Journal of Statistical Planning and Inference 139 (2009) 696 – 710

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Likelihood-based confidence sets for partially identified parameters

Zhiwei Zhang

Division of Biostatistics, Center for Devices and Radiological Health, U.S. Food and Drug Administration, HFZ-550, Rockville, MD 20850, USA[1]

ARTICLE INFO

Article history:
Received 4 December 2006
Accepted 7 August 2007
Available online 29 May 2008

Keywords:
Censoring
Confidence set
Likelihood
Missing data
Partial identification
Verification bias

ABSTRACT

There has been growing interest in partial identification of probability distributions and parameters. This paper considers statistical inference on parameters that are partially identified because data are incompletely observed, due to nonresponse or censoring, for instance. A method based on likelihood ratios is proposed for constructing confidence sets for partially identified parameters. The method can be used to estimate a proportion or a mean in the presence of missing data, without assuming missing-at-random or modeling the missing-data mechanism. It can also be used to estimate a survival probability with censored data without assuming independent censoring or modeling the censoring mechanism. A version of the verification bias problem is studied as well.

Published by Elsevier B.V

1. Introduction

Identifiability of a parameter means, roughly, that the parameter relates to the distribution of the observed data in a one-to-one fashion. Without identifiability, a consistent point estimator does not exist, and many good properties of a point estimator become impossible. Consequently, most of the modern theory of statistical inference requires that the inferential target be identifiable. When the parameter of interest is not fully identifiable from the observed data, it is common practice to impose additional assumptions or constraints that reduce the probability model and help identify the parameter. Such assumptions rely on external information such as prior knowledge and cannot be validated with the data alone. In reality, reliable external information is often unavailable, and identifying assumptions frequently driven by practical rather than scientific considerations. Because different assumptions may lead to different conclusions, it makes sense to compare results obtained under different assumptions if no single identifying assumption is strongly preferred. This practical approach, called sensitivity analysis, nonetheless lacks scientific rigor. It is usually impossible to enumerate all possible identifying assumptions, and often difficult to conduct and interpret a sensitivity analysis in a systematic and objective manner.

Example (*Proportion*). To fix ideas, consider the problem of estimating a proportion with missing data. Suppose that X is a Bernoulli variable, and that the parameter of interest is $\gamma = P(X = 1)$, the probability of success. If X is always observed, then γ is completely identified from a random sample of X. Suppose, however, that X is potentially missing, which may happen because of nonresponse in surveys, for instance. Let R be the observation indicator, so $R = 1$ if X is observed and 0 otherwise. Without additional assumptions, γ is not identifiable. One common identifying assumption is missing completely at random (MCAR) in the sense of Rubin (1976), namely that R and X are independent. Alternatively, a selection model could be specified for the conditional distribution of R given X, or a pattern mixture model for the conditional distribution of X given R; see, for example,

E-mail address: zhiwei.zhang@fda.hhs.gov.

[1] The views expressed in this article do not necessarily represent those of the U.S. Food and Drug Administration.

0378-3758/$ - see front matter Published by Elsevier B.V
doi:10.1016/j.jspi.2007.08.009

Little and Rubin (2002). None of these assumptions can be tested with the observed data, and it is natural to ask exactly what can be inferred about γ without imposing untestable assumptions.

Questions like this can be addressed under a different approach which treats the parameter of interest as partially identified. The notion of partial identification has originated and become increasingly popular in the econometrics literature; see Manski (2003) for an extensive survey. Much of the early research on partial identification has focused on characterization of identification regions (or "bounds" for one-dimensional parameters). However, inferential procedures have been developed for one-dimensional parameters by Horowitz and Manski (1998, 2000), who propose Bonferroni-type and bootstrap confidence intervals for the identification region, and Imbens and Manski (2004), who provide a general construction of Wald-type confidence intervals for the parameter of interest. More recent research efforts have been directed toward general vector-valued parameters. Andrews et al. (2004) consider estimation of the profit function of firms from "revealed choice" data on entry in a cross-section of oligopoly markets. Here the identification region is characterized by a set of moment inequalities, and a bootstrap procedure is used to obtain confidence sets for the partially identified parameter. Chernozhukov et al. (2004) and Shaikh (2005) study problems where the identification region is defined as the set of minima of a criterion function, and use a sample analogue of the criterion function to construct confidence sets for the identification region and for the parameter of interest. For models composed of a finite number of moment inequalities, Rosen (2006) develops confidence sets for a partially identified parameter by inverting a Wald-type test of composite hypotheses. Beresteanu and Molinari (2006) regard empirical estimators of identification regions as set-valued random variables and establish an asymptotic theory which can be used to derive tests and confidence sets.

