Ann. Inst. Statist. Math.
Vol. 57, No. 3, 425–442 (2005)

CENTRAL LIMIT THEOREM FOR ASYMMETRIC KERNEL FUNCTIONALS

MARCELO FERNANDES[1] AND PAULO KLINGER MONTEIRO[2]

[1] *Economics Department, Queen Mary, University of London, Mile End Road, London, E1 4NS, U.K.*, e-mail: m.fernandes@qmul.ac.uk
[2] *Graduate School of Economics, Getulio Vargas Foundation, Praia de Botafogo, 190, 22250-900 Rio de Janeiro, RJ, Brazil*, e-mail: pklm@fgv.br

(Received October 23, 2000; revised May 28, 2004)

Abstract. Asymmetric kernels are quite useful for the estimation of density functions with bounded support. Gamma kernels are designed to handle density functions whose supports are bounded from one end only, whereas beta kernels are particularly convenient for the estimation of density functions with compact support. These asymmetric kernels are nonnegative and free of boundary bias. Moreover, their shape varies according to the location of the data point, thus also changing the amount of smoothing. This paper applies the central limit theorem for degenerate U-statistics to compute the limiting distribution of a class of asymmetric kernel functionals.

Key words and phrases: Asymmetric kernel, beta kernel, boundary bias, central limit theorem, density estimation, gamma kernel, U-statistic theory.

1. Introduction

Fixed kernels are not appropriate to estimate density functions whose supports are bounded in view that they engender boundary bias due to the allocation of weight outside the support in the event that smoothing is applied near the boundary. A proper asymmetric kernel never assigns weight outside the density support and therefore should produce better estimates of the density near the boundary. Chen (1999, 2000) shows indeed that replacing fixed with asymmetric kernels substantially increases the precision of density estimation close to the boundary. In particular, beta kernels are particularly appropriate to estimate densities with compact support, whereas gamma kernels are more convenient to handle density functions whose supports are bounded from one end only. These asymmetric kernels are nonnegative and free of boundary bias. Moreover, their shape varies according to the location of the data point, thus also changing the amount of smoothing.

This paper derives the asymptotic behavior of asymmetric kernel functionals by applying a central limit theorem for degenerate U-statistics with variable kernel. The motivation is simple. It is often the case that one must derive the limiting distribution of density functionals such as

$$I_n = \int_A \varphi(x)[\hat{f}(x) - f(x)]^2 dx, \tag{1.1}$$

where $\varphi(\cdot)$ is a bounded regular function and $\hat{f}$ is an asymmetric kernel estimate of the true density f with support A. Examples abound in econometrics and statistics. Indeed, a central limit theorem for the density functional (1.1) is useful to study the order of closeness between the integrated square error and the mean integrated squared error in the ambit of nonparametric kernel estimation of densities with bounded support. Although there are sharp results for nonparametric density estimation based on fixed kernels, e.g. Bickel and Rosenblatt (1973) and Hall (1984), no results are available for asymmetric kernel density estimation.

Further, goodness-of-fit test statistics are usually driven by second-order asymptotics (e.g., Aït-Sahalia *et al.* (2001)), hence density functionals such as (1.1) arise very naturally in that context. Consider, for instance, one of the goodness-of-fit tests advanced by Fernandes and Grammig (2005) for duration models, which gauges how large is

$$\Lambda(f,\theta) = \int_0^\infty [\Gamma_\theta(x) - \Gamma_f(x)]^2 f(x)dx, \tag{1.2}$$

where $\Gamma_\theta(\cdot)$ and $\Gamma_f(\cdot)$ denote the parametric and nonparametric hazard rate functions, respectively. It follows from the functional delta method that the asymptotic behavior of (1.2) is driven by the leading term of the second functional derivative, namely

$$\int_0^\infty \frac{\Gamma_f(x)}{1-F(x)}[\hat{f}(x) - f(x)]^2 dx.$$

As duration data are nonnegative by definition, gamma kernels are called for so as to avoid boundary bias in the density estimation.

Let $X_1, \ldots, X_n$ be a random sample from an unknown probability density function f defined on a support A, which is either $A = [0,\infty)$ or $A = [0,1]$. The nonparametric estimator $\hat{f}$ of the density function f uses the appropriate asymmetric kernel, namely, the gamma kernel for $A = [0,\infty)$ and the beta kernel for $A = [0,1]$. As in any kernel density estimation, the smoothing bandwidth, say b, converges to zero as the sample size grows. We are now ready to formulate the main result.

THEOREM 1.1. *Suppose that φ is a bounded regular function and the density function f and its first and second derivatives are bounded and square integrable on A. Assuming further that b is of order $o(n^{-4/9})$, it then ensues that*

$$nb^{1/4}I_n - \frac{b^{-1/4}}{2\sqrt{\pi}}\mathbb{E}_{X_1}[c_A^{-1/2}(X_1)\varphi(X_1)] \xrightarrow{d} N(0,\sigma_A^2), \tag{1.3}$$

where the subscript of the expectation operator $\mathbb{E}$ indicates the random quantity that it refers to, $\sigma_A^2 = \frac{1}{\sqrt{2\pi}}\mathbb{E}_{X_1}[c_A^{-1/2}(X_1)\varphi^2(X_1)f(X_1)]$, and the boundary correction depends on the support A with $c_{[0,\infty)}(X_1) = X_1$ and $c_{[0,1]}(X_1) = X_1(1-X_1)$.

As is apparent, the boundary correction $c_A(\cdot)$ is the sole distinction between the limiting distributions of the two asymmetric kernel functionals. Because the support is bounded only from below in the gamma kernel case, it suffices to control for values of x close to the origin. In the context of beta kernels, it is necessary to deal with x in the vicinity of both boundaries of the unit interval. We defer until the next section, which reviews the properties of beta and gamma kernels, to comment upon the bandwidth condition $b = o(n^{-4/9})$. Lastly, we split the proof of the theorem into two parts: Sections 3 and 4 demonstrate the result for the gamma and beta kernel functionals, respectively.

2. Asymmetric kernels

Instead of the usual nonparametric kernel density estimator

$$\tilde{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

where K is a fixed kernel function and h is a smoothing bandwidth, consider the asymmetric kernel estimator

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_A(X_i),$$

where $K_A(\cdot)$ corresponds either to the gamma kernel

$$K_{x/b+1,b}(u) = \frac{u^{x/b}\exp(-u/b)}{\Gamma(x/b+1)b^{x/b+1}}\mathbb{1}(u \geq 0)$$

or to the beta kernel

$$K_{x/b+1,(1-x)/b+1}(u) = \frac{u^{x/b}(1-u)^{(1-x)/b}}{B(x/b+1,(1-x)/b+1)}\mathbb{1}(0 \leq u \leq 1)$$

according to the density support A. As usual, the smoothing bandwidth b converges to zero as the sample size n grows.

Asymmetric kernel estimators are boundary bias free in that the bias is of order $O(b)$ both near the boundaries and in the interior of the support (Chen (1999, 2000)). The absence of boundary bias is due to the fact that asymmetric kernels have the same support of the underlying density, and hence no weight is assigned outside the density support. The trick is that asymmetric kernels are flexible enough to vary their shape and the amount of smoothing according to the location within the support.

On the other hand, the asymptotic variance of asymmetric kernels is of higher order $O(n^{-1}b^{-1})$ near the boundaries than in the interior, which is of order $O(n^{-1}b^{-1/2})$. Nonetheless, the impact on the integrated variance is negligible, so that it does not affect the mean integrated square error. Furthermore, the optimal bandwidth $b_* = O(n^{-2/5}) = O(h_*^2)$, where h_* is the optimal bandwidth for fixed nonnegative kernel estimators (Chen (1999, 2000)). Accordingly, both beta and gamma kernel density estimators achieve the optimal rate of convergence for the mean integrated squared error of nonnegative kernels.

