Bias Deconvoving Kernel Distribution Estimators

Ioannides, Dimitris
University of Macedonia, Department of Economics
Egnatias 156
Thessaloniki (54006), Greece
E-mail: dimioan@uom.gr

1 Introduction

Measurement error potentially affects all statistical analysis, because it causes the probability distribution that generates the observable data to deviate from that which generates the unobservable. The effect of measurement error on the properties of estimators and testing procedures, particularly in the context of parametric models are studied by many authors. Fuller (1987) give detailed accounts of this work and comprehensive literature. The main effect of the error in the parametric estimators is that these are biased. On the opposite side, in the context of nonparametric models Fan (1991) ,proposed unbiased nonparametric estimators for the density function, distribution function and nonparametric regression function . The present paper focuses on the estimation of distribution functions and studies the impact of measurement error in the context of nonparametric estimation. This problem was first studied in Fan (1991), in Ioannides and Papanastasiou (2001), Hesse (1995) and more recently by Hall and . Lahiri (2009) and . Kulik (2009). All the above authors studied asymptotic properties of the nonparametric estimator for the distribution proposed by Fan (1991) in the i.i.d case or in the dependent case assuming that the variance σ^2 of the measurement error is fixed . Delaigle (2008) study a related problem by considering that the variance σ^2 of the measurement error goes to 0 . Here, we are using a modify estimator of Fan which is biased and we calculate its bias in terms of the error variance σ^2. Thus for $\sigma^2 \to 0$ the bias is negligible , and the main source of error for the approximation of the target distribution comes from its nonparametric kernel estimator . For σ^2 large there is an additional error which is calculated from the second derivative estimator. Our paper is organized as follows. In Section 2 the construction of our estimator is given. Our results are presented in Section 3 with the general assumptions under which these are obtained.

2 Construction of the estimator

The deconvolution kernel estimator was first considered by Carroll & hall (1988). Let us recall briefly here its construction. the variable of interest X is measured with error and is not directly observable. Instead X is observed through

$$Y = X + \varepsilon \qquad (2.1)$$

It is assumed that the variable ε has a known distribution and is independent of X . Given a random sample could we estimate the density of X ?

Denote by $\Phi_Y(t)$, $\Phi_X(t)$ and $\Phi_\varepsilon(t)$) the characteristic function of Y (resp. X and ε). We have $\Phi_Y(t) = \Phi_X(t)\Phi_\varepsilon(t)$.

If we assume to have a non-vanishing characteristic function, then by Fourier inversion, the density of X is given by

$$f_x(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\frac{\Phi_Y(t)}{\Phi_\varepsilon(t)}dt \qquad (2.2)$$

where

$$\Phi_Y(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(\text{-itu}) \mathrm{f}_Y(\mathrm{u})du .$$

A classical nonparametric estimator of $\mathrm{f}_Y(u)$ is given by, $\hat{f}_{nY}(u) = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{u-Y_i}{h_n})$ (2.3)

Where K is the kernel and h_n is the bandwidth. In order to obtain an estimator for $\Phi_Y(t)$, we replace $\mathrm{f}_X(x)$ by its nonparametric estimator. Using this last estimator in equation (2.2), we get an estimator of (2.2) by $\hat{f}_{nX}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} W_n(\frac{x-Y_i}{h_n})$, (2.4)

where the deconvolution kernel is given by

$$\mathrm{W}_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\Phi_\mathrm{Y}(t)}{\Phi_\varepsilon(t/\mathrm{h_n})}dt$$

In this paper we are interesting to estimate the distribution function of X. This was done by Fan (1991) leading to:

$$\hat{F}_n(x) = \int_A \hat{f}_{nX}(z)dz \qquad (2.5)$$

with $A = (-\infty, x]$ We can rewrite (2.5) as :

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} D_n(Y_i) \qquad (2.6)$$

where $D_n(Y_i) = \int_A W_n(\frac{z-Y_i}{h_n})dz$

with $A = (-\infty, x]$. Under some conditions $Wn(x)$ is integrable, and thus as an estimator for the distribution of X can be considered the modify estimator of Fan $\overline{F_n}(x) = \frac{\hat{F}n(x)}{\int_{-\infty}^{+\infty} W_n(u)du}$.

