Statistics 583, Problem Set 5 - revised, Solution

Wellner; 5/2715

1. The expression for the jackknife variance estimator for the median, in the display (1) on page 11 (3rd line from the bottom) in chapter 8 was derived under the assumption $n = 2m$ and that $T(\mathbb{F}_n) = X_{(m)}$ if $n = 2m - 1$, $T(\mathbb{F}_n) = (X_{(m)} + X_{(m+1)})/2$ if $n = 2m$.
(a) Derive the first equality in (1), page 11, using this definition of the sample median.
(b) Derive versions of the development in (1), page 11, using $T(F) = F^{-1}(1/2)$ (strictly). Does the asymptotic result in (1) still hold? Here is some further explanation of what I mean by "strictly" here: let $T_1(\mathbb{F}_n) = X_m$ if $n = 2m-1$, $T_1(\mathbb{F}_n) = (X_{(m)} + X_{(m+1)})/2$ if $n = 2m$. This is one common definition of the median, and this is the definition used in (a). Let $T_2(\mathbb{F}_n) = \mathbb{F}_n^{-1}(1/2)$. This is my favorite definition of the median. Note that $T_2(\mathbb{F}_n) = T_1(\mathbb{F}_n)$ if $n = 2m-1$, but $T_2(\mathbb{F}_n) \neq T_1(\mathbb{F}_n)$ if $n = 2m$. (What is the value of $T_2(\mathbb{F}_n)$ in this case?) T_2 is the definition of the median to be considered in 2(b)!

Solution: (a). For $n = 2m$,

$$T_{n,i} = \begin{cases} X_{(m+1)} & \text{if} \quad i \le m \\ X_{(m)} & \text{if} \quad i > m \end{cases}$$

and $T_{n,\cdot} = (X_{(m)} + X_{(m+1)})/2$. Hence

$$\begin{aligned} n\widehat{\mathrm{Var}}_n &= (n-1)\left\{ m(X_{(m+1)} - \frac{1}{2}(X_{(m)} + X_{(m+1)}))^2 \right. \\ &\quad + \left. m(X_{(m)} - \frac{1}{2}(X_{(m)} + X_{(m+1)}))^2 \right\} \\ &= n(n-1)\{\frac{X_{(m+1)} - X_{(m)}}{2}\}^2 . \qquad (1) \end{aligned}$$

(b). When $n = 2m$ and $T(F) = F^{-1}(1/2)$, we have $T(\mathbb{F}_n) = X_{(m)}$ and $T_{n,i}$ are exactly as in (a) above. Hence (1) continues to hold.
When $n = 2m - 1$, then $T(\mathbb{F}_n) = X_{(m)}$,

$$T_{n,i} = \begin{cases} X_{(m)} & \text{if} \quad i \le m-1 \\ X_{(m-1)} & \text{if} \quad i \ge m \end{cases} ,$$

and $T_{n,\cdot} = \{(m-1)X_{(m)} + mX_{(m-1)}\}/(2m-1)$. Therefore

$$\begin{aligned} n\widehat{\mathrm{Var}}_n &= (n-1)\Bigg\{(m-1)\{X_{(m)} - \frac{1}{2m-1}[(m-1)X_{(m)} + mX_{(m-1)}]\}^2 \\ &\quad + m\{X_{(m-1)} - \frac{1}{2m-1}[(m-1)X_{(m)} + mX_{(m-1)}]\}^2\Bigg\} \\ &= \frac{(n-1)^2(n+1)}{n}\left\{\frac{X_{(m)} - X_{(m-1)}}{2}\right\}^2 \\ &\to_d \frac{1}{4f^2(F^{-1}(1/2))}\left(\frac{\chi_2^2}{2}\right)^2 \end{aligned}$$

just as before.