It is readily seen that the bandwidth condition $b = o(n^{-4/9})$ induces undersmoothing in the density estimation. Other limiting conditions on the bandwidth are also applicable, but they would result in different terms for the bias in (1.3). For instance, the bandwidth condition $b = O(n^{-2/5})$ that entails asymptotically optimal pointwise density estimates yields an additional term driving the asymptotic distribution of (1.1). Accordingly, that would lead to another component in the variance whose estimation would require the estimate of the second-order derivative of the density function as in Härdle and Mammen (1993), for example. As an alternative, Härdle and Mammen (1993) and Chen *et al.* (2003) show that one may replace $f(x)$ by $\mathbb{E}\hat{f}(x)$ in (1.1) so as to avoid undersmoothing.

3. Gamma kernel functionals

In this section, we show that the asymptotic behavior of gamma kernel functionals of the form (1.1) is indeed as claimed in (1.3). The proof builds on U-statistic theory in that we decompose the functional so as to force the emergence of a degenerate U-statistic. Let $r_n(x, X_1) = \varphi^{1/2}(x) K_{x/b+1,b}(X_1)$ and $\breve{r}_n(x, X_1) = r_n(x, X_1) - \mathbb{E}_{X_1}[r_n(x, X_1)]$. In addition, denote by $\int$ the integral over the density support $A = [0, \infty)$ and by $\oint$ multiple integrals. Letting the absence of subscript in the expectation operator denote the expectation over $(X_1, \ldots, X_n)$ then yields

$$\begin{aligned} I_n &= \int \varphi(x)[\hat{f}(x) - \mathbb{E}\hat{f}(x)]^2 dx + \int \varphi(x)[\mathbb{E}\hat{f}(x) - f(x)]^2 dx \\ &\quad + 2\int \varphi(x)[\hat{f}(x) - \mathbb{E}\hat{f}(x)][\mathbb{E}\hat{f}(x) - f(x)] dx, \end{aligned}$$

or equivalently, $I_n = I_{1n} + I_{2n} + I_{3n} + I_{4n}$, where

$$\begin{aligned} I_{1n} &= \frac{2}{n^2} \sum_{i<j} \int \breve{r}_n(x, X_i) \breve{r}_n(x, X_j) dx \\ I_{2n} &= \frac{1}{n^2} \sum_{i} \int \breve{r}_n^2(x, X_i) dx \\ I_{3n} &= \int \varphi(x)[\mathbb{E}\hat{f}(x) - f(x)]^2 dx \\ I_{4n} &= 2\int \varphi(x)[\hat{f}(x) - \mathbb{E}\hat{f}(x)][\mathbb{E}\hat{f}(x) - f(x)] dx. \end{aligned}$$

The first term stands for a degenerate U-statistic and contributes to the asymptotic variance as well as to the limiting Gaussian distribution, whereas the second term defines the asymptotic mean. In turn, the third and the fourth terms are negligible provided that the bandwidth b is of order $o(n^{-4/9})$.

We start by deriving the first and second moments of $r_n(x, X_1)$. Observe that

$$\mathbb{E}_{X_1}[r_n(x, X_1)] = \varphi^{1/2}(x) \int K_{x/b+1,b}(x_1) f(x_1) dx_1 = \varphi^{1/2}(x) \mathbb{E}_\zeta[f(\zeta)],$$

where ζ has a gamma distribution $\mathcal{G}(x/b + 1, b)$. Applying a Taylor expansion yields

$$\mathbb{E}_\zeta[f(\zeta)] = f(x) + O(b),$$

illustrating the fact that the gamma kernel density estimation has a uniform bias of order $O(b)$ as singled out by Chen (2000). Put differently, the order of magnitude of the bias does not depend on the location within the density support and, as a consequence, $\mathbb{E}_{X_1}[r_n(x, X_1)] = \varphi^{1/2}(x) f(x) + O(b)$.

We compute the second moment of $r_n(x, X_1)$ in a similar fashion. It ensues from the properties of the gamma density that

$$\begin{aligned} \mathbb{E}_{X_1}[r_n^2(x, X_1)] &= \varphi(x) \int K^2_{x/b+1,b}(x_1) f(x_1) dx_1 \\ &= \varphi(x) B_b(x) \mathbb{E}_\eta[f(\eta)], \end{aligned}$$

where $\eta \sim \mathcal{G}(2x/b + 1, b/2)$ and

$$B_b(x) = \frac{\Gamma(2x/b + 1)/b}{2^{2x/b+1} \Gamma^2(x/b + 1)}. \tag{3.1}$$

Given that $\mathbb{E}_\eta[f(\eta)] = f(x) + O(b)$, it then follows that

$$\begin{aligned}\mathbb{E}(I_{2n}) &= \frac{1}{n}\int \mathbb{E}_{X_1}[r_n^2(x, X_1)]dx - \frac{1}{n}\int \mathbb{E}^2_{X_1}[r_n(x, X_1)]dx \\ &= \frac{1}{n}\int \varphi(x)B_b(x)[f(x) + O(b)]dx + O(n^{-1}) \\ &= \frac{1}{n}\int \varphi(x)B_b(x)f(x)dx + O(n^{-1}).\end{aligned}$$

Let $R(z) \equiv \sqrt{2\pi}\exp(-z)z^{z+1/2}/\Gamma(z+1)$. By rewriting (3.1) in terms of the R-function, one obtains

$$B_b(x) = \frac{R^2(x/b)}{R(2x/b)}\frac{b^{-1/2}x^{-1/2}}{2\sqrt{\pi}}. \tag{3.2}$$

According to Brown and Chen's (1999) Lemma 3, $R(z)$ is a monotonic increasing function that converges to one as $z \to \infty$ and $R(z) \leq 1$ for any $z > 0$. Moreover, if x/b is large enough, the difference between one and the first fraction in (3.2) is negligible and hence one may consider that $\sqrt{bx}B_b(x) = 1/(2\sqrt{\pi})$ (Chen (2000)). It therefore follows that

$$\begin{aligned}\mathbb{E}(I_{2n}) &= \frac{1}{n}\int \varphi(x)B_b(x)f(x)dx + O(n^{-1}) \\ &= \frac{1}{n}\int_0^b \varphi(x)B_b(x)f(x)dx + \frac{1}{n}\int_b^\infty \varphi(x)B_b(x)f(x)dx + O(n^{-1}).\end{aligned}$$

It then suffices to write the first integral in terms of $\omega = x/b$ to show that the first term is at most of order $O(1/n)$, so that

$$\mathbb{E}(I_{2n}) = \frac{b^{-1/2}}{n}\frac{1}{2\sqrt{\pi}}\int_0^\infty x^{-1/2}\varphi(x)f(x)dx + O(n^{-1})$$

by the dominated convergence theorem, provided that $\mathbb{E}[x^{-1/2}\varphi(x)]$ is finite. It therefore ensues that

$$nb^{1/4}\mathbb{E}(I_{2n}) - \frac{b^{-1/4}}{2\sqrt{\pi}}\mathbb{E}_{X_1}[X_1^{-1/2}\varphi(X_1)] = o(1).$$

In addition, we show in the Appendix that

$$V(I_{2n}) = O(n^{-3}b^{-2} + n^{-3}b^{-1}). \tag{3.3}$$

It then follows that $V(nb^{1/4}I_{2n}) = n^2b^{1/2}V(I_{2n}) = O(n^{-1}b^{-3/2})$, which is by assumption of order $o(1)$. Applying Chebyshev's inequality thus results in

$$nb^{1/4}I_{2n} - \frac{b^{-1/4}}{2\sqrt{\pi}}\mathbb{E}_{X_1}[X_1^{-1/2}\varphi(X_1)] = o_p(1).$$

The fact that $b = o(n^{-4/9})$ also ensures that the third and fourth terms are negligible if properly normalized. Indeed, I_{3n} is proportional to the integrated squared bias of the gamma kernel density estimation, hence it is of order $O(b^2)$. The bandwidth condition then guarantees that $nb^{1/4}I_{3n} = O(nb^{9/4}) = o(1)$. Further,

$$\mathbb{E}(I_{4n}) = 2\int \varphi(x)\mathbb{E}[\hat{f}(x) - \mathbb{E}\hat{f}(x)][\mathbb{E}\hat{f}(x) - f(x)]dx = 0,$$

whereas $\mathbb{E}(I_{4n}^2) = O(n^{-1}b^2)$ analogously to Hall ((1984), Lemma 1). This means that $\mathbb{E}(n^2 b^{1/2} I_{4n}^2) = O(nb^{5/2}) = o(n^{-1/10})$, and thus it stems from Chebyshev's inequality that $nb^{1/4}I_{4n} = o_p(1)$.

