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 $X_1,\ldots,X_n$ 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

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