Econometric Theory, **33**, 2017, 1259–1263.
doi:10.1017/S026646661600030X

KERNEL ESTIMATION WHEN DENSITY MAY NOT EXIST: A CORRIGENDUM

VICTORIA ZINDE-WALSH
McGill University and CIREQ

The paper "Kernel estimation when density may not exist" (Zinde-Walsh, 2008) considered density as a generalized function given by a functional on a space of smooth functions; this made it possible to establish the limit properties of the kernel estimator without assuming the existence of the density function. This note corrects an error in that paper in the derivation of the variance of the kernel estimator. The corrected result is that in the space of generalized functions the parametric rate of convergence of the kernel density estimator to the limit Gaussian process is achievable.

1. INTRODUCTION

Mass points are well documented in wage distributions (Fortin, Lemieux, and Firpo, 2011; Brochu, Green, Lemieux, and Townsend, 2015). Spikes in density or mass points in the distribution occur in financial data; for example, the density function for stock returns is shown to diverge at the median (Han, Cho, and Phillips, 2011). There are numerous examples of singularities and spikes in distributions studied in the physical sciences. In such cases the usual limit theory for nonparametric kernel estimation does not apply. Nevertheless nonparametric kernel estimators may still provide useful information and certain linear random functionals will deliver asymptotic results suitable for conducting inference. Classes of functionals, which are known as random generalized functions, were employed for this purpose by Zinde-Walsh (2008) and are briefly introduced here in Section 2. The paper by Zinde-Walsh (2008) provided the result that the bias of the kernel estimator functional is determined solely by the kernel order. Unfortunately, an error in the derivation of variance gave an incorrect rate. Here in Section 3 the corrected derivation of the variance is given. The corrected result demonstrates that the convergence rate to the limit Gaussian process (with undersmoothing) is parametric.

The support of the Social Sciences and Humanities Research Council of Canada (SSHRC) and the *Fonds québecois de la recherche sur la société et la culture* (FRQSC) is gratefully acknowledged. I thank John Galbraith, Yanqin Fan, the Editor and referees for helpful advice. Address correspondence to Victoria Zinde-Walsh, Department of Economics, McGill University, 855 Sherbrooke Street West, Montreal, QC H3A 2T7, Canada; e-mail: victoria.zinde-walsh@mcgill.ca.

© Cambridge University Press 2016

https://doi.org/10.1017/S026646661600030X Published online by Cambridge University Press

2. THE GENERALIZED FUNCTIONS SET-UP FOR THE KERNEL ESTIMATOR

We start by defining a generic linear topological space of well-behaved functions on R^k. Suppose that a function ψ has bounded support and is m times continuously differentiable on R^k. Then define D_m as a linear space of such functions; $m \geq 0$ could be finite or infinite. The topology is defined by convergence for sequences of functions $\psi_n \in D_m$: $\psi_n \to \psi$ in D_m if the supports are contained in the same bounded set and $\lim_{n\to\infty} \sup \left|\psi_n^{(l)}(x) - \psi^{(l)}(x)\right| = 0$, $l = 0, \ldots, m$; thus uniform convergence is required for the function sequence ($l = 0$) and for sequences of derivatives of all orders. Any such space is denoted D.

Linear continuous functionals on D are called generalized functions; the space of generalized functions is thus the dual space D^*. The restrictive conditions and convergence requirements on the space D give rise to more liberal convergence in the dual space: sequences of functions that are divergent in the usual norms of function spaces, such as uniform or L_p may converge as elements of D^*. The classical references are the books by Gelfand and Shilov (1964a,b, vols. 1,2); the necessary results are summarized in Appendix B of Zinde-Walsh (2008).

For any $\psi \in D$ denote the value of the functional $f \in D^*$ by (f, ψ). For any ordinary locally integrable function $f(x)$ on R^k, the value of the functional is defined as $(f, \psi) = \int f(x)\psi(x)dx$ with the integral to be taken over R^k. The bounded support of the continuous ψ and the local integrability of f ensures the existence of the integral. For the distribution function F on R^k, then, $(F, \psi) = \int F(x)\psi(x)dx$; since the distribution function is locally integrable this integral exists for any $\psi \in D$.

Define the operator ∂ which for any suitably differentiable function g provides the function ∂g which has point-wise values $\partial g(x) = \frac{\partial^k g(x_1,\ldots,x_k)}{\partial x_1 \ldots \partial x_k}$. Suppose that $D \subseteq D_k$, then for any $\psi \in D$, the function $\partial\psi$ is a continuous function with bounded support. If the distribution F were absolutely continuous, then the density function would be given by $f = \partial F$. But even if F is not absolutely continuous, $f = \partial F$ can be defined as a generalized function on $D \subseteq D_k$ as follows. The value of the functional $f = \partial F$ applied to $\psi \in D$ is defined as $(f, \psi) = (-1)^k (F, \partial\psi)$ and computed as $(-1)^k \int \ldots \int F(x_1, \ldots, x_k) \frac{\partial^k \psi(x_1,\ldots,x_k)}{\partial x_1 \ldots \partial x_k} dx_1 \ldots dx_k$. The integral always exists and defines the generalized density function by providing the values of the functional (f, ψ). Integration by parts establishes the fact that this definition is consistent with the functional given by the density when it exists as an ordinary integrable function. Moreover, the differentiation operator ∂ is continuous in D^*.