Finally, recall that $I_{1n} = \sum_{i<j} H_n(X_i, X_j)$, where

$$H_n(X_i, X_j) = \frac{2}{n^2} \int \breve{r}_n(x, X_i)\breve{r}_n(x, X_j)dx.$$

As is apparent, I_{1n} is a degenerate U-statistic as $H_n(X_i, X_j)$ is symmetric, centered, and $\mathbb{E}[H_n(X_i, X_j) \mid X_j] = 0$ almost surely. To see why, note that

$$\begin{aligned}\mathbb{E}[H_n(X_i, X_j) \mid X_j] &= \frac{2}{n^2} \int \breve{r}_n(x, X_j)\mathbb{E}[\breve{r}_n(x, X_i) \mid X_j]dx \\ &= \frac{2}{n^2} \int \breve{r}_n(x, X_j)\mathbb{E}[\breve{r}_n(x, X_i)]dx,\end{aligned}$$

which equals zero because the mean of $\breve{r}_n(x, X_i)$ is by construction zero. We then apply Theorem 4.7.3 of Koroljuk and Borovskich ((1994), p. 163) for degenerate U-statistics, which states that if, for some $k > 1$,

$$\frac{\mathbb{E}_{X_1,X_2}[G_n^k(X_1, X_2)] + n^{1-k}\mathbb{E}_{X_1,X_2}[H_n^{2k}(X_1, X_2)]}{\mathbb{E}^k_{X_1,X_2}[H_n^2(X_1, X_2)]} \longrightarrow 0, \tag{3.4}$$

where $G_n(X_1, X_2) = \mathbb{E}_{X_3}[H_n(X_3, X_1)H_n(X_3, X_2)]$, then

$$nb^{1/4}I_{1n} \xrightarrow{d} N\left(0, \frac{n^4 b^{1/2}}{2}\mathbb{E}_{X_1,X_2}[H_n^2(X_1, X_2)]\right).$$

We start by establishing that the denominator of (3.4) is of order $O(n^{-4k}b^{-k/2})$ as a by-product of the derivation of the asymptotic variance σ_G^2 of $nb^{1/4}I_{1n}$. Indeed,

$$\begin{aligned}\sigma_G^2 &\equiv \frac{n^4 b^{1/2}}{2}\mathbb{E}_{X_1,X_2}[H_n^2(X_1, X_2)] \\ &= 2b^{1/2} \oint \left[\int \breve{r}_n(x, x_1)\breve{r}_n(x, x_2)dx\right]^2 f(x_1, x_2)d(x_1, x_2) \\ &= 2b^{1/2} \oint \breve{r}_n(x, x_1)\breve{r}_n(x, x_2)\breve{r}_n(y, x_1)\breve{r}_n(y, x_2)f(x_1)f(x_2)d(x, y, x_1, x_2) \\ &= 2b^{1/2} \oint \left[\int \breve{r}_n(x, x_1)\breve{r}_n(y, x_1)f(x_1)dx_1\right]^2 d(x, y) \\ &= 2b^{1/2} \oint \mathbb{E}^2_{X_1}[\breve{r}_n(x, X_1)\breve{r}_n(y, X_1)]d(x, y) \\ &= 2b^{1/2} \oint \gamma_n^2(x, y)d(x, y),\end{aligned}$$

where $\gamma_n(x, y) = C_n(x, y) - \bar{C}_n(x, y)$, $C_n(x, y) \equiv \mathbb{E}_{X_1}[r_n(x, X_1)r_n(y, X_1)]$, and $\bar{C}_n(x, y) \equiv \mathbb{E}_{X_1}[r_n(x, X_1)]\mathbb{E}_{X_1}[r_n(y, X_1)]$. It is easy to show that the latter equals $\varphi^{1/2}(x)\varphi^{1/2}(y)f(x)f(y) + O(b)$, whereas

$$C_n(x, y) = \varphi^{1/2}(x)\varphi^{1/2}(y) \int K_{x/b+1,b}(x_1)K_{y/b+1,b}(x_1)dF(x_1)$$

$$= \varphi^{1/2}(x)\varphi^{1/2}(y)B_b(x,y)\int K_{(x+y)/b+1,b/2}(x_1)dF(x_1)$$
$$= \varphi^{1/2}(x)\varphi^{1/2}(y)B_b(x,y)\mathbb{E}_\varsigma[f(\varsigma)],$$

where $\varsigma \sim \mathcal{G}[(x+y)/b+1, b/2]$ and

$$B_b(x,y) = \frac{\Gamma[(x+y)/b+1]}{\Gamma(x/b+1)\Gamma(y/b+1)}\frac{b^{-1}}{2^{(x+y)/b+1}}. \tag{3.5}$$

As before, we rewrite (3.5) in terms of the R-function so as to get

$$B_b(x,y) = \frac{R\left(\frac{x}{b}\right)R\left(\frac{y}{b}\right)}{R\left(\dfrac{x+y}{b}\right)}\left(\frac{x+y}{2x}\right)^{x/b+1/4}\left(\frac{x+y}{2y}\right)^{y/b+1/4}\frac{b^{-1/2}x^{-1/4}y^{-1/4}}{2\sqrt{\pi}}$$
$$= \bar{B}_b(x,y)[\Delta(x,y)]^{1/b},$$

with

$$\bar{B}_b(x,y) = \frac{R\left(\frac{x}{b}\right)R\left(\frac{y}{b}\right)}{R\left(\dfrac{x+y}{b}\right)}\left(\frac{x+y}{2x}\right)^{1/4}\left(\frac{x+y}{2y}\right)^{1/4}\frac{b^{-1/2}x^{-1/4}y^{-1/4}}{2\sqrt{\pi}} \tag{3.6}$$

$$\Delta(x,y) = \left(\frac{x+y}{2x}\right)^{x}\left(\frac{x+y}{2y}\right)^{y}. \tag{3.7}$$

Applying a Taylor expansion to $\mathbb{E}_\varsigma[f(\varsigma)]$ then yields

$$\sigma_G^2 = 2b^{1/2}\oint \varphi(x)\varphi(y)\left\{B_b(x,y)\left[f\left(\frac{x+y}{2}\right)+O(b)\right]-f(x)f(y)\right\}^2 d(x,y)$$
$$= 2b^{1/2}\oint \varphi(x)\varphi(y)B_b^2(x,y)\left[f\left(\frac{x+y}{2}\right)\right]^2 d(x,y)\{1+o(1)\}$$
$$= 2b^{1/2}\oint \varphi(x)\varphi(y)\bar{B}_b^2(x,y)[\Delta(x,y)]^{2/b}\left[f\left(\frac{x+y}{2}\right)\right]^2 d(x,y)\{1+o(1)\},$$

which reduces to

$$\sigma_G^2 = 2\sqrt{2\pi}b\int x^{1/2}\varphi^2(x)\bar{B}_b^2(x,x)f(x)dF(x)\{1+o(1)\},$$

by Lemma A.4 in the Appendix.

