Journal of Statistical Planning and Inference 139 (2009) 828 – 842

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Simultaneous confidence bands for isotonic functions

Allan R. Sampson[a,∗,1], Harshinder Singh[b,2], Lyn R. Whitaker[c]

[a]*Department of Statistics, University of Pittsburgh, Pittsburgh, PA, USA*
[b]*West Virginia University, USA*
[c]*Operations Research Department, Naval Postgraduate School, Monterey, CA, USA*

ARTICLE INFO

Article history:
Received 19 July 2006
Received in revised form
18 December 2007
Accepted 12 May 2008
Available online 23 June 2008

Keywords:
Simultaneous confidence bands
Order restricted inference
Isotonic regression
Bandwidth

ABSTRACT

Procedures for simultaneous confidence bands are provided for bivariate regression data $\{x, Y_x, x \in \mathscr{X}\}$, where EY_x is an unknown monotonic function of x. A new bandwidth procedure is introduced which generalizes the previously introduced procedures of Korn [1982. Confidence bands for isotonic dose–response curves. Appl. Statist. 31, 59–63] and Lee [1996. On estimation for monotone dose–response curves. J. Amer. Statist. Assoc. 91, 1110–1119]. To gain insight into the comparability and applicability of these procedures a unifying geometric framework is used. This framework leads to certain optimality results for these banding procedures and allows for extensions to settings that have not been previously considered.

1. Introduction

Selected aspects of the theoretical and algorithmic development of statistical inference under order restrictions have been well developed over the past 50 years (see Barlow et al., 1972; Robertson et al., 1988). Estimation algorithms, hypothesis tests and related distribution theory results are areas that have received the most attention. On the other hand, data analysis techniques for order restricted inference have substantially lagged in development. The reasons for this are difficult to pinpoint, but as a result, the relative lack of applicable order restricted procedures make it difficult to examine the practicality of these procedures. This dichotomy of results is striking in the context of the explosive growth of computational methods in statistics.

One aspect of computational statistics that has received much consideration is the fitting of various types of semi-parametric and nonparametric curves and surfaces to data (e.g., Hastie and Tibshirani, 1990). The usefulness of any curve fitting procedure is enhanced by being able to describe the uncertainty of the fit assuming some reasonable error structure for the data. An effective and useful way to represent this uncertainty is to provide simultaneous confidence bands for the population curve (see Hastie and Tibshirani, 1990, Sections 3.8 and 5.4 for bands for generalized additive models).

Order restricted inference provides general nonparametric techniques to fit increasing or decreasing (isotonic) curves to data where we are only willing to assume the isotonicity of relationships. For bivariate data, which is the focus of this paper, highly efficient estimation algorithms are available. However, there does not seem to be a substantial use made of isotonic regression curves in data analysis . We suspect that among the reasons for this are the possible concerns about the naturally occurring "flat spots" of the isotonic regression line caused by pooling and the need for appropriate simultaneous confidence bands. For regression-like data where there are large amounts of continuous data, it is known that the problematic issue of "flat spots"

∗ Corresponding author.
E-mail address: asampson@pitt.edu (A.R. Sampson).

1 Research supported in part by the National Science Foundation, Grant No. DMS-0072207.

2 Harshinder Singh unfortunately passed away between the time of the preparation of this manuscript and its appearance; West Virginia University was his affiliation when he died.

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.05.045

decreases substantially. Moreover, as we show in this paper that there are simultaneous confidence banding procedures that provide useful and interpretable bands, particularly in this regression-like setting. With these confidence bands and our S-PLUS code implementing them, we provide tools for effective use of isotonic curve fitting.

Our specific purposes in this paper are multifold. Firstly, we want to give a unified presentation of three univariate simultaneous inference procedures of Korn (1982), Lee (1996), and Schoenfeld (1986), developed by them for the context of comparing ordered means for multiple treatment groups. The geometry of these three procedures is particularly explored to gain insights into applicability. Secondly, it is this geometry that suggests several generalizations. Here we focus on a natural generalization of Korn's (1982) and Lee's (1996) procedures, where we introduce the notion of "bandwidth" for simultaneous confidence bands. This notion conveys the same sense as the bandwidth ideas employed in the "smoother" context. The bandwidth confidence procedure is fully developed and shown to encompass both Korn's and Lee's procedures. Thirdly, the various procedures are compared, particularly with a focus on application in the regression setting, with heuristics provided for choice of the bandwidth.

