Supplement to "Large-sample study of the kernel density estimators under multiplicative censoring" by M. Asgharian, M. Carone and V. Fakoor.

Additional technical details: proof of lemmas.

Proof of Lemma 1. a) By the definition of $U_{m,n}$ and (2.4), we have that $\|\hat{G} - G\|_\infty = \|U_{m,n}\|_\infty/\sqrt{k} \leq \|\mathcal{F}_{m,n}^{-1}\|\|W_{m,n}\|_\infty/\sqrt{k}$. Using (2.3), we may write that $\|W_{m,n}\|_\infty/\sqrt{k} \leq \|G_m - G\|_\infty + \|F_n - F\|_\infty$, from which the result follows using the law of the iterated logarithm for empirical distribution functions and the uniform boundedness of $\|\mathcal{F}_{m,n}^{-1}\|$.

b) We may write that

$$\|\hat{f} - f\|_{[a_{m,n},\infty)} \leq \|\hat{f} - f\|_{[a_{m,n},\gamma_{m,n})}\mathbb{I}_{[0,\gamma_{m,n})}(a_{m,n}) + \|\hat{f} - f\|_{[\gamma_{m,n},\infty)}$$

and use that

$$\begin{aligned}\|\hat{f} - f\|_{[a_{m,n},\gamma_{m,n})} &\leq \sup_{a_{m,n}\leq t<\gamma_{m,n}} \left|\int_{t\leq z<\gamma_{m,n}} \frac{1}{z} d\left[\hat{G}(z) - G(z)\right]\right| \\ &\qquad + \left|\int_{z\geq\gamma_{m,n}} \frac{1}{z} d\left[\hat{G}(z) - G(z)\right]\right| \\ &\leq \left[F_U(\gamma_{m,n}) - F_U(a_{m,n})\right]/\mu_U + \|\hat{f} - f\|_{[\gamma_{m,n},\infty)} \;.\end{aligned}$$

In the last inequality above, we use that $\hat{G}$ vanishes below $\gamma_{m,n}$. Because we may also show that $\sup_{\gamma_{m,n}\leq s<\infty} |\hat{f}(s) - f(s)| \leq 2\|\hat{G} - G\|_{[\gamma_{m,n},\infty)}/\gamma_{m,n}$ using integration by parts, the conclusion follows from a).

Proof of Lemma 2. Choose $\epsilon \in (0, 1)$. By Lemma 1.2.1. of [1], there exists a constant $C = C(\epsilon) > 0$ such that

$$\begin{aligned}&\mathrm{pr}\left(\sup_{0\leq x\leq 1} |\mathcal{W}_n(x)| \geq \sqrt{3\log n}\right) \\ &\qquad \leq \mathrm{pr}\left(\sup_{0\leq x\leq 1}\sup_{0\leq y\leq 1} |\mathcal{W}_n(x + y) - \mathcal{W}_n(y)| \geq \sqrt{3\log n}\right) \\ &\qquad \leq 2C\exp\left(-\frac{3\log n}{2+\epsilon}\right) \;.\end{aligned}$$

The result follows from the Borel-Cantelli lemma. Alternatively, the reflection principle may be used along with results from [4].

Proof of Lemma 3. We first note that

$$\begin{aligned}\|\mathcal{F}_{m,n,\epsilon}-\mathcal{F}_\epsilon\| &= \|(\hat{p}-p)\mathcal{I}+(1-\hat{p})\mathcal{G}_{m,n,\epsilon}-(1-p)\mathcal{G}_\epsilon\| \\ &\leq |\hat{p}-p|\,(1+\|\mathcal{G}_\epsilon\|)+\|\mathcal{G}_{m,n,\epsilon}-\mathcal{G}_\epsilon\| \ .\end{aligned}$$