The second and third terms on the right-hand side of (3.6) equal one if $x=y$, resulting in $\bar{B}_b(x,x) = B_b(x)$. It then holds that

$$\sigma_G^2 = 2\sqrt{2\pi}b\int_0^\infty x^{1/2}\varphi^2(x)B_b^2(x)f(x)dF(x)\{1+o(1)\}$$
$$= 2\sqrt{2\pi}b\int_0^b x^{1/2}\varphi^2(x)B_b^2(x)f(x)dF(x)+o(1)$$
$$+ 2\sqrt{2\pi}b\int_b^\infty x^{1/2}\varphi^2(x)B_b^2(x)f(x)dF(x)$$

$$= 2\sqrt{2\pi}b \int_b^\infty x^{1/2}\varphi^2(x)\left(\frac{b^{-1}x^{-1}}{4\pi}\right) f(x)dF(x) + o(1)$$
$$= \frac{1}{\sqrt{2\pi}}\mathbb{E}_{X_1}[X_1^{-1/2}\varphi^2(X_1)f(X_1)] + o(1),$$

for b small enough.

Denote by Ψ_n and Λ_n the first and second terms of the numerator in (3.4), respectively. We show that $\Psi_n \le O(n^{-4k}b^{1-5k/4})$ and then demonstrate that $\Lambda_n = O(n^{1-5k}b^{(1-2k)/2})$, ensuring that (3.4) holds for $1 < k < 4/3$. In what follows, we treat k as an integer. This is without any loss of generality given that, for any $k > 1$, there exists an integer $k_1 > 1$ such that $k = k_1/k_2$ and, by Jensen's inequality, $\mathbb{E}(Z^k) \le [\mathbb{E}(Z^{k_1})]^{1/k_2}$ for any nonnegative random variable Z. Observe that

$$\begin{aligned}\Psi_n &\equiv \mathbb{E}_{X_1,X_2}\{\mathbb{E}^k_{X_3}[H_n(X_3,X_1)H_n(X_3,X_2)]\}\\ &= 4^k n^{-4k}\mathbb{E}_{X_1,X_2}\left\{\mathbb{E}^k_{X_3}\left[\oint \breve{r}_n(x,X_3)\breve{r}_n(y,X_3)\breve{r}_n(x,X_1)\breve{r}_n(y,X_2)d(x,y)\right]\right\}\\ &= 4^k n^{-4k}\mathbb{E}_{X_1,X_2}\left\{\left[\oint \gamma_n(x,y)\breve{r}_n(x,X_1)\breve{r}_n(y,X_2)d(x,y)\right]^k\right\}.\end{aligned}$$

Jensen's inequality then ensures that

$$\begin{aligned}\Psi_n &\le 4^k n^{-4k}\oint\left[\frac{\gamma_n(x,y)}{f(x)f(y)}\right]^k |\breve{r}_n(x,x_1)\breve{r}_n(y,x_2)|^k dF(x,y)dF(x_1)dF(x_2)\\ &= 4^k n^{-4k}\oint\left[\frac{\gamma_n(x,y)}{f(x)f(y)}\right]^k \mathbb{E}_{X_1}|\breve{r}_n(x,X_1)|^k\mathbb{E}_{X_1}|\breve{r}_n(y,x_1)|^k dF(x,y).\end{aligned}$$

However, centering implies that $\mathbb{E}_{X_1}|\breve{r}_n(u,X_1)|^k$ is at most of the same order of $\mathbb{E}_{X_1}[r_n^k(u,X_1)] = O(B_b^{(k)}(u))$, where

$$B_b^{(k)}(u) = \frac{R^k(u/b)}{R(ku/b)}(2\pi b)^{(1-k)/2}k^{-1/2} \le O(b^{(1-k)/2}).$$

It thus follows that

$$\Psi_n \le O(n^{-4k}b^{-k/4}[B_b^{(k)}(x)]^2) = O(n^{-4k}b^{1-5k/4})$$

in view that $\gamma_n(x,y) = O(b^{-1/4})$. In turn, the second term of the numerator in (3.4) reads

$$\begin{aligned}\Lambda_n &\equiv n^{1-k}\mathbb{E}_{X_1,X_2}[H_n^{2k}(X_1,X_2)]\\ &= 2^{2k}n^{1-5k}\oint\left[\int \breve{r}_n(y,x_1)\breve{r}_n(y,x_2)dy\right]^{2k} dF(x_1)dF(x_2)\\ &= 2^{2k}n^{1-5k}\oint\prod_{i=1}^{2k}\breve{r}_n(y_i,x_1)\breve{r}_n(y_i,x_2)dy_i dF(x_1)dF(x_2)\\ &= 2^{2k}n^{1-5k}\oint\mathbb{E}^2_{X_1}\left[\prod_{i=1}^{2k}\breve{r}_n(y_i,X_1)\right]d(y_1,\ldots,y_{2k}),\end{aligned}$$

which is of order $O(n^{1-5k}b^{(1-2k)/2})$ by Lemma A.5 in the Appendix. The quantity $nb^{1/4}I_{1n}/\sigma_G$ therefore weakly converges to a standard normal distribution, which implies that

$$nb^{1/4}I_n - \frac{b^{-1/4}}{2\sqrt{\pi}}\mathbb{E}_{X_1}[X_1^{-1/2}\varphi(X_1)] \xrightarrow{d} N(0,\sigma_G^2).$$

This proves the main result for the gamma kernel functionals.

4. Beta kernel functionals

In the sequel, we derive in similar fashion the limiting distribution of the beta kernel functional using the decomposition $I_n = I_{1n} + I_{2n} + I_{3n} + I_{4n}$. The only difference is that $r_n(x, X_1)$ now represents $\varphi^{1/2}(x)K_{x/b+1,(1-x)/b+1}(X_1)$ and $\int$ denotes the integral over the unit interval $A = [0, 1]$. Again, the first term stands for a degenerate U-statistic and contributes with the asymptotic variance, whereas the second term provides the asymptotic bias. The third and the fourth terms are, once more, negligible as long as the bandwidth b is of order $o(n^{-4/9})$. As before, this assumption precludes the use of the optimal bandwidth that is of order $O(n^{-2/5})$ as shown in Chen (1999).

To begin, note that $\mathbb{E}_{X_1}[r_n(x, X_1)] = \varphi^{1/2}(x)\mathbb{E}_\zeta[f(\zeta)]$, where ζ has a beta distribution $\mathcal{B}(x/b + 1, (1 - x)/b + 1)$. Chen (1999) demonstrates that the beta kernel density estimation has a uniform bias of order $O(b)$, and hence it turns out that $\mathbb{E}_{X_1}[r_n(x, X_1)] = \varphi^{1/2}(x)f(x) + O(b)$.

The second moment of $r_n(x, X_1)$ reads

$$\begin{aligned}\mathbb{E}_{X_1}[r_n^2(x, X_1)] &= \varphi(x)\int K^2_{x/b+1,(1-x)/b+1}(x_1)f(x_1)dx_1\\ &= \varphi(x)A_b(x)\mathbb{E}_\eta[f(\eta)],\end{aligned}$$

where $\eta \sim \mathcal{B}(2x/b + 1, 2(1 - x)/b + 1)$ and

$$A_b(x) = \frac{B[2x/b + 1, 2(1 - x)/b + 1]}{B^2[x/b + 1, (1 - x)/b + 1]}.$$

Chen (1999) shows that $\mathbb{E}_\eta[f(\eta)] = f(x) + O(b)$, thus

$$\begin{aligned}\mathbb{E}(I_{2n}) &= \frac{1}{n}\int \mathbb{E}_{X_1}[r_n^2(x, X_1)]dx - \frac{1}{n}\int \mathbb{E}^2_{X_1}[r_n(x, X_1)]dx\\ &= \frac{1}{n}\int \varphi(x)A_b(x)[f(x) + O(b)]dx + O(n^{-1}).\end{aligned}$$

For b small enough, Chen (1999) shows that $A_b(x)$ may be approximated according to the location of x within the support. More precisely, x/b and $(1-x)/b$ grows without bound as b shrinks to zero in the interior of the support, whereas either x/b or $(1-x)/b$ converges to some nonnegative constant c in the boundaries. The approximation is such that

$$A_b(x) \sim \begin{cases} \frac{1}{2\sqrt{\pi}}b^{-1/2}[x(1-x)]^{-1/2} & \text{if } \ x/b \text{ and } (1-x)/b \to \infty \\ \frac{\Gamma(2c+1)/b}{2^{2c+1}\Gamma^2(c+1)} & \text{if } \ x/b \text{ or } (1-x)/b \to c, \end{cases}$$

which implies that $A_b(x)$ is of larger order near the boundary. As before, there is no impact in $\mathbb{E}(I_{2n})$.

