Statistics and Probability Letters 96 (2015) 68–74

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

# A general set up for minimum disparity estimation

Arun Kumar Kuchibhotla, Ayanendranath Basu *

*Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata - 700 108, India*

ARTICLE INFO

*Article history:*
Received 3 April 2014
Received in revised form 29 August 2014
Accepted 29 August 2014
Available online 6 September 2014

*Keywords:*
Disparity
Robustness
Efficiency
Kernel density estimation
Residual adjustment function

ABSTRACT

Lindsay (1994) provided a general set up in discrete models for minimum disparity estimation. Such a set up eludes us in continuous models. We provide such a general result and hence fill up a major gap in the literature.



## 1. Introduction

In density-based minimum distance estimation one minimizes a discrepancy between a nonparametric density estimate obtained from the sample and the parametric model density. In discrete models a natural choice for this nonparametric estimate is provided by observed relative frequencies in the sample. In continuous models one needs a technique like kernel density estimation; this complicates the method and one now has to worry about the conditions on the model and the conditions on the kernel bandwidth.

Beran's (1977) work was the first useful application of the density-based minimum distance estimation method with a robustness motivation. Minimum Hellinger distance estimation was further studied by Tamura and Boos (1986) and Simpson (1987, 1989). Lindsay (1994) presented a comprehensive treatment of density-based minimum distance estimation in discrete models covering the class of chi-square type distances, called disparities.

Basu and Lindsay (1994) provided a similar treatment in the continuous case where the model was smoothed with the same kernel. This reduces the dependence on the bandwidth. However this method requires a transparent kernel for full asymptotic efficiency, which may be difficult to obtain.

A complete framework of minimum disparity estimation for continuous models, which would parallel Lindsay's (1994) work for discrete models, does not exist in the literature. The closest approximation to a general framework that the literature offers is the work by Park and Basu (2004). Although this work covers a large number of candidate disparities, it imposes strong conditions on the disparities, and as a result this framework also excludes a fair number of useful disparities, including the Hellinger distance.

In this paper we provide this general framework under simple conditions. Our proof consolidates the existing elements in the literature into a set of more inclusive assumptions. Our framework includes practically all disparities which are of interest in robust estimation; our conditions are even more accessible than Lindsay's discrete model conditions. We do not

* Corresponding author.
*E-mail addresses:* karun3kumar@gmail.com (A.K. Kuchibhotla), ayanbasu@isical.ac.in (A. Basu).

http://dx.doi.org/10.1016/j.spl.2014.08.020
0167-7152/

describe the robustness of the estimators here as we feel that the breakdown point results of Park and Basu (2004) are sufficiently general.

In Section 2, we introduce minimum disparity estimation. In Section 3, we state the required assumptions and derive the asymptotic distribution of the estimator. This is done by proving the asymptotic normality of the estimating function and by establishing the uniform convergence of the derivative of the estimating function to a non-stochastic function.

## 2. Minimum disparity estimation

Let $\mathcal{G}$ represent the class of all distribution functions having densities with respect to the Lebesgue measure. We assume that the true distribution $G$ and the model $\mathcal{F}_\theta = \{F_\theta : \theta \in \Theta\}$ belong to $\mathcal{G}$. Let $g$ and $f_\theta$ be the corresponding densities. Let $X_1, X_2, \ldots, X_n$ be a random sample from $G$ which is modelled by $\mathcal{F}_\theta$. Our aim is to estimate the parameter $\theta$ by choosing the model density which gives the closest fit to the data. We assume that the support of $f_\theta$ is independent of $\theta$ which is the same as the support of $g$.