Remark: The only case left out in (a) and (b) is that of an odd sample size, $n = 2m-1$ in part (a). In this case,

$$T_{n,i} = \begin{cases} (X_{(m)} + X_{(m+1)})/2 & \text{if } i \le m-1 \\ (X_{(m-1)} + X_{(m+1)})/2 & \text{if } i = m \\ (X_{(m-1)} + X_{(m)})/2 & \text{if } i \ge m+1 \end{cases}.$$

Thus

$$\begin{aligned} T_{n,\cdot} &= \frac{1}{n}\Bigg\{\frac{(m-1)}{2}(X_{(m)} + X_{(m+1)}) \\ &\quad + \frac{1}{2}(X_{(m-1)} + X_{(m+1)}) + \frac{(m-1)}{2}(X_{(m-1)} + X_{(m)})\Bigg\}. \end{aligned}$$

The analysis from this point proceeds not just by algebra, but by careful grouping of terms and observing which terms are negligible. I will not present a full analysis here, but will record the result:

$$\begin{aligned} n\widehat{\mathrm{Var}}_n &= \frac{(m-1)m^2}{2n^3}\left\{n(X_{(m+1)} - X_{(m-1)})\right\}^2 + o_p(1) \\ &\to_d \frac{1}{4f^2(F^{-1}(1/2))}(\frac{\chi_4^2}{4})^2 \end{aligned}$$

since, with $g \equiv F^{-1}$,

$$n(X_{(m+1)} - X_{(m-1)}) \to_d g'(1/2)W$$

where $W =_d Y_1 + Y_2 \sim \mathrm{Gamma}(2,1)$ for independent exponential rv's $Y_1,\ Y_2$, so that $2W \sim \chi_4^2$. Thus for this definition of the sample median, it is true that $n\widehat{\mathrm{Var}}_n = O_p(1)$ for the full sequence of nonnegative integers n but it converges in distribution to one limit as $n = 2m \to \infty$ and a different limit as $n = 2m-1 \to \infty$.

2. (a) Wasserman, problem 3.8.3, page 39, modified. Show that the claimed expression for v_{boot} given in the display for this problem is incorrect and find the correct expression. Here $v_{boot} = Var_{\mathbb{F}_n}(T_n)$ where $T_n = \overline{X}_n^2$. [Hint: see Dodd and Korn, *The American Statistician* **61** (2007), 127 - 131, and especially their appendix B, pages 130-131. Apparently the formula given by Wasserman in his problem is from Shao and Tu (1995), page 10; as noted by Dodd and Korn, the expression in Shao and Tu is incorrect.]
(b) Explain how the resulting formulas relate to how you would estimate the variance of $\overline{X}_n^2$ via the delta method.

Solution: (a) This is explained quite well in the appendix of the paper by Dodd and Korn (2007).
(b) The first term of the exact finite sample variance expression

$$Var(\overline{X}^2) = \frac{4\mu^2\sigma^2}{n} + \frac{2\sigma^4}{n^2} + \frac{4\mu\mu_3}{n^2} + \frac{\mu_4 - 3\sigma^4}{n^3}$$

corresponds exactly to what we would get from the delta method: with $g(x) = x^2$ we have $g'(x) = 2x$ and hence

$$\sqrt{n}(\overline{X}_n^2 - \mu^2) \to_d g'(\mu)\sigma Z \sim N(0, 4\mu^2\sigma^2)$$

where $Z \sim N(0,1)$. Thus the delta-method estimator of $Var(\overline{X}_n^2)$ is just $4\overline{X}_n^2 S_n^2$ where S_n is the sample variance. The bootstrap estimator of variance refines this (as shown by Dodd and Korn) by correctly capturing the n^{-2} term when $\mu \neq 0$. When $\mu = 0$, then neither the (first order) delta method nor the (nonparametric) bootstrap tells the complete story.

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. with density p on $\mathbb{R}$. The kernel density estimator $\hat{p}_n$ of p using the kernel k and bandwidth b is defined by

$$\hat{p}_n(x;b) = \frac{1}{b}\int k\left(\frac{x-y}{b}\right) d\mathbb{F}_n(y) = \frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{x-X_i}{b}\right).$$

We assume that $\int k(v)dv = 1$ and $k(v) \geq 0$.
(a) For a fixed point $x \in \mathbb{R}$ calculate $E\hat{p}_n(x;b)$ and give an expression for the bias, $E\hat{p}_n(x;b) - p(x)$.
(b) Again for a fixed x, calculate $Var(\hat{p}_n(x;b))$.
(c) Discuss how the bias and variance change as the bandwidth b increases.
(d) Assuming that $\int vk(v)dv = 0$ and $\int v^2 k(v)dv = 1$, compute

$$\hat{\mu}_n = \int x\hat{p}_n(x)dx \quad \text{and} \quad \hat{\sigma}_n^2 \equiv \int (x - \hat{\mu}_n)^2 \hat{p}_n(x)dx.$$

How do $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ compare to $\overline{X}_n$ and S_n^2?

