E C O N O M I C S   B U L L E T I N

k nearest-neighbor estimation of inverse density weighted expectations

David Jacho-Chávez
Indiana University

Abstract

This letter considers the problem of estimating expected values of functions that are inversely weighted by an unknown density using the k-Nearest Neighbor method. L^2-consistency is established. The proposed estimator is also shown to be asymptotically semiparametric efficient. Some limited Monte Carlo experiments show that the proposed estimator performs as good as alternative methods in finite sample applications.

We would like to thank Juan Carlos Escanciano for providing comments and suggestions. We also acknowledge the usage of the Libra and Quarry High Performance Clusters at Indiana University where the computations were performed. All errors are our own.
Citation: Jacho-Chávez, David, (2008) "k nearest-neighbor estimation of inverse density weighted expectations." *Economics Bulletin*, Vol. 3, No. 48 pp. 1-6
Submitted: August 3, 2008. **Accepted:** August 21, 2008.
URL: http://economicsbulletin.vanderbilt.edu/2008/volume3/EB-08C40005A.pdf

1 Introduction

In this letter we address the problem of estimating quantities of the form

$$\theta_0 = \mathbb{E}\left[\frac{Y}{f(X)}\right], \tag{1.1}$$

where $f(X)$ represents the unknown marginal density of a continuous scalar random variable X, $Y \in \mathbb{R}$, and $\mathbb{E}$ represents expectation with respect the joint distribution of (Y, X). This problem is important because many existing semiparametric estimators of limited dependent variable models make use of inverse density-weighted expectations like 1.1, e.g. Lewbel (1998), Lewbel (2000), Lewbel (2006), and Khan and Lewbel (2007).

If $\{Y_i, X_i\}_{i=1}^n$ represents a random sample from this distribution, a natural estimator of 1.1 is

$$\theta_n = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{f_n(X_i)}, \tag{1.2}$$

where $f_n(X_i)$ denotes the k-Nearest Neighbor (k-NN) density estimator of $f(X_i)$, i.e.

$$f_n(X_i) = \frac{k}{2nR_n(X_i, k)}, \tag{1.3}$$

and

$$R_i \equiv R_n(X_i, k) \stackrel{\text{def}}{=} \text{the Euclidean distance between } X_i \text{ and the } k\text{-th nearest}$$
$$\text{neighbor of } X_i \text{ among all the } X_j\text{'s for } j \neq i$$

for $j = 1, \ldots, n$. Therefore, estimator 1.2 can be re-written as

$$\theta_n = \frac{2}{k}\sum_{i=1}^{n} Y_i R_n(X_i, k), \tag{1.4}$$

The usage of nonparametric k-NN estimator of $f(X)$, in place of a kernel estimator for example is particularly helpful in 1.2, because 1.4 is theoretically easier to handle than 1.2, since it does not involve the ratio of two random quantities. Another important advantage of the k-NN approach is its local adaptation, a property that is not enjoyed by the kernel method for example.

1.1 An Ordered Data Estimator

The drawbacks of the kernel method partly motivated Lewbel and Schennach (2007) to propose an estimator of 1.1 based on nearest neighbor *spacings* as follows:

$$\widetilde{\theta}_n = \sum_{i=1}^{n-k} y_{[i]}\left(x_{[i+k]} - x_{[i]}\right)/k, \tag{1.5}$$

where $(y_{[i]}, x_{[i]})$ denote the ith observation when the data are sorted in increasing order of x, i.e. $x_{[i]}$ is the ith order statistics and $y_{[i]}$ is its concomitant. Lewbel and Schennach (2007) showed that under certain regularity conditions, $\sqrt{n}(\widetilde{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathbb{E}[\mathrm{var}(Y|X)/f^2(X)])$ when $k = \mathrm{o}(\ln n)$ as $k \to \infty$. They derived the semiparametric efficiency bound for regular estimators of θ_0 and proves that $\widetilde{\theta}_n$ achieves it.

Although similar in nature, estimators 1.4 and 1.5 are fundamentally different. In particular, k in 1.4 refers to the k-th order statistic from the (conditionally on X_i) i.i.d. sample $\{\|X_i - X_j\|\}_{j=1}^{n-1}$ with $i \neq j$, while k in 1.5 refers to the k-th order statistic from the original i.i.d. sample $\{X_i\}_{i=1}^n$. These differences also make their limiting distribution theory not applicable for fixed or increasing k.