Let $\kappa = b^{1-\epsilon}$, where $0 < \epsilon < 1$. Then,

$$\begin{aligned}\mathbb{E}(I_{2n}) &= \frac{1}{n}\int \varphi(x)A_b(x)[f(x)+O(b)]dx + O(n^{-1})\\ &= \frac{1}{n}\int_0^{\kappa} + \int_{\kappa}^{1-\kappa} + \int_{1-\kappa}^{1} \varphi(x)A_b(x)[f(x)+O(b)]dx + O(n^{-1})\\ &= \frac{1}{2\sqrt{\pi}n}\int_{\kappa}^{1-\kappa} b^{-1/2}[x(1-x)]^{-1/2}\varphi(x)[f(x)+O(b)]dx\\ &\quad + O(n^{-1}b^{-\epsilon}+n^{-1})\\ &= \frac{b^{-1/2}}{2\sqrt{\pi}n}\int_0^1 \varphi(x)[x(1-x)]^{-1/2}f(x)dx + o(n^{-1}b^{-1/2})\end{aligned}$$

as long as ϵ is properly chosen and $\mathbb{E}_{X_1}\{[X_1(1-X_1)]^{-1/2}\varphi(X_1)\}$ is finite. Further, as in the previous section, $V(I_{2n}) = O(n^{-3}b^{-2}+n^{-3}b^{-1})$, which means that $V(nb^{1/4}I_{2n}) = O(n^{-1}b^{-5/2}) = o(1)$. It thus ensues from the Chebyshev's inequality that $nb^{1/4}I_{2n} - \frac{b^{-1/4}}{2\sqrt{\pi}}\mathbb{E}_{X_1}\{[X_1(1-X_1)]^{-1/2}\varphi(X_1)\} = o_p(1)$.

Applying the same techniques as in the previous section, it is straightforward to demonstrate that the third and fourth terms are negligible provided that the bandwidth is of order $o(n^{-4/9})$. It remains to compute the variance of the degenerate U-statistic $I_{1n} = \sum_{i<j} H_n(X_i, X_j)$, viz.

$$\sigma_B^2 = \frac{n^4 b^{1/2}}{2}\mathbb{E}_{X_1,X_2}[H_n^2(X_1,X_2)] = 2b^{1/2}\oint [C_n(x,y) - \bar{C}_n(x,y)]^2 d(x,y),$$

where $\bar{C}_n(x,y) = \varphi^{1/2}(x)\varphi^{1/2}(y)f(x)f(y) + O(b)$ as before. In turn,

$$\begin{aligned}C_n(x,y) &= \varphi^{1/2}(x)\varphi^{1/2}(y)\int K_{x/b+1,(1-x)/b+1}(x_1)K_{y/b+1,(1-y)/b+1}(x_1)dF(x_1)\\ &= \varphi^{1/2}(x)\varphi^{1/2}(y)A_b(x,y)\int K_{(x+y)/b+1,(1-x-y)/b+1}(x_1)dF(x_1)\\ &= \varphi^{1/2}(x)\varphi^{1/2}(y)A_b(x,y)\mathbb{E}_{\varsigma}[f(\varsigma)],\end{aligned}$$

where $\varsigma \sim \mathcal{B}[(x+y)/b+1, (2-x-y)/b+1]$ and

$$A_b(x,y) = \frac{B[(x+y)/b+1,(1-x-y)/b+1]}{B[x/b+1,(1-x)/b+1]B[y/b+1,(1-y)/b+1]}.$$

Expressing the beta functions in terms of the R-function then yields

$$A_b(x,y) = \bar{A}_b(x,y)[\bar{\Delta}(x,y)]^{1/b},$$

where $\bar{\Delta}(u,v) \equiv \Delta(u,v)\Delta(1-u,1-v)$ and

$$\bar{A}_b(x,y) = R_b(x,y)D(x,y)\frac{b^{-1/2}x^{-1/4}y^{-1/4}}{2\sqrt{\pi}}$$

with

$$R_b(x,y) = \frac{R(x/b)R(y/b)}{R\left(\frac{x+y}{b}\right)}\frac{R\left(\frac{1-x}{b}\right)R\left(\frac{1-y}{b}\right)}{R\left(\frac{2-x-y}{b}\right)}\frac{R(2/b)}{R^2(1/b)}$$

$$D(x,y) = \left(\frac{x+y}{2x}\right)^{1/4}\left(\frac{x+y}{2y}\right)^{1/4}\left(\frac{2-x+y}{2(1-x)}\right)^{1/4}\left(\frac{2-x+y}{2(1-y)}\right)^{1/4}.$$

It then follows that

$$\begin{aligned}
\sigma_B^2 &= 2b \oint \varphi(x)\varphi(y) A_b^2(x,y) \mathbb{E}_\varsigma^2[f(\varsigma)] d(x,y) + O(b) \\
&= 2b \oint \varphi(x)\varphi(y) \left\{ \bar{A}_b(x,y)[\bar{\Delta}(x,y)]^{1/b} f\left(\frac{x+y}{2}\right)\right\}^2 d(x,y) + O(b)
\end{aligned}$$

by Taylor expanding $\mathbb{E}_\varsigma[f(\varsigma)]$. However, the term $\bar{\Delta}(x,y)$ works similarly to the term $\Delta(x,y)$ in Lemma A.4, yielding

$$\begin{aligned}
\sigma_B^2 &= 2\sqrt{2\pi} b \int \varphi^2(x) \bar{A}_b^2(x,x)[x(1-x)]^{1/2} f(x) dF(x)\{1+o(1)\} \\
&= \frac{1}{\sqrt{2\pi}} \int \varphi^2(x)[x(1-x)]^{-1/2} f(x) dF(x)\{1+o(1)\} \\
&= \frac{1}{\sqrt{2\pi}} \mathbb{E}\{\varphi^2(x)[x(1-x)]^{-1/2} f(x)\}\{1+o(1)\}
\end{aligned}$$

for b small enough. Applying then Koroljuk and Borovskich's (1994) central limit theorem for degenerate U-statistics gives way to

$$nb^{1/4} I_n - \frac{b^{-1/4}}{\sqrt{2\pi}} \mathbb{E}_{X_1}\{[X_1(1-X_1)]^{-1/2}\varphi(X_1)\} \xrightarrow{d} N(0, \sigma_B^2),$$

which completes the proof. To demonstrate that condition (3.4) holds, it suffices to apply the same technique as before. The only difference is that, instead of $B_b^{(k)}(\cdot)$, there will be an analogous residual term, say $A_b^{(k)}(\cdot)$, stemming from the extraction of a unique beta density out of the product of k beta kernels.

Acknowledgements

We are indebted to Sergio O. Parreiras, four anonymous referees, and seminar participants at the 5th World Congress of the Bernoulli Society for Probability Theory and Mathematical Statistics and the 63rd Annual Meeting of the Institute of Mathematical Statistics (Guanajuato, 2000) for valuable comments. The authors acknowledges with gratitude the financial support from CNPq-Brazil. The usual disclaimer applies.

