Journal of Statistical Planning and Inference 142 (2012) 2351–2368

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

M-estimators for isotonic regression

Enrique E. Álvarez [a,*], Víctor J. Yohai [b]

[a] University of La Plata and CONICET, Argentina
[b] University of Buenos Aires and CONICET, Argentina

ARTICLE INFO

Article history:
Received 13 March 2011
Received in revised form
20 February 2012
Accepted 21 February 2012
Available online 28 February 2012

Keywords:
Isotonic regression
M-estimators
Robust estimates

ABSTRACT

In this paper we propose a family of robust estimates for isotonic regression: isotonic M-estimators. We show that their asymptotic distribution is, up to an scalar factor, the same as that of Brunk's classical isotonic estimator. We also derive the influence function and the breakdown point of these estimates. Finally we perform a Monte Carlo study that shows that the proposed family includes estimators that are simultaneously highly efficient under Gaussian errors and highly robust when the error distribution has heavy tails.

1. Introduction

Let $x_1, \dots, x_n$ be independent random variables collected along observation points $t_1 \le \cdots \le t_n$ according to the model:

$$x_j = \mu(t_j) + u_j, \tag{1}$$

where the u_j's are i.i.d. random variables with distribution G. In *isotonic regression* the trend term $\mu(t)$ is monotone non-decreasing, i.e., $\mu(t_1) \le \cdots \le \mu(t_n)$, but it is otherwise arbitrary. In this set-up, the classical estimator of $\mu(t)$ is the function g which minimizes the L_2 distance between the vector of observed and fitted responses, i.e, it minimizes

$$\sum_{j=1}^{n} [x_j - g(t_j)]^2 \tag{2}$$

in the class $\mathcal{G}$ of non-decreasing piecewise continuous functions. It is trivial but noteworthy that Eq. (2) posits a finite dimensional convex constrained optimization problem. Its solution was first proposed by Brunk (1958) and has received extensive attention in the statistical literature (see, e.g., Robertson et al., 1988 for a comprehensive account). It is also worth noting that any piecewise continuous non-decreasing function which agrees with the optimizer of (2) at the t_j's will be a solution. For that reason, in order to achieve uniqueness, it is traditional to restrict further the class $\mathcal{G}_0$ to the subset of piecewise constant non-decreasing functions. Another valid choice consists of the interpolation at the knots with non-decreasing cubic splines or any other piecewise continuous monotone function, e.g., Meyer (1996). We will call this estimator the L_2 *isotonic estimator*.

* Corresponding author.
E-mail addresses: ealvarez@mate.unlp.edu.ar, enriqueealvarez@fibertel.com.ar (E.E. Álvarez).

0378-3758/$ - see front matter © 2012 Published by Elsevier B.V.
doi:10.1016/j.jspi.2012.02.051

The sensitivity of this estimator to extreme observations (outliers) was noted by Wang and Huang (2002), who propose minimizing instead using the L_1 norm, i.e., minimizing

$$\sum_{j=1}^{n} |x_j - g(t_j)|.$$

This estimator will be call here L_1 *Isotonic estimator*. Wang and Huang (2002) developed the asymptotic distribution of the trend estimator at a given observation point t_0 and obtained the asymptotic relative efficiency of this estimator compared with the classical L_2 estimator. Interestingly, this efficiency turned out to be $2/\pi = 0.637$, the same as in the i.i.d. location problem.

In this paper we will propose instead a robust *isotonic* M-estimator aimed at balancing robustness with efficiency. Specifically we shall seek the minimizer of

$$\sum_{j=1}^{n} \rho\left(\frac{x_j - g(t_j)}{\widehat{\sigma}_n}\right), \tag{3}$$

where $\widehat{\sigma}_n$ is an estimator of the error scale previously obtained and ρ satisfies the following properties:

A1 (i) $\rho(x)$ is non-decreasing in $|x|$, (ii) $\rho(0) = 0$, (iii) ρ is even, (iv) $\rho(x)$ is strictly increasing for $x > 0$, (v) ρ has two continuous derivatives and $\psi = \rho'$ is bounded and monotone non-decreasing, and (vi) $E(\psi(u/\sigma_0) = 0$.