The first summand above is almost surely of order $\mathcal{O}(\sqrt{\log\log k/k})$ in view of (A1) and the fact that $\|\mathcal{G}_\epsilon\|<\infty$. We have that $\|\mathcal{G}_{m,n,\epsilon}-\mathcal{G}_\epsilon\|\leq\|\hat{f}-f\|_{[0,\tau-\epsilon]}\|\mathcal{A}\circ\mathcal{I}_\epsilon\|+\|\mathcal{G}_{m,n,\epsilon}-\hat{f}(\mathcal{A}\circ\mathcal{I}_\epsilon)\|$. Using integration by parts, for t in $[\gamma_{m,n},\infty)$, we may write that

$$\begin{aligned}
&|\mathcal{G}_{m,n,\epsilon}(u)(t)-\hat{f}(t)\,(\mathcal{A}\circ\mathcal{I}_\epsilon)\,(u)(t)| \\
&\leq \hat{f}(t)\left\{\left|\int_{\gamma_{m,n}<y\leq t} y\left(\int_{y\leq z\leq\tau-\epsilon}\frac{u(z)}{z^2}\,dz\right)d\left[\frac{1}{\hat{f}(y)}-\frac{1}{f(y)}\right]\right|+\right. \\
&\qquad\left.\left|\int_{0<y\leq\gamma_{m,n}} y\left(\int_{y\leq z\leq\tau-\epsilon}\frac{u(z)}{z^2}\,dz\right)d\left[\frac{1}{f(y)}\right]\right|\right\} \\
&\leq \hat{f}(t)\left\{\left|\left[\left[\frac{1}{f(y)}-\frac{1}{\hat{f}(y)}\right]y\int_{y\leq z\leq\tau-\epsilon}\frac{u(z)}{z^2}dz\right]_{y=\gamma_{m,n}}^{t}\right|+\right. \\
&\qquad\left.\left|\int_{\gamma_{m,n}<y\leq t}\left[\frac{1}{f(y)}-\frac{1}{\hat{f}(y)}\right]\int_{y\leq z}\frac{u(z)}{z^2}dzdy\right|\right\}+ \\
&\quad \hat{f}(t)\left\{\left|\int_{\gamma_{m,n}<y\leq t}\left[\frac{1}{f(y)}-\frac{1}{\hat{f}(y)}\right]\frac{u(y)}{y}dy\right|+\right. \\
&\qquad\left.\left|\int_{0<y\leq\gamma_{m,n}} y\left(\int_{y\leq z\leq\tau-\epsilon}\frac{u(z)}{z^2}\,dz\right)d\left[\frac{1}{f(y)}\right]\right|\right\} \\
&\leq \|u\|_{[0,\tau-\epsilon]}\left\{2\left\|1-\frac{\hat{f}}{f}\right\|_{[\gamma_{m,n},\infty)}\left[1+\int_{\gamma_{m,n}}^{t}\frac{dy}{y}\right]+\right. \\
&\qquad\left.\frac{\hat{f}(\gamma_{m,n})\,(f(0)-f(\gamma_{m,n}))}{f(0)f(\gamma_{m,n})}\right\} \\
&\leq \|u\|_{[0,\tau-\epsilon]}\left\{\frac{2\|\hat{f}-f\|_{[\gamma_{m,n},\infty)}}{f(\tau-\epsilon)}\left[1+(\log(\tau-\epsilon)-\log(\gamma_{m,n}))\right]+\right. \\
&\qquad\left.\frac{\hat{f}(\gamma_{m,n})F_U(\gamma_{m,n})}{f(\gamma_{m,n})}\right\} \ .
\end{aligned}$$

Similarly, we may show that the inequality

$$|\mathcal{G}_{m,n,\epsilon}(u)(t) - \hat{f}(t)\,(\mathcal{A} \circ \mathcal{I}_\epsilon)\,(u)(t)| \;\leq\; \frac{\|u\|_{[0,\tau-\epsilon]}\hat{f}(\gamma_{m,n})F_U(\gamma_{m,n})}{f(\gamma_{m,n})}$$

holds for t in $[0, \gamma_{m,n})$. In view of part b) of Lemma 1 and the fact that $\|\mathcal{A}\| < \infty$, we conclude that both $\|\mathcal{G}_{m,n,\epsilon} - \mathcal{G}_\epsilon\|$ and $\|\mathcal{F}_{m,n,\epsilon} - \mathcal{F}_\epsilon\|$ are almost surely of order