In the next section, it is shown that if $k = k(n)$ is a predetermined sequence of positive integers, not dependent on the sample $\{Y_i, X_i\}_{i=1}^n$, such that $k \to \infty$, and $k/n \to 0$ as $n \to \infty$, then estimator 1.4 is also $\sqrt{n}$–consistent, and semiparametric efficient. A small Monte Carlo experiment confirms these predictions.

2 Asymptotic Properties

The derivation of the asymptotic properties of θ_n defined in 1.2–1.3 relies on the following assumptions:
Assumption A:

(A1) $\inf_{x \in \mathcal{F}} f(x) \geq \eta > 0$, where Ω represents the finite support of $f(x)$.

(A2) (i) $f(x)$ is continuously differentiable up to the second order over the interior of $\otimes$. (ii) $\mathbb{E}[\mathrm{var}(Y|X = x)/f^2(X)] < \infty$.

(A3) As $n \to \infty$, $k \to \infty$, and $k/n \to 0$.

Assumption A1 is usually made for inverse density weighted estimators such as 1.2, and can be relaxed at the expense of more complicated proofs. Assumption A2-(i) is somehow unusual since no continuity or smoothness are necessary for $m(x) \equiv E[Y|X = x]$, but differentiability of $f(x)$ is assumed. Assumption A2-(ii) ensures $\sqrt{n}$-consistency. Assumption A3 is necessary in pointwise asymptotic theory for k-NN estimators, since it guarantees that both bias and variance of k-NN converge to zero as sample size increases.

Theorem 1 *Under assumption A1–A3,*

$$E[\theta_n] = \theta_0 \{1 + \mathrm{o}_p(1)\},$$
$$\mathrm{var}\left(\sqrt{n}\theta_n\right) = \mathbb{E}[\mathrm{var}(Y|X)/f^2(X)]\{1 + \mathrm{o}_p(1)\}$$

as $n \to \infty$.

Some remarks are in order:

Remark 1 *Theorem 1 ensures the L^2-convergence of 1.2. Furthermore, it also implies that* $\mathrm{var}(\sqrt{n}\theta_n) \to \mathrm{var}(\sqrt{n}\widetilde{\theta}_n)$ *as $n \to \infty$.*

Remark 2 *We have been unable to establish limiting distribution theory for θ_n, since the statistical dependence among $\{Y_i R_n(X_i, k)\}_{i=1}^n$ is of a form that is not covered by standard central limit theory for dependent processes. However, we conjecture that results in Bickel and Breiman (1983) can be extended to establish asymptotic normality of statistics of this form.*

Remark 3 *From the computational point of view, if T_0 is the total number of operations necessary to sort an n-dimensional array, θ_n would require at least nT_0 such operations for its calculation.*

Remark 4 *Unlike the ordered estimator discussed in the previous section, θ_n can easily be adapted to handle vector-valued X's.*

An important issue for both estimators is how k can be chosen in a given application. Techniques presented in Jacho-Chávez (2007) can potentially answer this question, and remains a topic for future research.

3 Monte Carlo Results

We consider the data generating process used in Lewbel and Schennach (2007). We draw x_i, ε_i as independent standard normals and construct $y_i = 2x_i(1+\varepsilon_i)\mathbb{I}(0 < x_i < 1)$, $i = 1, 2, \ldots, n$. We then estimate $\theta = E[y/f(x)]$ by computing 1.4 and 1.5 using each of the 10000 constructed samples $\{y_i, x_i\}_{i=1}^n$ with $n = 200$, 400, and 600. Figure 1 shows the results. The bias, variance and Mean Squared Error (MSE) of estimators 1.4 (black line) and 1.5 (gray line) are presented for different values of $k = 1, \ldots, 40$. Their performance are comparable in terms of MSE and variances at different values of k. This observation reinforces the notion that k plays different roles in the construction of both estimators. As predicted by theorem 1, both estimators have comparable variances, but θ_n seems to be unbiased for this design, explaining why large values of k seems to improve its precision. As expected, the MSE decreases for both estimators as sample size increases.

Appendix

Let $\|u\|$ denote the Euclidean norm of the vector u. Set $S_r = \{v : \|v - x\| < r\}$, a ball centered at x with radius r. $G(r) = \Pr\{X_i \in S_r\}$ is defined accordingly.

