Journal of Econometrics 178 (2014) 613–623

Contents lists available at ScienceDirect

Journal of Econometrics

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

On empirical likelihood statistical functions

Ao Yuan *, Jinfeng Xu, Gang Zheng

Georgetown University, Washington DC 20057, USA
New York University, NY 10016, USA
National Heart, Lung and Blood Institute, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:
Received 9 June 2010
Received in revised form
19 December 2012
Accepted 30 August 2013
Available online 9 September 2013

JEL classification:
C14

Keywords:
Empirical likelihood
Quantile estimation
Uniform SLLN
Uniform CLT

ABSTRACT

We consider the empirical likelihood method for estimation of distribution and quantile functions where side information is incorporated through moment conditions. We systematically study the asymptotic properties of the estimators, such as the uniform strong laws of large numbers and weak convergence over classes of functions. Two Monte Carlo examples are also given to illustrate the practical utility of the method.

1. Introduction

Empirical likelihood (EL) can be traced all the way back to Thomas and Grunkemeier (1975), but Owen (1988, 1990, 1991, 2001) formally established it as a nonparametric analogue of parametric likelihood, providing an effective way for interval estimation and goodness-of-fit test with appealing data-driven and range respecting features. DiCiccio et al. (1991) and Chen and Cui (2007) further showed that EL is generally Bartlett correctable and thus leads to more accurate inferential procedures than some other commonly used ones such as the bootstrap. The EL method has been extended and applied in many fields in statistics and econometrics such as estimating equations (Qin, 1993; Qin and Lawless, 1994), quantile and density estimation (Chen and Hall, 1993; Hall and Presnell, 1999; Chen, 1997; Zhang, 1997, 1998), generalized linear model (Kolaczyk, 1994), survival analysis (Murphy and van der Vaart, 1997; Adimari, 1997), nonparametric regression (Qin and Tsao, 2005), goodness-of-fit measure (Baggerly, 1998), inference in the presence of nuisance parameters (Lazar and Mykland, 1999), marginal and conditional likelihood (Qin and Zhang, 2005), finite population inference (Chen and Qin, 1993; Chen and Wu, 2002), specification tests (Kitamura, 2001; Kitamura et al., 2004), generalized methods of moments (Imbens, 2002), local polynomial fitting (Zhang and Liu, 2003), and dual likelihood (Bravo, 2004).

In many situations of practical interest, there usually exists some side information about the otherwise unknown distribution of the sample. This information may come from preliminary studies or be directly implied by an assumed theoretical model. For example, we might know that the distribution function is symmetric around a constant which can either be specified or unknown. Intuitively, the side information "symmetry" should be utilized when it comes to estimate the distribution function. An attractive feature of the EL method is that it can effectively incorporate such information through the moment functions. For the problem where the information about distribution function F and a parametric vector θ associated with F are available in the form of some unbiased estimating functions, Qin and Lawless (1994) formally established the asymptotic normality of the EL estimators of F and θ. When the side information does not involve an unknown parameter vector, Chen and Qin (1993) and Zhang (1997) proposed the EL-based estimator of the quantile function and established its large sample properties.

In this paper, we systematically study the asymptotic properties of the EL-based estimators of distribution and quantile functions in the presence of side information which may involve an unknown parameter vector. For the EL-based estimator of the distribution function, we establish the uniform strong laws of large numbers and weak convergence, over classes of functions, i.e. the P-Glivenko–Cantelli and P-Donsker classes respectively. The latter allows us to construct nonparametric confidence bands for the distribution function and its related functionals. The EL-based estimator of the quantile function is obtained by direct inversion and the asymptotic results parallel to those of the distribution function are

* Correspondence to: 2216 Sixth Street, N.W., Suite 206, Washington DC 20059, USA. Tel.: +1 202 806 4361.
E-mail address: yuanao@hotmail.com (A. Yuan).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.08.037

established. Two Monte Carlo examples are also given to illustrate that for distribution and quantile functions, the estimation accuracy can generally be improved by incorporating the side information. The rest of the paper is organized as follows. In Section 2, we study the EL-based estimation of the distribution function, in the presence of side information. Under the same setting, we consider the EL-based quantile estimation in Section 3. Section 4 provides two Monte Carlo examples to illustrate the results. All the technical proofs are relegated to the Appendix.