$$\mathcal{O}\left(\frac{\log(1/\gamma_{m,n})}{\gamma_{m,n}f(\tau-\epsilon)}\sqrt{\frac{\log\log k}{k}} + \frac{F_U(\gamma_{m,n})}{\gamma_{m,n}f(\gamma_{m,n})}\sqrt{\frac{\log\log k}{k}} + F_U(\gamma_{m,n})\right) .$$

Define the operator $\mathcal{M}_\epsilon : D[0,\tau] \to D[0,\tau]$ as

$$\mathcal{M}_\epsilon(u)(t) = (1-p)f(t)\int_{0<y\leq t} y\left(\int_{\tau-\epsilon<z\leq\tau}\frac{u(z)}{z^2}dz\right)d\left[\frac{1}{f(y)}\right]$$

and observe that $\mathcal{F}_\epsilon = \mathcal{F} - \mathcal{M}_\epsilon = \mathcal{F} \circ \left(\mathcal{I} - \mathcal{F}^{-1} \circ \mathcal{M}_\epsilon\right)$, where $\mathcal{F}^{-1}$ exists and has norm $\|\mathcal{F}^{-1}\| \leq 2/p^2$ by Lemma 3 of [5]. It is possible to show that

$$\|\mathcal{M}_\epsilon\| \leq (1-p)\left(\frac{\epsilon}{\tau-\epsilon}\right)$$

and thus, provided $\epsilon < \tau p^2/(p^2 - 2p + 2)$, $\|\mathcal{F}^{-1} \circ \mathcal{M}_\epsilon\| \leq \|\mathcal{F}^{-1}\|\|\mathcal{M}_\epsilon\| < 1$ and $\mathcal{I} - \mathcal{F}^{-1} \circ \mathcal{M}_\epsilon$ is invertible. In such case, $\mathcal{F}_\epsilon$ is invertible with inverse $\mathcal{F}_\epsilon^{-1} = (\mathcal{I} - \mathcal{F}^{-1} \circ \mathcal{M}_\epsilon)^{-1} \circ \mathcal{F}^{-1}$ of norm

$$\|\mathcal{F}_\epsilon^{-1}\| \leq \frac{\|\mathcal{F}^{-1}\|}{1 - \|\mathcal{F}^{-1} \circ \mathcal{M}_\epsilon\|} \leq \frac{2\tau - 2\epsilon}{p^2\tau - (p^2 - 2p + 2)\epsilon} ,$$

the latter bound decreasing monotonically to $2/p^2$ as ϵ goes to zero. Similarly, one can show that, provided $\epsilon < \tau\hat{p}^2/(\hat{p}^2 - 2\hat{p} + 2)$, $\mathcal{F}_{m,n,\epsilon}$ is invertible with inverse $\mathcal{F}_{m,n,\epsilon}^{-1}$ of uniformly bounded norm.

The rate of convergence found above is preserved for the inverse operators since, for some $C > 0$, we have that

$$\left\|\mathcal{F}_{m,n,\epsilon}^{-1} - \mathcal{F}_\epsilon^{-1}\right\| = \left\|\mathcal{F}_{m,n,\epsilon}^{-1} \circ (\mathcal{F}_\epsilon - \mathcal{F}_{m,n,\epsilon}) \circ \mathcal{F}_\epsilon^{-1}\right\| \leq C\,\|\mathcal{F}_{m,n,\epsilon} - \mathcal{F}_\epsilon\|$$

in view of the boundedness of $\mathcal{F}_{m,n,\epsilon}^{-1}$ and the fact that the image of a continuous function under $\mathcal{F}_\epsilon^{-1}$ is a continuous function.