Appendix

Derivation of (3.3). Because $V(I_{2n})$ is at most of the same order of $\mathbb{E}(I_{2n}^2)$, it suffices to compute the order of the latter. Thus,

$$\begin{aligned}
\mathbb{E}(I_{2n}^2) &= \frac{1}{n^3} \mathbb{E}_{X_1}\left[\int \check{r}_n^2(x, X_1) dx\right]^2 \\
&= \frac{1}{n^3} \mathbb{E}_{X_1} \oint \check{r}_n^2(x, X_1) \check{r}_n^2(y, X_1) d(x,y)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^3} \oint \mathbb{E}_{X_1}[r_n^2(x, X_1) r_n^2(y, X_1)] d(x, y) \\
&- \frac{1}{n^3} \oint \mathbb{E}_{X_1}[r_n^2(x, X_1)] \mathbb{E}_{X_1}^2[r_n(y, X_1)] d(x, y) \\
&- \frac{1}{n^3} \oint \mathbb{E}_{X_1}[r_n^2(y, X_1)] \mathbb{E}_{X_1}^2[r_n(x, X_1)] d(x, y) \\
&+ \frac{1}{n^3} \oint \mathbb{E}_{X_1}^2[r_n(x, X_1)] \mathbb{E}_{X_1}^2[r_n(y, X_1)] d(x, y).
\end{aligned}$$

It is readily seen that the second and third terms are of order $O(n^{-3}b^{-1/2})$, whereas the fourth term is $O(n^{-3})$. It then remains to show that the first term is of order $O(n^{-3}b^{-2})$. Notice that

$$\begin{aligned}
\Upsilon_n &= \mathbb{E}_{X_1}\left[\oint r_n^2(x, X_1) r_n^2(y, X_1) d(x, y)\right] \\
&= \mathbb{E}_{X_1}\left[\oint \varphi(x)\varphi(y) K^2_{x/b+1,b}(X_1) K^2_{y/b+1,b}(X_1) d(x, y)\right] \\
&= \mathbb{E}_{X_1}\left[\oint \varphi(x)\varphi(y) B_b(x) B_b(y) K_{2x/b+1,b/2}(X_1) K_{2y/b+1,b/2}(X_1) d(x, y)\right] \\
&= \mathbb{E}_{X_1}\left[\oint \varphi(x)\varphi(y) B_b(x) B_b(y) B_{b/2}(x, y) K_{2(x+y)/b+1,b/4}(X_1) d(x, y)\right].
\end{aligned}$$

This means that

$$\Upsilon_n = \oint \varphi(x)\varphi(y) B_b(x) B_b(y) B_{b/2}(x, y) \mathbb{E}_\omega[f(\omega)] d(x, y),$$

where $\omega \sim \mathcal{G}(2(x+y)/b + 1, b/4)$. Applying a Taylor expansion results in $\mathbb{E}_\omega[f(\omega)] = f(\frac{x+y}{2}) + O(b) = O(1)$, whereas rewriting $B_b(\cdot)$ in terms of the R-function yields

$$B_b(u) = \frac{R^2(u/b)}{R(2u/b)} \frac{u^{-1/2}b^{-1/2}}{2\sqrt{\pi}} \leq \frac{u^{-1/2}b^{-1/2}}{2\sqrt{\pi}}.$$

Similarly, one obtains

$$B_{b/2}(x, y) = \frac{R\left(\frac{2x}{b}\right) R\left(\frac{2y}{b}\right)}{R\left(\frac{2(x+y)}{b}\right)} \left(\frac{x+y}{2x}\right)^{2x/b+1/4} \left(\frac{x+y}{2y}\right)^{2y/b+1/4} \frac{b^{-1/2}x^{-1/4}y^{-1/4}}{\sqrt{2\pi}}.$$

Because $R(\cdot)$ is a monotonic increasing function that never exceeds one (Brown and Chen (1999), Lemma 3) and $\Delta(x, y) \leq 1$, it then follows that

$$B_{b/2}(x, y) \leq \left(\frac{x+y}{2x}\right)^{1/4} \left(\frac{x+y}{2y}\right)^{1/4} \frac{b^{-1/2}x^{-1/4}y^{-1/4}}{\sqrt{2\pi}} = \left(\frac{x+y}{xy}\right)^{1/2} \frac{b^{-1/2}}{2\sqrt{\pi}}.$$

Because both x and y are at most of $O(b^{1-\varepsilon})$, where $\varepsilon > 0$, the first term is at most of order $O(b^{\varepsilon/2-1/2})$. It then follows that $B_{b/2}(x, y) \leq O(b^{-1})$, which implies that Υ_n is at most of order $O(b^{-2})$, completing the proof.

LEMMA A.1. *If $y \geq x$, then $\bar{\Delta}(y) \equiv \log \Delta(x,y) \geq -\frac{1}{4x}(y-x)^2$.*

PROOF. The argument relies on the third-order Taylor series expansion of $\bar{\Delta}(y) = (x+y)\log(x+y) - (x+y)\log 2 - x\log x - y\log y$ around x. It follows from $\bar{\Delta}'(y) = \log(x+y) - \log 2 - \log y$, $\bar{\Delta}''(y) = (x+y)^{-1} - y^{-1} < 0$ and $\bar{\Delta}'''(y) = -(x+y)^{-2} + y^{-2} > 0$ that $\bar{\Delta}(x) = \bar{\Delta}'(x) = 0$, $\bar{\Delta}''(x) = -1/(2x)$ and $\bar{\Delta}'''(x) = 3/(4x^2)$. It thus holds that, for $x \leq x^* \leq y$,

$$\bar{\Delta}(y) = -\frac{(y-x)^2}{4x} + \frac{(y-x)^3}{6}\bar{\Delta}'''(x^*).$$

The result then immediately follows from the fact that $\bar{\Delta}'''(\cdot) \geq 0$. □

LEMMA A.2. *If $t > 0$, then $\bar{\Delta}(y) \geq \frac{1}{8x^2}(y-x)^2(y-x)$ for $x(1-t) \leq y \leq x$.*

PROOF. As before, the following Taylor expansion holds

$$\bar{\Delta}(y) = -\frac{(y-x)^2}{4x} + \frac{(y-x)^3}{6}\bar{\Delta}'''(x^*),$$

where $x \leq x^* \leq x(1-t)$. As $\bar{\Delta}'''$ is a decreasing function, $\bar{\Delta}'''(x^*) \leq \bar{\Delta}'''(x)$ and thus $\bar{\Delta}'''(x^*)(y-x)^3 \geq \bar{\Delta}'''(x)(y-x)^3$ for $y \leq x$. It then ensues that

$$\bar{\Delta}(y) \geq -\frac{(y-x)^2}{4x} + \frac{(y-x)^3}{8x^2} = \frac{(y-x)^2(y-x)}{8x^2},$$

completing the proof. □

LEMMA A.3. *For every y, $\bar{\Delta}(y) \leq \frac{1}{8x^2}(y-x)^2(y-x)$.*

PROOF. In view that $\bar{\Delta}''''(y) = 2(x+y)^{-3} - 2y^{-3} < 0$, the result readily follows from the fourth-order Taylor expansion

$$\bar{\Delta}(y) = -\frac{(y-x)^2}{4x} + \frac{(y-x)^2(y-x)}{8x^2} + \bar{\Delta}''''(x^*)\frac{(y-x)^4}{24},$$

where x^* lies in the interval between x and y. □

LEMMA A.4. *Let $\ell(b) = \int_0^\infty \int_0^\infty \Delta^{2/b}(x,y)g(x,y)d(x,y)$ with Δ as in (3.7) and $\tilde{g}(y) = \int_0^\infty g(x,y)dx$ such that $|\tilde{g}|_\infty$ exists. It then holds that $\ell(b) = (2\pi b)^{1/2}\int_0^\infty x^{1/2}g(x,x)dx$.*

PROOF. Let $0 < t < 1$, then

$$\ell(b) \geq \int_0^\infty \int_x^\infty \Delta^{2/b}(x,y)g(x,y)dydx + \int_0^\infty \int_{x(1-t)}^x \Delta^{2/b}(x,y)g(x,y)dydx.$$

Applying Lemmas A.1 and A.2 yields

$$\ell(b) \geq \int_0^\infty \int_x^\infty \exp\left[-\frac{(y-x)^2}{2xb}\right] g(x,y)dydx$$

$$+\int_0^\infty \int_{x(1-t)}^x \exp\left[-\frac{(y-x)^2(y-3x)}{4x^2b}\right] g(x,y)dydx$$
$$\geq \int_0^\infty \int_x^\infty \exp\left[-\frac{(y-x)^2}{2xb}\right] g(x,y)dydx$$
$$+\int_0^\infty \int_{x(1-t)}^x \exp\left[-\frac{(y-x)^2(t+2)}{4xb}\right] g(x,y)dydx.$$