Our more general long-term goals are to make available computational tools that allow us to realistically explore the potential usefulness of isotonic regression, both in the univariate setting considered here and ultimately extending them to the multivariate setting. The tools provided in this paper, the illustrative applications we consider, and the extensions of our procedure, for example, to non-monotonic and non-normal data suggest that these proposed techniques yield meaningful procedures.

Section 2 provides the conceptual framework for constructing confidence bands. Sections 3, 4, and 5 discuss, respectively, the simultaneous procedures of Korn (1982), Lee (1996), and Schoenfeld (1986) from the perspective of the framework. In Section 6, we provide a new banding procedure which generalizes Korn's and Lee's procedures. We conclude with a discussion in Section 7.

2. Framework for construction of confidence bands

For the observed data, $(x_i, Y_{x_i}(j)), j = 1, \ldots, n_{x_i}, i = 1, \ldots, k$, we suppose the underlying model:

$$Y_x(j) \overset{\text{indep}}{\sim} \mathrm{N}(\phi(x), \sigma^2), \quad j = 1, \ldots, n_x, \;\; x \in \mathscr{X},$$

where $\mathscr{X} = \{x_1, \ldots, x_k\}$ is a finite linearly ordered set, (for $x_i, x_j \in \mathscr{X}$ then $x_i < x_j$ or $x_j < x_i$), ϕ is an unknown function nondecreasing in x, and $\sigma^2 > 0$. We alternatively denote $\phi(x_i)$ by $\mu_i,\ i = 1, \ldots, k$. More specifically, $\mathscr{X}$ can be a set of real numbers or a set of ordered categories, so that we are examining in the first instance regression-like data and in the second instance data with an ordinal independent variable, e.g., dose–response data, where doses are designated low, medium, and high. Note that we have formulated our model in terms of $\mathscr{X}$ being a finite linearly ordered set to accommodate the realities of a regression-like data set. However, as we point out later in this section, we can view $\mathscr{X}$ simply as a linearly ordered set and can extend our finite sample bands to handle a continuous setting.

The estimation of $\phi(x)$ has been widely studied (e.g., Robertson et al., 1988). The usual estimator is $\hat{\phi}$, the isotonic regression of $\{Y_x(j)\}$ on $\mathscr{X}$, where $\hat{\phi}$ minimizes, among nondecreasing $\phi(x)$,

$$\sum_{x \in \mathscr{X}} \sum_{j=1}^{n_x} [Y_x(j) - \phi(x)]^2.$$

Equivalently, we find $\hat{\phi}$ by minimizing

$$\sum_{x \in \mathscr{X}} [\bar{Y}_x - \phi(x)]^2 n_x,$$

over $\phi(x)$ nondecreasing, where $\bar{Y}_x = \sum_{j=1}^{n_x} Y_x(j)/n_x$. Let $\hat{\sigma}$ be an estimator of σ, which is independent of $\{\bar{Y}_x, x \in \mathscr{X}\}$. A standard estimator of σ^2, which ignores the monotonicity of $\phi(x)$, is $\sum_{x \in \mathscr{X}} \sum_{j=1}^{n_x} (Y_x(j) - \bar{Y}_x)^2/(n-k)$, where $n = \sum_{x \in \mathscr{X}} n_x$; this is clearly an unbiased estimator when all n_x's are greater than 1. A more generalized discussion of estimation of σ^2 taking into account the monotonicity of $\phi(x)$ is given by Sampson et al. (2003). In this paper we assume that no matter how it is chosen, $\hat{\sigma}^2$ is meaningful, that $v\hat{\sigma}^2/\sigma^2$ has a χ^2_v distribution, and the standard independence condition holds. For the examples that we consider, either we assume for convenience that σ^2 is known to be 1, or suggest suitable approximations.

Our goal throughout this paper is to produce confidence bands around $\phi(x), x \in \mathscr{X}$, with an eye to their application in a "regression" setting. This translates into the need to produce a rectangular confidence region for $\phi(x), x \in \mathscr{X}$, where the sides of the region are aligned with the axes and, moreover, the upper confidence band for $x \in \mathscr{X}$ is nondecreasing and similarly so the lower confidence band. The concern to produce such bands or rectangular regions has been discussed in a related context by Hastie and Tibshirani (1990, Section 3.8.2).

Note that throughout this paper, even if not explicitly stated, we always assume that *all* rectangles that we consider are aligned with the axes.