The kernel estimator of the density is defined next. The assumptions on the kernel and bandwidth are standard.

Assumption. (a) $K_s(w)$ is a bounded function which is an l-th order kernel with $l \geq 2$; (b) The bandwidth (vector) h has components $h_1, \ldots, h_k$ dependent on n; $\bar{h} = \max\{h_1, \ldots, h_k\}$ converges to zero as $n \to \infty$; (c) For $x_i = (x_{i1}, \ldots, x_{ik})$ and $x = (x_1, \ldots, x_k)$ the function $K(\frac{x_i - x}{h})$ is a multivariate product kernel

https://doi.org/10.1017/S026646661600030X Published online by Cambridge University Press

function $K\left(\frac{x_i-x}{h}\right) = K_1\left(\frac{x_{i1}-x_1}{h_1}\right)\ldots K_k\left(\frac{x_{ik}-x_k}{h_k}\right)$ with each $K_s(w)$ satisfying Assumption A(a) and h satisfying Assumption A(b).

Assume further that $x_i, i = 1,\ldots,n$, is a random sample from the distribution F on R^k. The kernel density estimator is (as, e.g., in Li and Racine, 2007) $\hat{f}(x) = \frac{1}{n\Pi_{s=1}^k h_s}\sum_{i=1}^n K(\frac{x_i-x}{h})$. The kernel estimator is defined as a point-wise function but it can be also considered as a generalized function $\hat{f}$ on D with values of the functional computed for every $\psi \in D$ as

$$\left(\hat{f},\psi\right) = \int \hat{f}(x)\psi(x)dx. \tag{1}$$

The values of this linear continuous functional are random variables; thus $\hat{f}$ defines a random functional.

Random generalized functions are random linear continuous functionals on the specific spaces D; they were introduced by Gelfand and Vilenkin (1964); Appendix B of Zinde-Walsh (2008) summarizes some of the results from that book. In particular, a generalized function Gaussian process W arises if for the random functional W all the random vectors of values, $((W,\psi_1),\ldots,(W,\psi_v))'$ are Gaussian for any finite set of functions $\psi_1,\ldots,\psi_v \in D$ and the mean functional and the bilinear covariance functional exist. The mean functional and the covariance functional fully define the generalized function Gaussian process.

3. LIMIT PROCESS FOR THE KERNEL DENSITY ESTIMATOR

For smooth random functionals such as the integral in (1), a parametric rate of convergence is common under appropriate undersmoothing (see, e.g., Goldstein and Messer, 1992). Thus two points have to be established: (i) undersmoothing is achievable; (ii) the parametric rate is obtained in this case. Point (i) was proved in Theorem 2 in Zinde-Walsh (2008) which established the convergence rate $O(\bar{h}^l)$ for the generalized bias functional of the kernel estimator on $D \subset D_{k+l}$. Differentiability up to order k of the functions in D is required as shown above for the density to be defined as a generalized function; further differentiability is required to exploit the order of the kernel in the expansion of the bias functional. The Theorem below provides the parametric convergence rate to the limit Gaussian process; it corrects the error in Zinde-Walsh (2008).[1]

THEOREM. *For a kernel function K satisfying the Assumption, if $\overline{h} \to 0$ as $n \to \infty$ with $\overline{h}^{2l} n \to 0$, the sequence of generalized random processes $n^{\frac{1}{2}}\left(\hat{f}-f\right)$ converges to a generalized Gaussian process with mean functional zero and covariance functional C which for any $\psi_1, \psi_2 \in D_{k+l}$ yields*

$$(C,(\psi_1,\psi_2)) = cov\,(\psi_1,\psi_2)\,. \tag{2}$$

Proof. Denote $\frac{1}{\Pi_{s=1}^k h_s}$ by $\frac{1}{\Pi h}$. Define a generalized function e_{nhi} as a random functional with values $(e_{nhi},\psi) = \int \frac{1}{\Pi h} K\left(\frac{x_i-x}{h}\right)\psi(x)dx - (f,\psi)$; then

https://doi.org/10.1017/S026646661600030X Published online by Cambridge University Press

$\frac{1}{n}\Sigma_{i=1}^{n} e_{nhi} = \hat{f} - f$. By Theorem 2 in Zinde-Walsh (2008), $E(e_{nhi}) = O\left(\bar{h}^{l}\right)$ as $\bar{h} \to 0$. Then as $\bar{h} \to 0$

$$E((e_{nhi}, \psi_1)(e_{nhj}, \psi_2)) = \begin{cases} E\left(T_{ii}^{1} + T_{ii}^{2}\right); & i = j \\ O\left(\bar{h}^{2l}\right); & i \neq j \end{cases},$$

where $T_{ii}^{1} = \left(\int \frac{1}{\Pi h} K\left(\frac{x_i - x}{h}\right) \psi_1(x) dx\right)\left(\int \frac{1}{\Pi h} K\left(\frac{x_i - x}{h}\right) \psi_2(x) dx\right)$ and

$$T_{ii}^{2} = -\left(\int \frac{1}{\Pi h} K\left(\frac{x_i - x}{h}\right) \psi_1(x) dx\right) E\psi_2 - \left(\int \frac{1}{\Pi h} K\left(\frac{x_i - x}{h}\right) \psi_2(x) dx\right) E\psi_1 + E\psi_1 E\psi_2.$$