Letting $z=(2xb)^{-1/2}(y-x)$ results in

$$\int_0^\infty \int_x^\infty \exp\left[-\frac{(y-x)^2}{2xb}\right] g(x,y)dydx$$
$$=\int_0^\infty (2xb)^{1/2}\int_0^\infty \exp(-z^2)g(x,x+z(2xb)^{1/2})dzdx,$$

and hence

$$\liminf_{b\to 0} b^{-1/2}\int_0^\infty \int_x^\infty \exp\left[-\frac{(y-x)^2}{2xb}\right] g(x,y)dydx$$
$$\geq \int_0^\infty \sqrt{2x}\int_0^\infty \exp(-z^2)g(x,x)dzdx$$
$$=\sqrt{\pi/2}\int_0^\infty x^{1/2}g(x,x)dx.$$

Letting now $w=(4xb)^{-1/2}(t+2)^{1/2}(y-x)$ yields

$$\int_0^\infty \int_{x(1-t)}^x \exp\left[-\frac{(y-x)^2(t+2)}{4xb}\right] g(x,y)dydx$$
$$=\int_0^\infty \sqrt{\frac{4xb}{t+2}}\int_{-xt\sqrt{(t+2)/4xb}}^0 \exp(-w^2)g\left(x,x+w\sqrt{\frac{4xb}{t+2}}\right)dwdx,$$

and thus

$$\liminf_{b\to 0} b^{-1/2}\int_0^\infty \int_{x(1-t)}^x \exp\left[-\frac{(y-x)^2(t+2)}{4xb}\right] g(x,y)dydx$$
$$\geq \int_0^\infty \sqrt{\frac{4x}{t+2}}\int_{-\infty}^0 \exp(-w^2)g(x,x)dwdx$$
$$=\frac{\sqrt{\pi}}{2}\int_0^\infty \sqrt{\frac{4x}{t+2}}g(x,x)dx.$$

It then suffices to let t shrink to zero to appreciate that

$$\liminf_{b\to 0} b^{-1/2}\ell(b)\geq \sqrt{\pi/2}\int_0^\infty x^{1/2}g(x,x)dx.$$

On the other hand, for $t>0$, one may write

$$\ell(b)=\int_0^\infty \int_{x(t+1)}^\infty \Delta^{2/b}(x,y)g(x,y)dydx+\int_0^\infty \int_0^{x(t+1)} \Delta^{2/b}(x,y)g(x,y)dydx.$$

However, for every $y \geq x(t+1)$,

$$\bar{\Delta}(y) \leq \bar{\Delta}(x(t+1)) = x \log \frac{(t+2)^{t+2}}{2^{t+2}(t+1)^{t+1}},$$

so that $\Delta(x,y) \leq a_t^x$, where $a_t \equiv \frac{(t+2)^{t+2}}{2^{t+2}(t+1)^{t+1}} < 1$. The following inequalities then hold

$$\begin{aligned}\int_0^\infty \int_{x(t+1)}^\infty \Delta^{2/b}(x,y) g(x,y) dy dx &\leq \int_0^\infty \int_0^\infty a_t^{2x/b} g(x,y) dy dx \\ &\leq |\tilde{g}|_\infty \int_0^\infty a_t^{2x/b} dx = -\frac{b|\tilde{g}|_\infty}{2\log a_t},\end{aligned}$$

where $\tilde{g}(x) \equiv \int g(x,y)dy$. It also follows from Lemma A.3 that

$$\begin{aligned}&\int_0^\infty \int_0^{x(t+1)} \Delta^{2/b}(x,y) g(x,y) dy dx \\ &\quad\leq \int_0^\infty \int_0^{x(t+1)} \exp\left[-\frac{(y-x)^2(y-3x)}{4x^2 b}\right] g(x,y) dy dx \\ &\quad\leq \int_0^\infty \int_0^{x(t+1)} \exp\left[-\frac{(y-x)^2(2-t)}{4xb}\right] g(x,y) dy dx.\end{aligned}$$

Letting $\omega = (4xb)^{-1/2}(2-t)^{1/2}(y-x)$ results in

$$\begin{aligned}&\int_0^\infty \int_0^{x(t+1)} \Delta^{2/b}(x,y) g(x,y) dy dx \\ &\quad\leq \int_0^\infty \sqrt{\frac{4xb}{2-t}} \int_{-\infty}^\infty \exp(-\omega^2) g\left(x, x+\omega\sqrt{\frac{4xb}{2-t}}\right) d\omega dx.\end{aligned}$$

It now suffices to assume that $x^{1/2} g(x,y) \leq x^{1/2}\bar{g}(x) \in L^1$ for every $y \geq x$ to ensure that

$$b^{-1/2}\ell(b) \leq \int_0^\infty \sqrt{\frac{4x}{2-t}} \int_{-\infty}^\infty \exp(-\omega^2) g\left(x, x+\omega\sqrt{\frac{4xb}{2-t}}\right) d\omega dx + o(\sqrt{b})$$

and hence

$$\limsup_{b\to 0} b^{-1/2}\ell(b) \leq \frac{\sqrt{\pi}}{2} \int_0^\infty \sqrt{\frac{4x}{2-t}} g(x,x) dx,$$

which yields the desired result for $t \to 0$. □

LEMMA A.5. *Let $x_{(m)} = (x_1, \ldots, x_m)$, $s_m = \sum_{j=1}^m x_j$ and*

$$\Delta_m(x_{(m)}) = \frac{(s_{m-1}+x_m)^{s_{m-1}+x_m}}{m^{s_{m-1}+x_m}\prod_{i=1}^m x_i^{x_i}}.$$

Suppose that

$$g(x_{(m)}) = f^2(s_m/m) s_m \prod_{i=1}^m \frac{\varphi(x_i)}{x_i}$$

is bounded and there exists $h(x_1) \in L^1$ such that $g(x_{(m)}) \leq h(x_1)$ if $x_i \geq x_1$ for every $i \geq 2$. It then follows that $\oint \Delta_m^{1/b}(x_{(m)})g(x_{(m)})dx_{(m)} = O(b^{(m-1)/2})$.

PROOF. Let $\ell_m(b) = \oint \Delta_m^{1/b}(x_{(m)})g(x_{(m)})dx_{(m)}$ and

$$\begin{aligned} G_m(\alpha) &= \log \Delta_m(x_{(m-1)}, \alpha) \\ &= (s_{m-1} + \alpha)\log\left(\frac{s_{m-1}+\alpha}{m}\right) - \sum_{i=1}^{m-1} x_i \log x_i - \alpha \log \alpha. \end{aligned}$$

Differentiating $G_m(\alpha)$ with respect to α then gives way to

$$\begin{aligned} G'_m(\alpha) &= \log\left(\frac{s_{m-1}+\alpha}{m}\right) - \log \alpha \\ G''_m(\alpha) &= \frac{1}{s_{m-1}+\alpha} - \frac{1}{\alpha} = -\frac{s_{m-1}}{\alpha(s_{m-1}+\alpha)} < 0 \\ G'''_m(\alpha) &= -\frac{1}{(s_{m-1}+\alpha)^2} + \frac{1}{\alpha^2} > 0, \end{aligned}$$

whereas $G''''_m(\alpha) < 0$. Evaluating at $\alpha = \frac{s_{m-1}}{m-1}$ then yields

$$\begin{aligned} G_m\left(\frac{s_{m-1}}{m-1}\right) &= s_{m-1}\log\left(\frac{s_{m-1}}{m-1}\right) - \sum_{i=1}^{m-1} x_i \log x_i \\ G'_m\left(\frac{s_{m-1}}{m-1}\right) &= 0 \\ G''_m\left(\frac{s_{m-1}}{m-1}\right) &= -\frac{(m-1)^2}{m s_{m-1}} \\ G'''_m\left(\frac{s_{m-1}}{m-1}\right) &= -\left(\frac{m-1}{m s_{m-1}}\right)^2 + \left[\frac{m(m-1)}{m s_{m-1}}\right]^2 = \frac{(m-1)^2(m^2-1)}{m^2 s_{m-1}^2}. \end{aligned}$$