Lemma 1 *Let $h(r) = [r^\lambda G^\gamma(r)]^{-1}$, λ and γ are integers such that $E[h(R_i)] < \infty$, then*

$$\mathbb{E}[h(R_n(X_i, k))|X_i] = (2f(X_i))^\lambda \left(\frac{k}{n}\right)^{-\lambda-\gamma} \{1 + o_p(1)\}.$$

Proof. This corresponds to Lemma A.1 of Ouyang et al. (2006) with $q = 1$. See Liu and Lu (1997) for a detailed proof. ■

Proof of Theorem 1:

Let $\epsilon \equiv Y - m(x)$, then, by construction $\mathbb{E}[\epsilon|X = x] = 0$, and

$$\begin{aligned}\theta_n &= \frac{2}{k}\sum_{i=1}^{n} m(X_i) R_n(X_i, k) + \frac{2}{k}\sum_{i=1}^{n} \epsilon_i R_n(X_i, k) \\ &\equiv \mathcal{T}_{1n} + \mathcal{T}_{2n},\end{aligned}$$

where we have used the notation $\epsilon_i = Y_i - m(X_i)$, and the definition of $\mathcal{T}_{ln}$ $(l = 1, 2)$ should be apparent. Firstly, notice that $\mathbb{E}[\mathcal{T}_{2n}] = 0$ by the law of iterated expectations. It then follows from assumption A1 (i) that

$$\begin{aligned}\mathbb{E}[\mathcal{T}_{1n}] &= \frac{2}{k}\sum_{i=1}^{n} \mathbb{E}\left[m(X_i) R_n(X_i, k)\right] = 2\frac{n}{k}\mathbb{E}[m(X) E[R_n(X, k)|X]] \\ &= \frac{2n}{k}\mathbb{E}\left[\frac{m(X)}{2f(X)}\left(\frac{k}{n}\right)\right]\{1 + o_p(1)\} \qquad \text{(A-1)} \\ &= \mathbb{E}\left[\frac{E[Y|X]}{f(X)}\right]\{1 + o_p(1)\} = \theta_0 + o_p(\theta),\end{aligned}$$

where A-1 follows from Lemma 1 with $\lambda = -1$ and $\gamma = 0$. Similarly, by the tower property of conditional expectations (see Billingsley (1986, Theorem 34.3)), it follows that

$$\begin{aligned}\mathbb{E}[\mathcal{T}_{2n}^2] &= \frac{4}{k^2}\sum_{i=1}^{n} \mathbb{E}\left[\epsilon_i^2 R_n^2(X_i, k)\right] = \frac{4n}{k^2}\mathbb{E}\left[\mathrm{var}(Y|X) E[R_n^2(X, k)|X]\right] \\ &= \frac{4n}{k^2}\mathbb{E}\left[\frac{\mathrm{var}(Y|X)}{4f^2(X)}\left(\frac{k}{n}\right)^2\right]\{1 + o_p(1)\} \qquad \text{(A-2)} \\ &= n^{-1}\mathbb{E}\left[\frac{\mathrm{var}(Y|X)}{f^2(X)}\right]\{1 + o_p(1)\} = n^{-1}\sigma^2 + o_p\left(n^{-1}\sigma^2\right),\end{aligned}$$

where A-2 follows from Lemma 1 with $\lambda = -2$ and $\gamma = 0$.

References

Bickel, Peter J., and Leo Breiman, 1983, Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test, The Annals of Probability 11(1), 185–214.

Billingsley, Patrick, 1986, Probability and Measure Wiley series in probability and mathematical statistics, 2 ed. (Wiley, New York, NY).

Hayfield, Tristen, and Jeffrey S. Racine, 2006 *np: Nonparametric kernel smoothing methods for mixed datatypes.* R package version 0.12-1.

Jacho-Chávez, David T., 2007, Optimal bandwidth choice for estimation of inverse conditionaldensity-weighted expectations, Unpublished Manuscript.

Khan, Shakeeb, and Arthur Lewbel, 2007, Weighted and two-stage least squares estimation of semiparametric truncated regression models, Econometric Theory 23(2), 309–347.

Lewbel, Arthur, 1998, Semiparametric latent variable model estimation with endogenous or mismeasured regressors, Econometrica 66(1), 105–121.

_ 2000, Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables, Journal of Econometrics 97(1), 145–177.

_ 2006, Endogenous selection or treatment model estimation, Forthcoming in Journal of Econometrics.

Lewbel, Arthur, and Susanne M. Schennach, 2007, A simple ordered data estimator for inverse density weighted expectations, Journal of Econometrics 127(1), 189–211.

Liu, Zhenjuan, and Xuewen Lu, 1997, Root-n-consistent semiparametric estimation of partially linear models based on k-nn method, Econometric Reviews 16(4), 411–420.

Ouyang, Desheng, Dong Li, and Qi Li, 2006, Cross-validation and non-parametric k nearest–neighbour estimation, Econometrics Journal 9, 448–471.

Figure 1: Monte Carlo results