Consider T_{ii}^{1}. Changing the variable x to hw we obtain

$$T_{ii}^{1} = \int K(w)\, \psi_1(x_i - hw)\, dw \int K(w)\, \psi_2(x_i - hw)\, dw.$$

By Taylor expansion for $s = 1$ and 2 we obtain

$$\psi_s(x_i - hw) = \psi_s(x_i) + \sum_{\Sigma t_v = 1}^{l} \frac{1}{t_1! \ldots t_k!} \frac{\partial^{\Sigma t_v} \psi_s}{\partial x_1^{t_1} \ldots \partial_k^{t_k}}(x_i) \Pi(-w_v h_v)^{t_v} + \bar{h}^{l} R(\psi_s, x_i - hw, x_i),$$

where $\bar{h}^{l} R(\psi_s, x_i - hw, x_i)$ represents the remainder term. By continuity of the derivatives of ψ and bounded support of ψ, the terms in the expansion are all uniformly bounded, and as $\bar{h} \to 0$ for any ε there exists an $\bar{h}$, which is small enough that $|R(\psi_s, x_i - hw, x_i)| < \varepsilon$ for any $x_i \in R^k$. Substituting this expression into T_{ii}^{1} and using kernel order we find that $T_{ii}^{1} = \psi_1\psi_2 + \bar{h}^{l} R_i$ with uniformly bounded R_i. Then $E\left(T_{ii}^{1}\right) = E(\psi_1\psi_2) + O\left(\bar{h}^{l}\right)$. (There was an error in this expression in Zinde-Walsh, 2008). Similarly, for T_{ii}^{2} the expectation is $E(T_{ii}^{2}) = -E\psi_1 E\psi_2 + O\left(\bar{h}^{l}\right)$. This provides the limit of $E\left(T_{ii}^{1} + T_{ii}^{2}\right)$ as $E(\psi_1\psi_2) - E\psi_1 E\psi_2 = cov(\psi_1, \psi_2)$.

Define $\eta_{nhi} = n^{\frac{1}{2}}(e_{nhi} - Ee_{nhi})$ and consider $\eta_{nh} = \frac{1}{n}\Sigma \eta_{nhi}$. The generalized random function η_{nh} has expectation functional zero, in the covariance the terms with $i \neq j$ vanish and $E(\eta_{nh}, \psi_1)(\eta_{nh}, \psi_2)$ converges to $E(\psi_1\psi_2) - E\psi_1 E\psi_2$. By the usual arguments (similar to those in the proof of Theorem 3 in Zinde-Walsh, 2008) by appropriately standardizing the vector with components $(\eta_{nh}, \psi_s), s = 1, 2$ and applying the Cramer-Wold device the Lyapunov central limit theorem provides convergence to a limit generalized Gaussian process for the random functional $\sqrt{n}\left(\hat{f} - f\right)$ on D. ■

NOTE

1. That paper also considered the numerator of the Nadaraya-Watson estimator as a generalized random function. A similar correction of the error in the derivation of variance establishes a parametric rate. A complete treatment of the kernel estimator of conditional mean (not just the numerator) as a generalized random function is a subject of ongoing research.

https://doi.org/10.1017/S026646661600030X Published online by Cambridge University Press

REFERENCES

Brochu, P., D.A. Green, T. Lemieux, & J. Townsend (2015) The Minimum Wage, Turnover, and the Shape of the Wage Distribution. Working paper, University of British Columbia.

Fortin, N., T. Lemieux, & S. Firpo (2011) Decomposition methods in economics. In O. Ashenfelter & D. Card (eds.), *Handbook of Labor Economics*, vol. 4A, pp. 1–102. North-Holland.

Gelfand, I.M. & G.E. Shilov (1964a) *Generalized Functions, Volume 1, Properties and Operations*. Academic Press.

Gelfand, I.M. & G.E. Shilov (1964b) *Generalized Functions, Volume 2, Spaces of Test Functions and Generalized Functions*. Academic Press.

Gelfand, I.M. & N. Ya. Vilenkin (1964) *Generalized Functions, Volume 4, Applications of Harmonic Analysis*. Academic Press.

Goldstein, L. & K. Messer (1992) Optimal plug-in estimators for nonparametric functional estimation. *The Annals of Statistics* 20, 1306–1328.

Han, C., J.S. Cho, & P.C.B. Phillips (2011) Infinite density at the median and the typical shape of stock return distributions. *Journal of Business and Economic Statistics* 29, 282–294.

Zinde-Walsh, V. (2008) Kernel estimation when density may not exist. *Econometric Theory* 24, 696–725.

https://doi.org/10.1017/S026646661600030X Published online by Cambridge University Press