Solution: (a) By a straightforward application of the Fubini-Tonelli theorem

$$\begin{aligned} E\left\{\hat{p}_n(x;b)\right\} &= E\left\{\frac{1}{nb}\sum_{i=1}^{n} k\left(\frac{x-X_i}{b}\right)\right\} \\ &= \frac{1}{nb}\sum_{i=1}^{n} E\left\{k\left(\frac{x-X_i}{b}\right)\right\} \\ &= \frac{1}{b}\int k\left(\frac{x-y}{b}\right) dF(y) \\ &= \frac{1}{b}\int k\left(\frac{x-y}{b}\right) p(y)dy, \end{aligned}$$

and thus the bias is, by the change of variables $z \equiv (x-y)/b$

$$\begin{aligned} E\left\{\hat{p}_n(x;b)\right\} - p(x) &= \int \frac{1}{b}k\left(\frac{x-y}{b}\right) p(y)dy - p(x) \\ &= \int k(z)\left\{p(x-bz)-p(x)\right\} dz. \end{aligned}$$

(b) To calculate the variance, note that by independence it follows that

$$\begin{aligned} Var(\hat{p}_n(x;b)) &= \frac{1}{(nb)^2}\sum_{i=1}^{n} Var\left(k\left(\frac{x-X_i}{b}\right)\right) \\ &= \frac{1}{nb^2}Var\left(k\left(\frac{x-X}{b}\right)\right) \\ &= \frac{1}{nb^2}\left\{Ek^2\left(\frac{x-X}{b}\right) - [Ek\left(\frac{x-X}{b}\right)]^2\right\} \\ &\leq \frac{1}{nb^2}Ek^2\left(\frac{x-X}{b}\right) = \frac{1}{nb}\int k^2(z)p(x-bz)dz. \end{aligned}$$

(c) It is clear from (a) and (b) that the bias increases as b increases, while the variance decreases as b increases.

(d) Note that by the Fubini-Tonelli theorem

$$\begin{aligned}
\hat{\mu}_n &= \int x\hat{p}_n(x)dx = \int x\frac{1}{b}k\left(\frac{x-y}{b}\right)d\mathbb{F}_n(y)dx \\
&= \int\left\{\int x\frac{1}{b}k\left(\frac{x-y}{b}\right)dx\right\}d\mathbb{F}_n(y) \\
&= \int\left\{\int\left\{\frac{x-y}{b}k\left(\frac{x-y}{b}\right)+\frac{y}{b}k\left(\frac{x-y}{b}\right)\right\}dx\right\}d\mathbb{F}_n(y) \\
&= \int\left\{b\int vk(v)dv + y\int\frac{1}{b}k(\frac{x-y}{b}dx\right\}d\mathbb{F}_n(y) \\
&= \int\{b\cdot 0+y\}\,d\mathbb{F}_n(y) = \overline{X}_n.
\end{aligned}$$

Also,

$$\hat{\sigma}_n^2 \equiv \int(x-\hat{\mu}_n)^2\hat{p}_n(x)dx = \int x^2\hat{p}_n(x,b)dx - \hat{\mu}_n^2$$

where

$$\begin{aligned}
\int x^2\hat{p}_n(x,b)dx &= \int x^2\int\frac{1}{b}k\left(\frac{x-y}{b}\right)d\mathbb{F}_n(y)dx \\
&= \int\left\{\int x^2\frac{1}{b}k\left(\frac{x-y}{b}\right)dx\right\}d\mathbb{F}_n(y) \\
&= \int\left\{\int(x-y+y)^2\frac{1}{b}k\left(\frac{x-y}{b}\right)dx\right\}d\mathbb{F}_n(y) \\
&= \int\left\{\int\{(x-y)^2+2(x-y)y+y^2\}\frac{1}{b}k\left(\frac{x-y}{b}\right)dx\right\}d\mathbb{F}_n(y) \\
&= \int\left\{b^2\int v^2k(v)dv+0+y^2\right\}d\mathbb{F}_n(y) \\
&= b^2+\overline{X^2}_n.
\end{aligned}$$

