Journal of Statistical Planning and Inference 142 (2012) 2574–2585

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# Decomposable pseudodistances and applications in statistical estimation

Michel Broniatowski [a], Aida Toma [b,c,*], Igor Vajda [d,†]

[a] Laboratoire de Statistique Théorique et Appliquée, Université Paris 6, 4 Place Jussieu, 75005 Paris, France
[b] Department of Applied Mathematics, Bucharest Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania
[c] "Gheorghe Mihoc – Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Calea 13 Septembrie 13, Bucharest, Romania
[d] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodarenskou, Vezi 4, 18208 Prague, Czech Republic

ARTICLE INFO

*Article history:*
Received 8 September 2011
Received in revised form
23 March 2012
Accepted 23 March 2012
Available online 1 April 2012

*Keywords:*
Divergence
Pseudodistance
Influence function
Robustness
Parametric model
Regression

ABSTRACT

The aim of this paper is to introduce new statistical criteria for estimation, suitable for inference in models with common continuous support. This proposal is in the direct line of a renewed interest for divergence based inference tools imbedding the most classical ones, such as maximum likelihood, Chi-square or Kullback–Leibler. General pseudodistances with decomposable structure are considered, they allowing defining minimum pseudodistance estimators, without using nonparametric density estimators. A special class of pseudodistances indexed by $\alpha > 0$, leading for $\alpha \downarrow 0$ to the Kullback–Leibler divergence, is presented in detail. Corresponding estimation criteria are developed and asymptotic properties are studied. The estimation method is then extended to regression models. Finally, some examples based on Monte Carlo simulations are discussed.


## 1. Introduction

In parametric estimation, minimum divergence methods, i.e. methods which estimate the parameter by minimizing an estimate of some divergence between the assumed model density and the true density underlying the data, have been extensively studied (see Pardo, 2006 and references herein). Generally, in continuous models, the minimum divergence methods have the drawback that it is necessary to use some nonparametric density estimator. In order to remove this drawback, some proposals have been made in the literature. Among them, we recall the minimum density power divergence method introduced in Basu et al. (1998), and a minimum divergence method based on duality arguments, independently proposed in Liese and Vajda (2006) and in Broniatowski and Keziou (2009). Many statistical applications of these methods are now available. Among them we recall applications in two sample density ratio models (Keziou and Leoni-Aubin, 2008), in robust testing (Toma and Leoni-Aubin, 2010), applications for construction of tests of independence in copula models (Bouzebda and Keziou, 2010a; Bouzebda and Keziou, 2010b) or applications in construction of goodness of fit tests (Karagrigoriou and Mattheou, 2010).

The results obtained in the present paper follow this line of research. Without referring to all properties of the divergence criterions, we mainly quote their information processing property, i.e. the complete invariance with respect to the statistically sufficient transformations of the observation space. This property is useful but probably not unavoidable in

* Corresponding author at: Department of Applied Mathematics, Bucharest Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania.
*E-mail addresses:* michel.broniatowski@upmc.fr (M. Broniatowski), aida_toma@yahoo.com (A. Toma).
† Deceased

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.03.019

the minimum divergence estimation based on similarity between theoretical and empirical distributions. In this paper we admit general pseudodistances which may not satisfy the information processing property. The definition of the pseudodistance, which is at the start of this work, pertains to the willingness to define a simple frame including all commonly used statistical criterions, from maximum likelihood to the $L_2$ norm. Such a description is provided in unpublished Broniatowski and Vajda (2009). In the present paper we define a class of pseudodistances indexed by $\alpha>0$, leading for $\alpha\downarrow 0$ to the Kulback–Leibler divergence. The peculiar features of these pseudodistances recommend it as an appealing competing choice for defining estimation criteria. We argue that by defining and studying minimum pseudodistances estimators for classical parametric models and for regression models. We present such tools for inference with a special attention to limit properties and robustness, in a similar spirit as in Toma and Broniatowski (2011).

The outline of the paper is as follows. Section 2 introduces decomposable pseudodistances and define minimum pseudodistances estimators. Section 3 presents a special class of minimum pseudodistances estimators. For these estimators we study invariance properties, consistency, asymptotic normality and robustness. In Section 4, the estimation method is applied to linear models for which asymptotic and robustness properties are derived. Finally, in order to illustrate the performance in finite samples of the proposed method, we give some examples based on Monte Carlo simulations.

## 2. Decomposable pseudodistances and estimators

We consider $\mathcal{P}$ a parametric model with euclidian parameter space $\Theta$ and we assume that all the probability measures $P_\theta$ in $\mathcal{P}$ share the same support, which is included in $\mathbb{R}^d$. Every $P_\theta$ has a density $p_\theta$ with respect to the Lebesgue measure.