Proof of Lemma 4. Let $s \leq \tau - \epsilon$. Consider the sequence of Bernstein polynomials $P_{d_{m,n}}$ of order $d_{m,n}$ approximating $B_{Y,n}$, that is,

$$P_{d_{m,n}}(x) = \sum_{j=0}^{d_{m,n}} B_{Y,n}\left(\frac{j}{d_{m,n}}\right)\binom{d_{m,n}}{j}x^j(1-x)^{d_{m,n}-j} .$$

We may first write, for $s \geq \gamma_{m,n}$,

$$\int_0^s B_{Y,n}(F(y))d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right]$$

$$= -\int_0^{\gamma_{m,n}} B_{Y,n}(F(y))d\left[\frac{1}{f(y)}\right] + \int_{\gamma_{m,n}}^s B_{Y,n}(F(y))d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right].$$

By the MVT, setting $L = \|g\|_{[0,\tau]} + \|f\|_{[0,\tau]} < \infty$, we have that

$$\begin{aligned}
&\left|\hat{f}(s)\int_0^{\gamma_{m,n}} B_{Y,n}(F(y))d\left[\frac{1}{f(y)}\right]\right| \\
&\qquad \leq \hat{f}(s) \sup_{0\leq y\leq \gamma_{m,n}} |B_{Y,n}(F(y))| \left[\frac{f(0)-f(\gamma_{m,n})}{f(0)f(\gamma_{m,n})}\right] \\
&\qquad \leq \hat{f}(\gamma_{m,n}) \sup_{0\leq u\leq L\gamma_{m,n}} |B_{Y,n}(u)| F_U(\gamma_{m,n})/f(\gamma_{m,n}) \\
&\qquad = \mathcal{O}\left(\frac{\hat{f}(\gamma_{m,n})F_U(\gamma_{m,n})}{f(\gamma_{m,n})}\sqrt{\gamma_{m,n}\log(1/\gamma_{m,n})}\right) \ a.s.
\end{aligned}$$

Defining $\Delta_{m,n} = (B_{Y,n} - P_{d_{m,n}}) \circ F$, we may write

$$\int_{\gamma_{m,n}}^s B_{Y,n}(F(y))d\left[1/\hat{f}(y) - 1/f(y)\right]$$

as

$$\int_{\gamma_{m,n}}^s \Delta_{m,n}(y)d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right] + \int_{\gamma_{m,n}}^s P_{d_{m,n}}(F(y))d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right].$$

We may easily show that

$$\left|\hat{f}(s)\int_{\gamma_{m,n}}^s \Delta_{m,n}(y)d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right]\right| \leq 2\|\Delta_{m,n}\|_\infty\left[1 + \frac{\|\hat{f}-f\|_{[\gamma_{m,n},\infty)}}{f(\tau-\epsilon)}\right].$$

In view of Theorem 1.6.1 of [3], denoting by $\Omega(\phi,\delta)$ the modulus of continuity of ϕ with respect to bandwidth $\delta > 0$, we have that

$$\|\Delta_{m,n}\|_\infty \leq \frac{5}{4}\cdot\Omega\left(B_{Y,n}, \frac{1}{\sqrt{d_{m,n}}}\right) = \mathcal{O}\left(\sqrt{\frac{\log d_{m,n}}{\sqrt{d_{m,n}}}}\right) \ a.s.$$

We may then use integration by parts to show that

$$\left|\hat{f}(s)\int_{\gamma_{m,n}}^s P_{d_{m,n}}(F(y))d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right]\right|$$

$$\leq \frac{\|\hat{f}-f\|_{[\gamma_{m,n},\infty)}}{f(\tau-\epsilon)}\left\{2 \sup_{\gamma_{m,n}\leq y\leq \tau-\epsilon} |P_{d_{m,n}}(F(y))| \right.$$
$$\left. + \int_{\gamma_{m,n}}^{\tau-\epsilon} \left| P'_{d_{m,n}}(F(y)) \right| f(y)dy \right\}$$
$$\leq \frac{\|\hat{f}-f\|_{[\gamma_{m,n},\infty)}}{f(\tau-\epsilon)}\left\{2\left[\|\Delta_{m,n}\|_\infty + \|B_{Y,n}\|_{[0,1]}\right] \right.$$
$$\left. + \int_{\gamma_{m,n}}^{\tau-\epsilon} \left| P'_{d_{m,n}}(F(y)) \right| f(y)dy \right\}$$