We conclude that $\hat{\sigma}_n^2 = b^2 + S_n^2$. Thus the mean according to $\hat{p}_n(\cdot, b)$ agrees with the empirical mean $\overline{X}_n$, but the variance according to $\hat{p}_n(\cdot, b)$ exceeds the sample variance S_n^2 by b^2. Without worrying about the increase in variance we can accomplish (bootstrap) sampling from $\hat{p}_n(\cdot, b)$ based on a standard normal kernel $k = \phi$ by letting $X_1^*, \ldots, X_n^*$ be i.i.d. $\mathbb{F}_n$, letting $\epsilon_1, \ldots, \epsilon_n$ be i.i.d. $N(0,1)$ and independent of the X_i^*'s, and representing a bootstrap sample $Y_1^*, \ldots, Y_n^*$ from $\hat{p}_n(\cdot, b)$ as $Y_i^* = X_i^* + b\epsilon_i$, $i = 1, \ldots, n$.

4. Silverman (1981, 1983; see the course handout page for copies of these two papers) proposed a test for the number of modes of a density. This test is discussed on

pages 227-232 of Efron and Tibshirani (1998), *An Introduction to the Bootstrap*. Silverman's proposed test goes as follows: let $\hat{p}_n(\cdot; b)$ be the kernel estimator of p based on a standard Gaussian kernel k; i.e. $k(v) = \phi(v)$ where ϕ is the standard Gaussian density $\phi(v) = (2\pi)^{-1/2} \exp(-v^2/2)$. [Use of a Gaussian kernel is crucial in Silverman's test!] As b increases the density estimator $\hat{p}_n(\cdot; b)$ becomes smoother and has fewer modes. In fact for a Gaussian kernel, the number of modes is a monotone non-increasing function of the bandwidth b. See Figure 16.2 on page 228 of Efron and Tibshirani (1998) for an illustration of this. Consider testing $H_0 :\ p$ has one mode versus $H_1 :\ p$ has 2 or more modes. Since the number of modes decreases as b increases, there is a smallest value of b such that $\hat{p}(\cdot; b)$ has one mode. Call this $\hat{b}_1$. Now we use $\hat{p}_n(\cdot; \hat{b}_1)$ as the estimated null distribution for our test of H_0 versus H_1. As noted in Efron and Tibshirani, it seems reasonable to adjust $\hat{p}_n(\cdot; \hat{b}_1)$ slightly to adjust for the fact noted in problem 3(d) above that the variance under $\hat{p}_n(\cdot; \hat{b}_1)$ is somewhat larger than the sample variance. We call the resulting estimator $\hat{q}_n(\cdot; \hat{b}_1)$. A reasonable test statistic is $\hat{b}_1$: if this is large, then a greater amount of smoothing is required to obtain one mode, and this supports the alternative hypothesis. Now the test is carried out via bootstrap resampling from the fitted model under the null hypothesis; see Efron and Tibshirani (1998) for details.
(a) Describe this bootstrap testing procedure from the perspective of estimation of some functional of the true distribution and our discussion in sections 8.2 and 8.3, distinguishing carefully between the ideal bootstrap and the Monte-Carlo implementation of the bootstrap.
(b) Verify that the resampling scheme outlined on page 232 of Efron and Tibshirani accomplishes the desired adjustment of $\hat{p}_n(\cdot; \hat{b}_1)$ so that the resulting $\hat{q}_n(\cdot; \hat{b}_1)$ has variance very nearly equal to the sample variance.
(c) Find at least one alternative test of multimodality of a univariate density p that has been proposed since (1983).

Solution: (a) The basic test rule is to reject H_0 if

$$\hat{b}_1 = \inf\{b > 0 :\ \hat{p}_n(\cdot, b) \text{ has one mode}\}.$$

In this case the "ideal model-based bootstrap is to choose $c_\alpha(\underline{X})$ so that $P_{\hat{p}_n(\cdot,\hat{b}_1}(\hat{b}_1 > c_\alpha) = \alpha$ for a given type one error $\alpha \in (0, 1/2)$. To implement this via a Monte-Carlo sampling scheme we proceed by sampling from $\hat{p}_n(\cdot, \hat{b}_1)$ as follows: let $X^*_{1,j}, \ldots, X^*_{n,j}$ be i.i.d. $\hat{p}_n(\cdot, \hat{b}_1)$ for $j = 1, \ldots, B$, and calculate $\hat{b}^*_{1,j} \equiv \hat{b}_1(\mathbb{F}^*_{n,j})$, $j = 1, \ldots, B$, and choose $\hat{c}_\alpha$ so that