We denote by $\mathcal{P}_{\text{emp}}$ the class of probability measures induced by samples, namely the class of all probability measures $P_n \coloneqq (1/n)\sum_{i=1}^n \delta_{X_i}$, where $X_1,\dots,X_n$ is sampled according to a distribution on $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$, not necessarily in $\mathcal{P}$. In addition to the previous notation it is useful to introduce a family of measures $\mathcal{P}_0$ associated to distributions generating the data when studying robustness properties. Often, such a measure is a mixture of some element in $\mathcal{P}$ with a Dirac measure at some point $x$ in $\mathbb{R}^d$. We also define $\mathcal{P}^+ \coloneqq \mathcal{P}\cup\mathcal{P}_0$.

**Definition 1.** We say that $\mathfrak{D}:\mathcal{P}\otimes\mathcal{P}^+\mapsto\mathbb{R}$ is a pseudodistance between probability measures $P\in\mathcal{P}=\{P_\theta:\theta\in\Theta\}$ and $Q\in\mathcal{P}^+$ if $\mathfrak{D}(P_\theta,Q)\geq 0$, for all $\theta\in\Theta$ and $Q\in\mathcal{P}^+$ and $\mathfrak{D}(P_\theta,P_{\tilde\theta})=0$ if and only if $\theta=\tilde\theta$.

**Definition 2.** A pseudodistance $\mathfrak{D}$ on $\mathcal{P}\otimes\mathcal{P}^+$ is called decomposable if there exist functionals $\mathfrak{D}^0:\mathcal{P}\mapsto\mathbb{R}$, $\mathfrak{D}^1:\mathcal{P}^+\mapsto\mathbb{R}$ and measurable mappings $\rho_\theta:\mathbb{R}^d\mapsto\mathbb{R},\theta\in\Theta$, such that for all $\theta\in\Theta$ and $Q\in\mathcal{P}^+$ the expectations $\int\rho_\theta\,dQ$ exist and

$$\mathfrak{D}(P_\theta,Q)=\mathfrak{D}^0(P_\theta)+\mathfrak{D}^1(Q)+\int\rho_\theta\,dQ. \tag{1}$$

A known class of pseudodistances has been introduced by Basu et al. (1998) and is called the class of power divergences. This class corresponds to

$$\mathfrak{D}(P_\theta,Q)=\int\left\{p_\theta^{\alpha+1}-\left(1+\frac{1}{\alpha}\right)p_\theta^\alpha q+\frac{1}{\alpha}q^{\alpha+1}\right\}d\lambda, \tag{2}$$

for $\alpha>0$. Note that the pseudodistances (2) are decomposable with

$$\mathfrak{D}^0(P_\theta)=\int p_\theta^{\alpha+1}\,d\lambda,\quad \mathfrak{D}^1(Q)=\frac{1}{\alpha}\int q^{\alpha+1}\,d\lambda,\quad \rho_\theta=-\left(1+\frac{1}{\alpha}\right)p_\theta^\alpha. \tag{3}$$

In the next section, we introduce a new class of pseudodistances from which a new statistical criterion for inference is deduced.

**Definition 3.** We say that a functional $T_{\mathfrak{D}}:\mathcal{Q}\mapsto\Theta$ for $\mathcal{Q}=\mathcal{P}^+\cup\mathcal{P}_{\text{emp}}$ defines a minimum pseudodistance estimator (briefly, min $\mathfrak{D}$-estimator), if the pseudodistance $\mathfrak{D}(P_\theta,Q)$ is decomposable on $\mathcal{P}\otimes\mathcal{P}^+$ and the parameters $T_{\mathfrak{D}}(Q)\in\Theta$ minimize $\mathfrak{D}^0(P_\theta)+\int\rho_\theta\,dQ$ on $\Theta$, in symbols

$$T_{\mathfrak{D}}(Q)=\arg\inf_\theta\left[\mathfrak{D}^0(P_\theta)+\int\rho_\theta\,dQ\right]\quad\text{for all } Q\in\mathcal{Q}. \tag{4}$$

In particular, for $Q=P_n\in\mathcal{P}_{\text{emp}}$

$$\widehat\theta_{\mathfrak{D},n}\coloneqq T_{\mathfrak{D}}(P_n)=\arg\inf_\theta\left[\mathfrak{D}^0(P_\theta)+\frac{1}{n}\sum_{i=1}^n\rho_\theta(X_i)\right]. \tag{5}$$

**Theorem 1.** *Every* min $\mathfrak{D}$*-estimator given by* (5) *is Fisher consistent in the sense that* $T_{\mathfrak{D}}(P_{\theta_0})=\theta_0$, *for all* $\theta_0\in\Theta$.

**Proof.** Consider fixed $\theta_0\in\Theta$. Then, by assumptions, $\mathfrak{D}^1(P_{\theta_0})$ is a finite constant. Therefore (4) together with the definition of pseudodistance implies

$$T_{\mathfrak{D}}(P_{\theta_0})=\arg\inf_\theta\left[\mathfrak{D}^0(P_\theta)+\int\rho_\theta\,dP_{\theta_0}\right]=\arg\inf_\theta\left[\mathfrak{D}^0(P_\theta)+\mathfrak{D}^1(P_{\theta_0})+\int\rho_\theta\,dP_{\theta_0}\right]=\arg\inf_\theta\mathfrak{D}(P_\theta,P_{\theta_0})=\theta_0.$$