Inspecting the proof of Theorem 2.1 of [2], we have that

$$\left| P'_{d_{m,n}}(F(t)) \right| \;\leq\; 2\cdot\Omega\left(B_{Y,n}, \sqrt{\frac{F(t)(1-F(t))}{d_{m,n}}}\right)\sqrt{\frac{d_{m,n}}{F(t)(1-F(t))}}\ ,$$

from which some algebraic manipulations yield that

$$\int_{\gamma_{m,n}}^{\tau-\epsilon} \left| P'_{d_{m,n}}(F(y)) \right| f(y)dy$$
$$= \mathcal{O}\left(d_{m,n}^{1/4}\left[\sqrt{\log d_{m,n}} + \sqrt{\log\left(1/\gamma_{m,n}\right)} + \sqrt{\log\left(1/\epsilon\right)}\right]\right) \quad a.s.$$

Some calculations indicate that the choice $d_{m,n} = k/\sqrt{\log\log k}$ leads to the upper bound for $\hat{f}(s)\left|\int_{\gamma_{m,n}}^{s} B_{Y,n}(F(y))d\left[1/\hat{f}(y) - 1/f(y)\right]\right|$ of least order, provided we have that $d_{m,n}\gamma_{m,n}\to\infty$: this optimal order is

$$\mathcal{O}\left(k^{-\frac{1}{4}}\log k/f(\tau-\epsilon)\right)\ .$$

Proof of Lemma 5. Define the following terms:

$$\mathcal{J}_1(s) = \sqrt{\hat{p}}\,|W_{X,m}(s) - B_{X,m}(G(s))|\ ,\ \mathcal{J}_2(s) = \left|\sqrt{\hat{p}} - \sqrt{p}\ \right| |B_{X,m}(G(s))|\ ,$$

$$\mathcal{J}_3(s) = \sqrt{1-\hat{p}}\hat{f}(s)\int_{0<y\leq s} |W_{Y,n}(y) - B_{Y,n}(F(y))|\, d\left[\frac{1}{\hat{f}(y)}\right]\ ,$$
$$\mathcal{J}_4(s) = \left|\left(\sqrt{1-\hat{p}} - \sqrt{1-p}\right)\hat{f}(s)\int_0^s B_{Y,n}(F(y))d\left[\frac{1}{\hat{f}(y)}\right]\right|\ ,$$

$$\mathcal{J}_5(s) = \sqrt{1-p}\left|\hat{f}(s)\int_0^s B_{Y,n}(F(y))d\left[\frac{1}{\hat{f}(y)} - \frac{1}{f(y)}\right]\right|$$

$$\text{and} \quad \mathcal{J}_6(s) = \sqrt{1-p}\left|\left(\hat{f}(s) - f(s)\right)\int_0^s B_{Y,n}(F(y))d\left[\frac{1}{f(y)}\right]\right| .$$

Define further $\mathcal{I}_r = \|\mathcal{J}_r\|_{[0,\tau-\epsilon]}$ for $r = 1, ..., 6$, and note that

$$\left\|W_{m,n} - W^0_{m,n}\right\|_{[0,\tau-\epsilon]} \le \sum_{r=1}^{6} \mathcal{I}_r .$$

From KMT, we have that both $\mathcal{I}_1$ and $\mathcal{I}_3$ are $\mathcal{O}\left(\log k/\sqrt{k}\right)$ almost surely. Using Lemma 1.4.1 of [1], (A2) and Lemma 2, we have that both $\mathcal{I}_2$ and $\mathcal{I}_4$ are $\mathcal{O}\left(\sqrt{\log k \log\log k/k}\right)$ almost surely. In view of Lemma 4, we have that

$$\mathcal{I}_5 = \mathcal{O}\left(\frac{k^{-\frac{1}{4}}\sqrt{\log k}(\log\log k)^{\frac{1}{4}}}{f(\tau-\epsilon)}\right) \quad a.s.$$