Note that in the case that $E(\psi(u/\sigma_0)) = c \neq 0$, Assumption A1 (vi) holds changing the trend function μ by $\mu + c$. If we assume errors with symmetric distribution, A1 (vi) holds automatically when ψ is even. Clearly, the L_2 choice corresponds to taking $\rho(x) = x^2$ while the L_1 option is akin to opting for $\rho(x) = |x|$. These two estimators do no require the scale estimator $\widehat{\sigma}_n$. Note also that the class of M-estimators satisfying A1 does not include estimators with a redescending choice for ψ. We believe that the strict differentiability conditions on ρ required in A1 are not strictly necessary, but they make the proofs for the asymptotic theory simpler. Moreover, some functions ρ which are not twice differentiable everywhere such as $|x|$ or the Huber's functions defined below in (7) can be approximated by functions satisfying A1.

The asymptotic distribution of the L_2 isotonic estimators at a given point was found by Brunk (1970) and Wright (1981) and the one of the L_1 estimator by Wang and Huang (2002). They prove that the distribution of these estimators conveniently normalized converge to the distribution of the slope at zero of the greatest convex minorant of the two-sided Brownian Motion with parabolic drift. In this paper, we prove a similar result for isotonic M-estimators. The focus of this paper is on estimation of the trend term at a single observation point t_0. We do not address the issue of distribution of the whole stochastic process $\{\hat{\mu}_n(t), t \in \mathcal{T}\}$. Recent research along those lines are given by Kulikova and Lopuhaä (2006) and a related result with smoothing was also obtained simultaneously in Pal and Woodroofe (2006). The isotonic regression problem has been also addressed with modern techniques of empirical processes. We refer the reader here to Van de Geer (2000, p. 58), who shows the uniform consistency of the classical isotonic regression estimator within a class of functions utilizing methods based on entropy with bracketing.

In order to place this paper within the Statistical Literature, it is worth mentioning that the literature on estimation of a monotone regression function is by now enormous. In an attempt to describe it, one can broadly distinguish between two main approaches, based on the type and extent of smoothness assumed for the underlying trend function, denoted by $\mu(\cdot)$. The first approach, which we will call the *raw isotonic regression* approach, only assumes that μ is monotonic, with the result that its estimate $\hat{\mu}$ is a non-decreasing piecewise constant function. As mentioned, the main difference within the raw approach is the choice of ρ-function in Eq. (3), which yields either classical isotonic regression, median isotonic regression or the estimators that we propose here in this paper. Those three options also share the property that when t_0 is a point for which $\mu'(t_0)$ exists and is positive, the asymptotic distribution of $n^{-1/3}[\hat{\mu}(t_0) - \mu(t_0)]$, is proportional to the slope at zero of the greatest convex minorant of two-sided Brownian motion with parabolic drift. Better rates could be obtained within the classical approach at the expense of stronger smoothness assumptions (e.g., Leurgans, 1982). In this paper we present estimators that are highly efficient under Gaussian errors as well as highly robust when the error distribution has heavy tails, and we measure the quantitative robustness by the a version of the influence function and by the asymptotic breakdown point.

An important advantage of the raw isotonic approach is that it facilitates graphically the detection of change-points or abrupt changes in the trend $\mu(\cdot)$, which is of interest in many applications, such as the Global Warming data presented in Section 7. However, the estimate $\hat{\mu}(\cdot)$, being piecewise constant, is less attractive visually than alternative methods that yield smooth curves as estimates. Moreover, the lack of smoothness of the raw isotonic estimate seems unnatural in such situations where abrupt changes appear physically unlikely, such as plant or children growth. Accordingly, there has been interest in the literature in combining smoothness and monotonicity. Friedman and Tibshirani (1984), Mukherjee (1988) and Mammen (1991) were early contributors, and Ramsey (1998) is a more recent method. All these methods, to which we will generically refer as the *smooth isotonic approach* need to assume that $\mu(\cdot)$ is smooth. In this line of work there are two main choices, namely the selection of a method for smoothing and a technique to achieve a monotone estimate. The usual options for smoothing include local polynomials, splines, and kernel estimators. Each of these methods regulate the smoothness of the resulting $\hat{\mu}(\cdot)$ in its own way, for example by choosing the kernels and the bandwidth in the kernel