2. The EL-based estimation of the distribution function

Let $X_1, \ldots, X_n$ be d-dimensional independent and identically distributed random vectors with an unknown distribution function F and a parameter θ associated with F which takes values in a compact parameter space $\Theta \subseteq R^q$. Denote the true values of F and θ by F_0 and θ_0 respectively. As in Qin and Lawless (1994), suppose that the side information involves an unknown parameter vector θ and can be summarized in the form of $r \geq q$ unbiased moment functions $g_j(x, \theta), j = 1, 2, \ldots, r$, such that $Eg(X, \theta_0) = 0$, where the expectation is for the distribution of X at the true parameter θ_0 when no confusion appears, and $g(X, \theta) = (g_1(X, \theta), g_2(X, \theta), \ldots, g_r(X, \theta))^T$. Let $w_1, w_2, \ldots, w_n$ be non-negative empirical weights allocated to the observations. The EL incorporating the side information as proposed in Qin and Lawless (1994) is $L(\theta) = \prod_{i=1}^n w_i(\theta)$ which maximizes

$$\prod_{i=1}^n w_i \quad \text{subject to } w_i \geq 0, \qquad \sum_{i=1}^n w_i = 1, \quad \text{and}$$

$$\sum_{i=1}^n w_i g(X_i, \theta) = 0;$$

which are given by

$$w_i(\theta) = \frac{1}{n} \frac{1}{1 + t_n^T(\theta) g(X_i, \theta)},$$

and $t_n(\theta) = (t_{n,1}(\theta), \ldots, t_{n,r}(\theta))^T$ are solutions of

$$\sum_{i=1}^n \frac{g(X_i, \theta)}{1 + t_n^T(\theta) g(X_i, \theta)} = 0.$$

Write $\Omega = E[g(X_1, \theta_0) g(X_1, \theta_0)^T]$, $L = E[\partial g(X_1, \theta_0)/\partial\theta]$, and $A(x) = E[g(X_1, \theta_0) I(X_1 \leq x)]$. Let $I(\cdot)$ be the indicator function. Without any side information, distribution function can be efficiently estimated by the empirical distribution function $F_n(x) = \sum_{i=1}^n I(X_i \leq x)/n$, where the inequality is meant componentwise. The asymptotic variance of $\sqrt{n}(F_n(x) - F_0(x))$ is $\sigma^2(x) = F_0(x)(1 - F_0(x))$. When θ_0 is known, the EL estimator of distribution function is defined as

$$\hat{F}_n(x) = \sum_{i=1}^n w_i(\theta_0) I(X_i \leq x),$$

with the associated EL process $\sqrt{n}(\hat{F}_n(\cdot) - F_0(\cdot))$. When θ_0 is unknown, it can be estimated by the maximum empirical likelihood estimate $\tilde{\theta}_n = \arg\max L(\theta)$. The corresponding EL estimator of distribution function and the associated EL process are then

$$\tilde{F}_n(x) = \sum_{i=1}^n w_i(\tilde{\theta}_n) I(X_i \leq x) \quad \text{and}$$

$$\sqrt{n}(\tilde{F}_n(\cdot) - F_0(\cdot)), \text{ respectively.}$$

It is natural to estimate F_0 by $\tilde{F}_n$ or $\hat{F}_n$ since the w_is are the weights that the EL distribution places on the observations X_is. As noted by a referee, alternatively, for any given x, write $\tau = F(x)$ ($\tau_0 = F_0(x)$ is the true value) and use the EL method to estimate $\pi = (\tau, \theta)$. The moment functions for π are $m(X, \pi) = (I(X \leq x) - \tau, g(X, \theta)^T)^T$. It is easy to see that the maximum empirical likelihood estimate of π is $(\tilde{F}_n(x), \tilde{\theta}_n)$ when θ_0 is unknown. As shown in Qin and Lawless (1994), the asymptotic variance of $\sqrt{n}(\tilde{F}_n(x) - F_0(x))$ is $\tilde{\sigma}^2(x) = F_0(x)(1 - F_0(x)) - A(x)^T(\Omega^{-1} - U)A(x)$, where U is given in Theorem 1. When θ_0 is known, $L = 0$ and the asymptotic variance of $\sqrt{n}(\hat{F}_n(x) - F_0(x))$ reduces to $\hat{\sigma}^2(x) = F_0(x)(1 - F_0(x)) - A(x)^T\Omega^{-1}A(x)$. The asymptotic variances are equal to the efficiency bounds for τ, as derived for the general GMM case by Chamberlain (1987) and thus $\tilde{F}_n$ and $\hat{F}_n$ are semiparametrically efficient. This shows that the weights $w_i(\tilde{\theta}_n)$ and $w_i(\theta_0)$ are optimal although they are both of order $1/n(1 + o_p(1))$ the same as those of the NPMLE. When $A(x) \neq 0$, $\hat{\sigma}^2(x) < \sigma^2(x)$ and this indicates that there is always an efficiency gain in incorporating the side information when θ_0 is known. When θ_0 is unknown, however, there could be no efficiency gain for some x. As illustrated in the first example of Section 4, when the side information is "symmetry" and the underlying distribution is the Laplace distribution, there is no efficiency gain in estimating $F_0(\theta_0)$ where θ_0 is the unknown symmetry center.