The present paper proposes a method for constructing confidence sets for a class of partially identified models where data are incompletely observed, due to nonresponse or censoring, for instance. The proposed method is based on a likelihood for the observed data and applies to vector-valued parameters. It relates naturally to the independent work of Chernozhukov et al. (2004) and Shaikh (2005) based on criterion functions, and the relationship will be discussed in Section 6 after the likelihood-based approach is described and illustrated.

In the next section, I set up the notation for the incomplete data problem and, for readers unfamiliar with partial identification, give a brief introduction to the concept and its implications. Then, in Section 3, the proposed likelihood-cased procedure is presented. The proportion example described earlier will be used throughout this discussion as an illustration. Three more examples (mean, censoring and verification bias) are studied in Section 4. Numerical results, which consist of a simulation study and a hypothetical example, are reported in Section 5. A discussion is given in Section 6. Proofs are either omitted or relegated to Appendix A.

2. Preliminaries

Partial identification due to incomplete data has been studied extensively, and many examples, including the one introduced earlier, can be found in Manski (2003). In this section I attempt to formulate the problem in some generality. The notation given below facilitates a general discussion of partial identification and will be useful in describing the proposed method. I also discuss the implications of partial identification on statistical inference. The results in this section are straightforward and may have been observed, but they seem worth noting to readers unfamiliar with the concept of partial identification, especially because much of this discussion appears difficult to locate in the literature.

2.1. Partial identification with incomplete data

Let Y^* be a random variable taking values in $\mathscr{Y}^*$ and let $\mathscr{P}^*$ denote the set of all possible distributions of Y^*. Suppose the parameter of interest is $\gamma = T(P^*)$, where T is a known function. The range of γ is a subset of a Euclidean space in the examples to follow, but this is not an absolute requirement. If Y^* is observed, then γ is trivially identifiable. Suppose, however, that Y^* is not directly observed; instead we observe $Y = h(Y^*)$, where $h : \mathscr{Y}^* \to \mathscr{Y}$ is a known transformation. Then the induced model for Y is $\mathscr{P} := \{P^* h^{-1} : P^* \in \mathscr{P}^*\}$. It is often convenient to parameterize $\mathscr{P}$ as $\{P_\theta : \theta \in \Theta\}$, where Θ may be finite- or infinite-dimensional. Assume that θ is identifiable in the usual sense that $\theta_1 = \theta_2$ whenever $P_{\theta_1} = P_{\theta_2}$. This allows us to define $S : \mathscr{P}^* \to \Theta$ by

$$P_{S(P^*)} = P^* h^{-1}, \quad P^* \in \mathscr{P}^*,$$

in other words, θ is determined by P^* through the known function S.

The parameters θ and γ need not determine each other; they are related through their common dependence on the underlying distribution P^* of Y^*. A pair of values (θ, γ) will be said to be *compatible* if they can possibly obtain together, that is, if there exists $P^* \in \mathscr{P}^*$ such that $S(P^*) = \theta$ and $T(P^*) = \gamma$. If the value of θ is known, then γ must be one of those values compatible with θ. The collection of such values is called the *identification region* for γ given θ. Formally, this is written as

$$\Gamma|_\theta = TS^{-1}(\theta) = \{T(P^*) : S(P^*) = \theta\},$$

where the superscript -1 denotes inverse image. Conversely, if the value of γ is given, then θ must be one of those values compatible with γ. This latter set of values is called the *inverse identification region* for θ given γ and is denoted by

$$\Theta|_\gamma = ST^{-1}(\gamma) = \{S(P^*) : T(P^*) = \gamma\}.$$