Let $C$ be a thrice differentiable convex function defined on $[-1, \infty)$, satisfying $C(0) = 0$. Define

$$\rho_C(g, f_\theta) = \int C\left(\frac{g(x)}{f_\theta(x)} - 1\right) f_\theta(x) dx.$$

This form describes the class of disparities between the densities $g$ and $f_\theta$. A simple application of Jensen's inequality shows that $\rho_C(g, f_\theta) \geq 0$ with equality if and only if $g = f_\theta$ identically. We denote by $T(G)$, the "best fitting parameter" which minimizes $\rho_C(g, f_\theta)$ over all $\theta \in \Theta$. We consider the minimum disparity estimator $\hat{\theta}_n$ of $\theta$ defined by

$$\hat{\theta}_n := \arg\min_\theta \rho_C(g_n, f_\theta),$$

where $g_n$ is a kernel density estimator obtained from the sample. Under differentiability of the model, $\hat{\theta}_n$ can be obtained as a root of the equation

$$\int A(\delta_n(x)) \nabla f_\theta(x) dx = 0, \tag{1}$$

where $\nabla$ represents the gradient with respect to $\theta$, $A(\delta) = C'(\delta)(\delta + 1) - C(\delta)$ and $\delta_n(x) + 1 = g_n(x)/f_\theta(x)$. This function $A(\cdot)$ is called the residual adjustment function (RAF) of the disparity and $\delta_n$ is referred to as the Pearson residual. The function $A(\cdot)$ plays a very crucial role in determining the robustness properties of the estimator. See Basu et al. (2011) for more details. Denote by $\Psi_n(\theta)$ the expression on the left-hand side of Eq. (1); $\Psi_n(\theta)$ is our estimating function.

## 3. Asymptotic distribution of the estimator

### 3.1. Assumptions

Consider the parametric set up of the previous section. We will estimate $\theta$ by minimizing the disparity $\rho_C(g_n, f_\theta)$ between a kernel density estimator $g_n$ and the model density $f_\theta$. The estimator $g_n$ based on independent and identically distributed observations $X_1, X_2, \ldots, X_n$ is given by

$$g_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where $K$ is a suitable kernel function and $h_n$ is the bandwidth. We will use the notation $u_\theta(x)$ for $\nabla \log f_\theta(x)$.

Here we present the set of conditions under which the necessary asymptotic results will be derived. In the following $A(\delta)$ will represent the residual adjustment function of the disparity $\rho_C$; for simplicity of notation we will drop the subscript $n$ from $h_n$, unless specifically demanded by the situation.

(A1) $A''(\delta)(\delta + 1)^\alpha$ is bounded for some fixed $\alpha$, i.e., $|A''(\delta)(\delta + 1)^\alpha| \leq M < \infty$ for some $\alpha$ and for all $\delta \geq -1$, where $1 + \delta(x) = g(x)/f_\theta(x)$. All the other instances of $\alpha$ in the assumptions relate to this specific value.
(A2) The density $g$ is twice differentiable and $g''$ is bounded. Also, $g$ is uniformly continuous.
(A3) $(n^{1/2}h)^{-1} \int |\lambda_n(x)| dx \to 0$ as $n \to \infty$, where $\lambda_n(x) = \mathbf{1}_{\{|x| \leq \tau_n\}} \frac{g^{1-\alpha}}{f_\theta^{1-\alpha}} |u_\theta|$, $\tau_n$ is a sequence such that $\tau_n \to \infty$ as $n \to \infty$. Here $\mathbf{1}_{\{A\}}$ denotes the indicator function of set $A$. Define $\varepsilon_n(x) = \mathbf{1}_{\{|x| > \tau_n\}} g^{1-\alpha} |u_\theta| / f_\theta^{1-\alpha}$.
(A4) The kernel $K$ is symmetric and is supported on a compact support denoted by $\Omega$; $h \to 0$, $nh \to \infty$, $nh/\log n \to \infty$, $n^{1/2}h^2 \to 0$ as $n \to \infty$.
(A5) $M_n \equiv \sup_{|x| \leq \tau_n} \sup_{t \in \Omega} g(x + ht)/g(x) = O(1)$, as $n \to \infty$.
(A6) $n \sup_{t \in \Omega} P(|X_1 - h_n t| > \tau_n) \to 0$ as $n \to \infty$.
(A7) For $\theta = T(G)$, the best fitting parameter,