$\sup_{|\alpha-\theta_0|\le\delta} \left|R_n(\alpha)\right| \ge \epsilon$$

tends to zero as $n \longrightarrow \infty$.

Remark 2.

(i) Using example 19.8 in [30], it is clear that the class of functions $\{\alpha \mapsto h(\theta,\alpha); \alpha \in \mathcal{U}\}$ is a Glivenko–Cantelli class of functions for all fixed θ, that (R.1) holds.
(ii) Conditions (R.2) and (R.3) are about usual regularity properties of the underlying model, they guarantee that we can interchange integration and differentiation and the existence of the variance–covariance matrices, they are similar to regularity conditions used in [15] and [5] in the frequentist case.
(iii) Condition (R.4) easily holds when there is enough smoothness. It requires that the remainder term of the expansion can be controlled in a particular way over a neighborhood of θ_0.

Define, for any α

$$t := \sqrt{n}(\alpha - \Delta_n), \quad \Delta_n := \theta_0 + S^{-1}U_n(\theta_0),$$

and $p^*_{\phi,n}(t)$ be the ϕ-posterior density of t.

The following theorem states that under some regularity conditions, for large n, $p^*_{\phi,n}(\cdot)$ is approximately a random normal density in the L_1 sense.

Theorem 1. *Let π be any prior that is continuous and positive at θ_0 with $\int |\theta|\pi(\theta)\,\mathrm{d}\theta < \infty$. Then under conditions (R.1–R.4)*

$$\int \left| p^*_{\phi,n}(t) - \left(\frac{\det S}{2\pi}\right)^{d/2} \exp\left\{-\frac{1}{2}t^\top St\right\}\right| \mathrm{d}t \xrightarrow{P} 0.$$

We now state the principal result of this section. Theorem 2 is concerned with the efficiency and asymptotic normality of the proposed estimates. See [14,26] and [17] for more on the consistency and efficiency of Bayes estimators.

Theorem 2. *Let π be any prior that is continuous and positive at θ_0 with $\int |\theta|\pi(\theta)\,\mathrm{d}\theta < \infty$. Assume that conditions (R.1–R.4) hold, then as n tends to infinity*

$$V^{-1/2}S\sqrt{n}\big(\widehat{\alpha}^*_\phi(\theta) - \theta_0\big) \xrightarrow{d} \mathcal{N}(0, I).$$

Remark 3. The very peculiar choice of the escort parameter defined through $\theta = \theta_0$ has the same limit properties as the MLE one. That is $S^\top V^{-1}S = I_{\theta_0}$ the information matrix, so that $\widehat{\alpha}^*_\phi(\theta_0)$ is consistent and asymptotically efficient. The consequence is that the value of the escort parameter should be taken as a consistent estimator of θ_0. If the data are subject to contamination, better results are obtained for estimators escorted by robust estimator of θ_0, see [7] and [1] for relevant discussion on this subject.

4. Concluding remarks

We have introduced a new estimation procedure in parametric models that combine divergences method with Bayesian analysis, it generalizes the expected a posteriori estimate. The proposed estimators are based on integration rather than optimization. These estimators are often much easier to compute in practice than the arg sup estimators (6), especially in the high-dimensional setting; see, for example, the discussion in [27].

In order to compute these estimators, using MCMC methods, we can draw a Markov chain,

$$\mathbb{S} = \big(\alpha^{(1)}; \alpha^{(2)}; \cdots; \alpha^{(B)}\big);$$

whose marginal density is approximately given by $p_{\phi,n}(\cdot)$, the ϕ-posterior distribution. Then the estimate $\widehat{\alpha}^*_\phi(\theta)$ is computed as

$$\widehat{\alpha}^*_\phi(\theta) = \frac{1}{B}\sum_{i=1}^{B} \alpha^{(i)}$$

where B is the number of MCMC draws.

Consider the construction of confidence intervals for the quantity $f(\theta_0)$, for a given continuously differentiable function $f : \Theta \longrightarrow \mathbb{R}$. Define

$$C_n(\epsilon) := \inf\left\{x : \int_{f(\alpha)\leq x} \alpha\, p_{\phi,n}(\alpha)\,\mathrm{d}\alpha \geq \epsilon\right\}.$$

Then the dual ϕ-divergence Bayes type estimator confidence interval is given by $[C_n(\frac{\epsilon}{2}); C_n(1-\frac{\epsilon}{2})]$. These confidence intervals can be constructed simply by taking the $\frac{\epsilon}{2}$th and $1-\frac{\epsilon}{2}$th quantiles of the MCMC sequence

$$f(\mathbb{S}) = \left(f\left(\alpha^{(1)}\right); f\left(\alpha^{(2)}\right); \cdots; f\left(\alpha^{(B)}\right)\right),$$

and thus are quite simple in practice.

The very peculiar choice of the escort parameter defined through $\theta = \theta_0$ has the same limit properties as the posterior mean. This result is of some relevance, since it leaves open the choice of the divergence, while keeping good asymptotic properties; we expect that it can also be used directly to provide robust inference, and we leave this study for a subsequent paper.