3 Assumptions and Results

The model that we assume for the data is as follows. The observable r.v.'s are

$$Y_j = X_j + \varepsilon_{j,}\ i = 1,\ldots,n,$$

where $X_1,\ldots,X_n$, are the first n r.v.'s from the stationary process $\{X_j\}, j \geq 1$, with continuous d.f. F_X and $\varepsilon_1,...\varepsilon_n$ the first n r.v.'s from the noise stationary process $\{\varepsilon_j\}, j \geq 1$, which is independent of the unobservable process $\{X_j\}, j \geq 1$, and has a known probability density function . The distribution of each Y_j is the convolution of the distributions of X_j and ε_j, and thus as a nonparametric estimator of F_X can be considered the modify estimator of Fan given in (2.6) , with the only difference that here our data $Y_1,...,Y_n$, are the first n r.v.'s from the stationary process $\{Y_j\}, j \geq 1$.

In this section , we establish the asymptotic normality for our estimator $\overline{F_n}(x)$. We impose the following assumptions which are summarized here for easy reference.

Assumption (A)

(i) The process $\{X_j, \varepsilon_j\}, j \geq 1$ is strictly stationary and $\rho-$mixing with mixing coefficient $\rho(j) = O(j^{-k})$, for $k > 2$.

(ii) The process $\{X_j\}, j \geq 1$ is strictly stationary and independent to the strictly stationary process $\{\varepsilon_j\}, j \geq 1$

(iii) The density f is twice differentiable and support on R.

Assumption (D)

We assume that the integral $\int_{-\infty}^{+\infty} W_n(u)du$ exists for every n, as $n \to \infty$.

Remark 3.1. Conditions under which Assumption (D) holds are given in Fan (1991).

Theorem 3.1. If assumptions (A) and (D) are satisfied and the second derivative of the density function is bounded, then: $E\overline{F}_n(x) - F_X(x) - \sigma^2 F_X''(x) \to 0$, as $n \to \infty$.

Proof : Working similar as in Ioannides & Papanastasiou (2001) (Lemma 3.1 (i)) and using relation (2.4) from A. Chesher(1991) our result follows immediately .

Theorem 3.2. If assumptions (A) and (D) are satisfied and the second derivative of the density function is bounded, then $Var(\overline{F_n}(x)) \to F_X(x)(1 - F_X(x))$

as $n \to \infty$.

Proof: Similar as in Ioannides & Papanastasiou (2001) (Lemma 3.1 (ii)).

Theorem 3.3: If assumptions (A) and (D) are satisfied and the second derivative of the density function is bounded, then : $\overline{F}n(x) \mapsto N(0,1)$, in distribution as $n \to \infty$.

Proof : The result follows easily if we are working similar as in Theorem 2.1 of Ioannides & Papanastasiou (2001).

REFERENCES (RÉFERENCES)

R.J Carroll, Measurement error in nonlinear models: a modern perspective, CRC Press, 2006.

A. Chesher, The effect of measurement error. Biometrika, 78(3), 451, 1991.

J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems. The Annals of Statistics, 19(3), 1257–1272, 1991.

4.W.A. Fuller,. et al., Measurement error models, Wiley Online Library, 1987.

P. Hall, and S.N. Lahiri, Estimation of distributions, moments and quantiles in deconvolution problems. Annals of Statistics, 36(5), 2110–2134, 2008.

C. Hesse, Distribution function estimation from noise observation, Pulp. Inst. Stat. Univ. Paris, XXXIX fasc. 3, 21-35, 1995.

D.A. Ioannides, and D.P. Papanastassiou, Estimating the distribution function of a stationary process involving measurement errors. Statistical inference for stochastic processes, 4(2), 181–198, 2001.

R. Kulik, Nonparametric deconvolution problem for dependent sequences. Electronic Journal of Statistics, 2, 722–740, 2008.

ABSTRACT

A nonparametric estimator for a smooth distribution function based on contaminated observations was first considered by Fan. A method is developed to establish asymptotic normality results for the modified estimator and for weakly stochastic processes corrupted by some noise process. Our estimator is biased, and the bias part of is calculated under some weak conditions. The asymptotic normality is obtained under very general assumptions on the error characteristic function, which generalizes previous conditions on this topic. Some simulations results are given.

Keywords: *deconvolution, nonparametric estimation, distribution function, noise distribution*

.