Further, we find that $\mathcal{I}_6 \le \|\hat{f} - f\|_{[0,\infty)} \sup_{0\le t\le 1}|\mathcal{W}_{Y,n}(t)|/f(\tau-\epsilon)$, which implies that

$$\mathcal{I}_6 = \mathcal{O}\left(\left[\gamma_{m,n}^{-1}\sqrt{\frac{\log\log k}{k}} + F_U(\gamma_{m,n})\right]\frac{\sqrt{\log k}}{f(\tau-\epsilon)}\right) = \mathcal{O}\left(\frac{k^{-\frac{\alpha-1}{2\alpha}}\sqrt{\log k}}{f(\tau-\epsilon)}\right)$$

almost surely. It is clear then, in view of the above, that $\mathcal{I}_6$ dominates for $\alpha \in (1,2)$, while $\mathcal{I}_5$ dominates for $\alpha \in [2,\infty)$.

Proof of Lemma 6. The result is a consequence of Lemma 1.4.1 of [1], Lemma 2 and (3.2).

Proof of Lemma 7. By Theorem 1, there exists a sequence of Gaussian processes $U^0_{m,n}$ such that, as $k \to \infty$,

$$\sup_{0\le s\le\tau-\eta}\left|U_{m,n}(s) - U^0_{m,n}(s)\right| = \mathcal{O}\left(\epsilon_{m,n}(\log k)^{\frac{3}{2}}\sqrt{\log\log k}\right) \quad a.s.$$

For any $s \in [0, \tau-\eta]$, we may use integration by parts to show that

$$\tilde{g}_{m,n}(s) - g_{m,n}(s) = \frac{1}{h_{m,n}}\int_0^\infty K\left(\frac{s-x}{h_{m,n}}\right)d\left[\hat{G}(x) - G(x)\right]$$

$$
\text{(S.1)} \qquad = B_{m,n}(s) + \mathcal{O}\left(\frac{\epsilon_{m,n}(\log k)^{\frac{3}{2}}\sqrt{\log\log k}}{\sqrt{k}h_{m,n}}\right) \quad a.s. \ ,
$$

where $B_{m,n}(s) = (\sqrt{k}h_{m,n})^{-1}\int_{-1}^{1} U^0_{m,n}(s - uh_{m,n})dK(u)$. We notice that, for large m and n,

$$
\sup_{0\le s\le\tau-\eta}\ \sup_{-1\le u\le 1}\left|U^0_{m,n}(s-uh_{m,n}) - U^0_{m,n}(s)\right|
$$

$$
\le \sup_{0\le x\le\tau-\eta}\ \sup_{0\le y\le h_{m,n}}\left|U^0_{m,n}(x+y) - U^0_{m,n}(x)\right| \ ,
$$

and thus, using Theorem 2 and (K1), we obtain that

(S.2)

$$
\limsup_{m,n\to\infty}\ \sup_{0\le s\le\tau-\eta}|B_{m,n}(s)| \le \limsup_{m,n\to\infty}\left\{\frac{\sqrt{h_{m,n}\log(1/h_{m,n})}}{\sqrt{k}h_{m,n}}V_K\right\} = 0 \quad a.s.
$$

The result follows from (S.1), (S.2) and (A6).

References.

[1] M. Csörgö and P. Révész. *Strong Approximations in Probability and Statistics*. Academic Press, 1981.

[2] Z. Ditzian. Derivatives of Bernstein polynomials and smoothness. *Proceedings of the American Mathematical Society*, 93(1):25–31, 1985.

[3] G.G. Lorentz. *Bernstein polynomials*. University of Toronto Press, Toronto, 1953.

[4] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, pages 1269–1283, 1990.

[5] Y. Vardi and C.H. Zhang. Large sample study of empirical distributions in a random-multiplicative censoring model. *The Annals of Statistics*, pages 1022–1039, 1992.