As in the proof of Lemma A.4, we let $0 < t < 1$ and decompose the integral in $\ell_m(b)$ into the sum of the integral over $x_m \geq s^*_{m-1} \equiv (1+t)s_{m-1}/(m-1)$ and the integral over $x_m \leq s^*_{m-1}$. In the first case, the fact that $x_m \geq s^*_{m-1}$ and $G'(s^*_{m-1}) \leq G'(\frac{s_{m-1}}{m-1}) = 0$ implies that

$$\begin{aligned} G_m(x_m) &\leq G_m(s^*_{m-1}) \\ &= (s_{m-1} + s^*_{m-1})\log\left(\frac{s_{m-1}+s^*_{m-1}}{m}\right) - \sum_{i=1}^{m-1} x_i \log x_i - s^*_{m-1}\log s^*_{m-1} \\ &= \frac{s_{m-1}(m+t)}{m-1}\log\left[\frac{s_{m-1}(m+t)}{m(m-1)}\right] - \sum_{i=1}^{m-1} x_i \log x_i \\ &\quad - \frac{s_{m-1}(1+t)}{m-1}\log\left[\frac{s_{m-1}(1+t)}{m-1}\right] \\ &\leq \frac{s_{m-1}(m+t)}{m-1}\log\left[\frac{s_{m-1}(m+t)}{m(m-1)}\right] - s_{m-1}\log\left(\frac{s_{m-1}}{m-1}\right) \end{aligned}$$

$$- \frac{s_{m-1}(1+t)}{m-1} \log \left[\frac{s_{m-1}(1+t)}{m-1} \right]$$
$$= \frac{s_{m-1}}{m-1} \log \left\{ \frac{[(m+t)/m]^{m+t}}{(1+t)^{1+t}} \right\}.$$

Letting a_t denote the term within curled brackets and

$$\tilde{g}(x_{(m-1)}) = \int g(x_{(m-1)}, x_m) dx_m$$

then yields

$$\begin{aligned}
&\oint_{x_m \geq s^*_{m-1}} \Delta_m^{1/b}(x_{(m)}) g(x_{(m)}) dx_{(m)} \\
&\quad \leq \oint_{x_m \geq s^*_{m-1}} a_t^{s_{m-1}/b(m-1)} g(x_{(m)}) dx_{(m)} \leq \oint a_t^{s_{m-1}/b(m-1)} \tilde{g}(x_{(m-1)}) dx_{(m-1)} \\
&\quad \leq |\tilde{g}|_\infty \oint a_t^{s_{m-1}/b(m-1)} dx_{(m-1)} = |\tilde{g}|_\infty \left\{ \int a_t^{x_1/b(m-1)} dx_1 \right\}^{m-1} \\
&\quad = |\tilde{g}|_\infty \left(\frac{m-1}{\log a_t} \right)^{m-1} b^{m-1} = O(b^{m-1}).
\end{aligned}$$

As for the integral over $x_m \leq s^*_{(m-1)}$, we first observe that

$$\begin{aligned}
G_m(\alpha) &= G_m\left(\frac{s_{m-1}}{m-1}\right) - \frac{1}{2!} \frac{(m-1)^2}{m s_{m-1}} \left(\alpha - \frac{s_{m-1}}{m-1}\right)^2 \\
&\quad + \frac{1}{3!} \frac{(m-1)^2(m^2-1)}{m^2 s^2_{m-1}} \left(\alpha - \frac{s_{m-1}}{m-1}\right)^3 + \frac{1}{4!} G''''(\xi) \left(\alpha - \frac{s_{m-1}}{m-1}\right)^4
\end{aligned}$$

for $\alpha \leq \xi \leq s^*_{(m-1)}$. It follows from $G''''(\xi) < 0$ that

$$\begin{aligned}
G_m(\alpha) &\leq G_m\left(\frac{s_{m-1}}{m-1}\right) - \frac{1}{2} \frac{(m-1)^2}{m s_{m-1}} \left(\alpha - \frac{s_{m-1}}{m-1}\right)^2 \\
&\quad + \frac{1}{6} \frac{(m-1)^2(m^2-1)}{m^2 s^2_{m-1}} \left(\alpha - \frac{s_{m-1}}{m-1}\right)^3 \\
&= G_m\left(\frac{s_{m-1}}{m-1}\right) - \frac{1}{2} \left(\alpha - \frac{s_{m-1}}{m-1}\right)^2 \frac{(m-1)^2}{m s_{m-1}} \left[1 - \frac{m^2-1}{3m s_{m-1}} \left(\alpha - \frac{s_{m-1}}{m-1}\right)\right].
\end{aligned}$$

However, $x_m \leq s^*_{(m-1)}$ implies that $x_m - \frac{s_{m-1}}{m-1} \leq \frac{s_{m-1}}{m-1} t$, and hence

$$1 - \frac{m^2-1}{3m s_{m-1}} \left(x_m - \frac{s_{m-1}}{m-1}\right) \geq 1 - \frac{m+1}{3m} t.$$

Now, because

$$\Delta_m^{1/b}(x_{(m)}) \leq \Delta_{m-1}^{1/b}(x_{(m-1)}) \exp\left[-\frac{1}{b} \left(x_m - \frac{s_{m-1}}{m-1}\right)^2 \frac{(m-1)^2}{2m s_{m-1}} \left(1 - \frac{m+1}{3m} t\right) \right]$$

for $x_m \le s^*_{m-1}$, changing variables to $z_m = b^{-1/2}(x_m - \frac{s_{m-1}}{m-1})$ results in

$$\begin{aligned}\oint_{x_m \le s^*_{m-1}} & \Delta_m^{1/b}(x_{(m)}) g(x_{(m)}) dx_{(m)} \\ & \le b^{1/2} \oint \Delta_{m-1}^{1/b}(x_{(m-1)}) \Bigg\{ \int_{-\infty}^{\infty} \exp\left[-z_m^2 \frac{(m-1)^2}{2ms_{m-1}} \left(1 - \frac{m+1}{3m} t\right)\right] \\ & \qquad \times g\left(x_{(m-1)}, \frac{s_{m-1}}{m-1} + z_m b^{1/2}\right) dz_m \Bigg\} dx_{(m-1)}.\end{aligned}$$

To complete the proof, it suffices to proceeding by induction so as to conclude that $\ell_m(b)$ is at most of order $O(b^{(m-1)/2})$. □

References

Aït-Sahalia, Y., Bickel, P. J. and Stoker, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option-implied volatilities, *Journal of Econometrics*, **105**, 363–412.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Annals of Statistics*, **1**, 1071–1095.

Brown, B. M. and Chen, S. X. (1999). Beta-Bernstein smoothing for regression curves with compact support, *Scandinavian Journal of Statistics*, **26**, 47–59.

Chen, S. X. (1999). Beta kernel estimators for density functions, *Computational Statistics and Data Analysis*, **31**, 131–145.

Chen, S. X. (2000). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics*, **52**, 471–480.

Chen, S. X., Härdle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series, *Journal of the Royal Statistical Society B*, **65**, 663–678.

Fernandes, M. and Grammig, J. (2005). Nonparametric specification tests for conditional duration models, *Journal of Econometrics*, **127**, 35–68.

Hall, P. (1984). Central limit theorem for integrated squared error multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, **14**, 1–16.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric vs. parametric regression fits, *Annals of Statistics*, **21**, 1926–1947.

Koroljuk, V. S. and Borovskich, Y. V. (1994). *Theory of U-statistics*, Kluwer Academic Publishers, Dordrecht.