Below we introduce some notations before the statements of our results. Let μ be a dominating measure for F_0 (For example we may take μ be the Lebesgue measure λ if F_0 is continuous, the counting measure C if F_0 is discrete, or $\mu = \lambda|A + C|A^c$ if F_0 is continuous on A and discrete on A^c). For ease of exposition we assume F_0 is on R^1, the case F_0 being on R^d ($d > 1$) is similar. Let $f = dF/d\mu$ be the Radon–Nikodym derivative of F_0 with respect to μ, and

$$\tilde{f}_n(x) = \frac{\tilde{F}_n(x + c_n) - \tilde{F}_n(x - c_n)}{2c_n},$$

$$\hat{f}_n(x) = \frac{\hat{F}_n(x + c_n) - \hat{F}_n(x - c_n)}{2c_n},$$

with $c_n \to 0$, $nc_n \to \infty$. Define $f_n(x)$ accordingly using $F_n(x)$. Note that $\tilde{f}_n(\cdot)$, $\hat{f}_n(\cdot)$ and $f_n(\cdot)$ are proper densities for each n. Now let $\tilde{\mathscr{P}}_n$, $\hat{\mathscr{P}}_n$ and $\mathscr{P}_n$ be the corresponding (random) probability measures for $\tilde{f}_n(\cdot)$, $\hat{f}_n(\cdot)$ and $f_n(\cdot)$. We will establish contiguity among these probability measures.

Let $\tilde{P}_n$, $\hat{P}_n$, P_n and P be the (random) probability measures induced by $\tilde{F}_n$, $\hat{F}_n$, F_n and F_0 respectively. For a function $h : R^d \mapsto R$, define $\tilde{P}_n h = \sum_{i=1}^n w_i(\tilde{\theta}_n) h(X_i)$, $Ph = Eh(X_1)$, $\tilde{\mathbb{G}}_n h = \sqrt{n}(\tilde{P}_n h - Ph) = \sqrt{n}(\sum_{i=1}^n w_i(\tilde{\theta}_n) h(X_i) - Eh(X_1))$, and $\mathbb{G}_n h = \sqrt{n}(P_n h - Ph) = \sqrt{n}(\sum_{i=1}^n n^{-1} h(X_i) - Eh(X_1))$. For any given h satisfying some regularity condition, it can be shown that $\tilde{P}_n h \to Ph$ (a.s.) and $\tilde{\mathbb{G}}_n h \xrightarrow{D} N(0, \tilde{\sigma}_h^2)$, where $\tilde{\sigma}_h^2 = Ph^2 - (Ph)^2 - P(g_0^T h)(\Omega^{-1} - U)P(g_0 h)$ and $g_0(x) = g(x, \theta_0)$. By contrast, $\mathbb{G}_n h \xrightarrow{D} N(0, \sigma_h^2)$ with $\sigma_h^2 = Ph^2 - (Ph)^2$. Hence when it comes to estimate Ph, incorporating the side information g reduces the asymptotic variance by the amount $P(g_0^T h)(\Omega^{-1} - U)P(g_0 h)$. For a given space $\mathscr{H}$, let $l^\infty(\mathscr{H})$ be the set of functions z on $\mathscr{H}$ with $\sup_{h\in} |z(h)| < \infty$. We aim to establish the uniform strong laws of large numbers and weak convergence of empirical likelihood process as random elements indexed by a class of functions $\mathscr{H}$, in the space $l^\infty(\mathscr{H})$, which may not be measurable. Accordingly, the resultant convergence results are in the sense of the outer measure P^* of P (van der Vaart and Wellner, 1996; hereafter VW). When $\mathscr{H}$ is measurable, the results are automatically in the sense of the measure P itself. For any probability measure Q, let $\|Q\|_{\mathscr{H}} = \sup\{|Qh| : h \in \mathscr{H}\}$. Define for any vector $k = (k_1, \ldots, k_d)$ of d integers and a function $h : \mathcal{X} \mapsto R$, $D^k h(x) = \partial^{|k|} h(x)/(\partial x_1^{k_1} \cdots \partial x_d^{k_d})$, where $\mathcal{X}$ is a subset of R^d and $|k| = k_1 + \cdots + k_d$. Let

$$\|h\|_s = \max_{|k| \leq s} \sup_{x \in \mathcal{X}} |D^k h(x)| + \max_{|k| = s} \sup_{x, y \in \mathcal{X}} |D^k h(x) - D^k h(y)|$$