The problem of the choice of the divergence remains an open question and needs more investigation.

Acknowledgements

The author is very grateful to the referee for his/her comments that have led to an improved version of the note. The author owes thanks to Salim Bouzebda, who has carefully read the first draft of this work and provided many insightful suggestions.

References

[1] S. Bouzebda, M. Cherfi, General bootstrap for dual ϕ-divergence estimates, J. Probab. Stat. (2012) 33 p., article ID 834107.
[2] S. Bouzebda, M. Cherfi, Dual divergence estimators of the tail index, ISRN Probab. Stat. (2012) 14 p., article ID 746203.
[3] S. Bouzebda, A. Keziou, New estimates and tests of independence in semiparametric copula models, Kybernetika 46 (1) (2010) 178–201.
[4] M. Broniatowski, A. Keziou, Minimization of ϕ-divergences on sets of signed measures, Studia Sci. Math. Hung. 43 (4) (2006) 403–442.
[5] M. Broniatowski, A. Keziou, Parametric estimation and tests through divergences and the duality technique, J. Multivar. Anal. 100 (1) (2009) 16–36.
[6] M. Broniatowski, I. Vajda, Several applications of divergence criteria in continuous families, Kybernetika 48 (4) (2012) 600–636.
[7] M. Cherfi, Dual divergences estimation for censored survival data, J. Stat. Plan. Inference 142 (7) (2012) 1746–1756.
[8] V. Chernozhukov, H. Hong, An MCMC approach to classical estimation, J. Econometrics 115 (2003) 293–346.
[9] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, J. R. Stat. Soc., Ser. B, Stat. Methodol. 46 (3) (1984) 440–464.
[10] D.K. Dey, L.R. Birmiwal, Robust Bayesian analysis using divergence measures, Stat. Probab. Lett. 20 (4) (1994) 287–294.
[11] J. Hanousek, Robust Bayesian type estimators and their asymptotic representation, Stat. Decis. 8 (1) (1990) 61–69.
[12] J. Hanousek, Generalized Bayesian-type estimators. Robust and sensitivity analysis, Kybernetika 30 (3) (1994) 271–278.
[13] G. Hooker, A. Vidyashankar, Bayesian model robustness via disparities, Test (2014), http://dx.doi.org/10.1007/s11749-014-0360-z, in press.
[14] I.A. Ibragimov, R.Z. Has'minskii, Statistical Estimation — Asymptotic Theory, Springer-Verlag, New York, 1981.
[15] A. Keziou, Dual representation of ϕ-divergences and applications, C. R. Acad. Sci. Paris, Ser. I 336 (10) (2003) 857–862.
[16] A. Keziou, S. Leoni-Aubin, On empirical likelihood for semiparametric two-sample density ratio models, J. Stat. Plan. Inference 138 (4) (2008) 915–928.
[17] E.L. Lehmann, G. Casella, Theory of Point Estimation, 2nd edition, Springer Texts in Statistics, Springer-Verlag, New York, 1998.
[18] F. Liese, I. Vajda, Convex Statistical Distances, Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], vol. 95, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, Germany, 1987, with German, French and Russian summaries.
[19] F. Liese, I. Vajda, On divergences and informations in statistics and information theory, IEEE Trans. Inf. Theory 52 (10) (2006) 4394–4412.
[20] B.G. Lindsay, Efficiency versus robustness: the case for minimum Hellinger distance and related methods, Ann. Statist. 22 (2) (1994) 1081–1114.
[21] L. Pardo, Statistical Inference Based on Divergence Measures, Statistics: Textbooks and Monographs, vol. 185, Chapman & Hall/CRC, Boca Raton, FL, USA, 2006.
[22] F. Peng, D. Dey, Bayesian analysis of outlier problems using divergence measures, Canad. J. Statist. 23 (2), 199–213.
[23] G. Ragusa, Bayesian properties of minimum divergence and generalized empirical likelihood methods, Unpublished manuscript, 2006.
[24] C. Robert, The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementations, Springer, New York, 2001.
[25] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, 2005.
[26] H. Strasser, Consistency of maximum likelihood and Bayes estimates, Ann. Statist. 9 (1981) 1107–1113.
[27] L. Tian, J.S. Liu, L.J. Wei, Implementation of estimating-function based inference procedures with Markov chain Monte Carlo samplers, J. Amer. Statist. Assoc. 102 (2007) 881–888.
[28] A. Toma, M. Broniatowski, Dual divergence estimators and tests: robustness results, J. Multivar. Anal. 102 (1) (2011) 20–36.
[29] A. Toma, S. Leoni-Aubin, Robust tests based on dual divergence estimators and saddlepoint approximations, J. Multivar. Anal. 101 (5) (2010) 1143–1155.
[30] A.W. van der Vaart, Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3, Cambridge University Press, Cambridge, UK, 1998.