$$\frac{1}{B} \sum_{j=1}^{B} 1\{\hat{b}^*_{1,j} > \hat{c}^*_\alpha\} = \alpha.$$

Then we reject H_0 if $\hat{b}_1 > \hat{c}^*_\alpha$.
(b) If $X_1^*, \ldots, X_n^*$ are i.i.d. $\mathbb{F}_n$ and $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. $N(0,1)$ and independent of the X_i^*'s, we now define

$$Z_i^* = \overline{X}_n^* + (1 + \hat{b}_1^2/S_n^2)^{-1/2}(X_i^* - \overline{X}_n^* + \hat{b}_1\epsilon_i)$$

for $i = 1, \ldots, n$. Then we compute

$$\begin{aligned} E(Z_i^*|\underline{X}) &= E(\overline{X}_n^*|\underline{X}) + \frac{1}{(1+\hat{b}_1^2/S_n^2)^{1/2}} E\left(X_i^* - \overline{X}_n^*|\underline{X}\right) \\ &= \overline{X}_n + 0 = \overline{X}_n. \end{aligned}$$

Furthermore,

$$\begin{aligned} Var(Z_i^*) &= Var(\overline{X}_n^*|\underline{X}) + \frac{1}{1+\hat{b}_1^2/S_n^2}\left\{Var(X_i^* - \overline{X}_n^*|\underline{X}) + \hat{b}_1^2\right\} \\ &= \frac{S_n^2}{n} + \frac{S_n^2}{S_n^2 + \hat{b}_1^2}(S_n^2 + \hat{b}_1^2) \\ &= \frac{n+1}{n} S_n^2 \approx S_n^2. \end{aligned}$$

(c) For a different univariate test, see:
• Hartigan, J. A. and Hartigan, P. M. (1985). The DIP test of unimodality. *Ann. Statist.* **13**, 70 - 84.
For a different approach based on log-concavity and mixing, see
• Walther, Guenther (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97**, 508513;
• Walther, Guenther (2001). Multiscale maximum likelihood analysis of a semi-parametric model, with applications. *Ann. Statist.* **29**, 12971319.
For multivariate mode tests, see:
• Rozl, Gregory Paul M.; Hartigan, J. A. (1994). The MAP test for multimodality. *J. Classification* **11**, 536.

5. **Optional bonus problem:** Consider a two-sample testing problem with $X_1, \ldots, X_m$ i.i.d. F and $Y_1, \ldots, Y_n$ i.i.d. G. Consider testing:
(1) $H : F = G$ versus $K_{Gaussian,1}$: $F(x) = \Phi((x-\mu)/\sigma)$, $G(y) = \Phi((y-\nu)/\sigma)$, $\mu \neq \nu$.
(2) $H : F = G$ versus $K_{Gaussian,2}$: $F(x) = \Phi((x-\mu)/\sigma)$, $G(y) = \Phi((y-\nu)/\tau)$, $\mu \neq \nu$, $\sigma \neq \tau$.
(3) $H : \mu_F = \mu_G$, F, G otherwise unknown, versus $K_{Gaussian,2}$ as in (2).

(a) Discuss appropriate bootstrap testing procedures for these three testing problems, including identification of which (estimated) distribution is involved for the resampling procedure.
(b) In which problems is a permutation test appropriate?
(c) Which of the three problems involves the largest null hypothesis?
Hint: see Efron and Tibshirani (1998), sections 16.1-16.3.

6. **Optional bonus problem:** (Hard!) On page 12, line 4 of Chapter 8 of the lecture notes, it is claimed that if $E_F|X|^r < \infty$ for some $r > 0$ and $f(F^{-1}(1/2)) > 0$, then for the median function $T(F) = F^{-1}(1/2)$ we have

$$nVar_F(T(\mathbb{F}_n)) \to \frac{1/4}{f^2(F^{-1}(1/2))}.$$

Prove (or disprove) this claim.

Solution: More general versions of this problem are discussed in Section 11.4.7, pages 474 - 480, Shorack and Wellner (1986). Unfortunately this material in SW (1986) is based to a substantial extent on an unpublished Stanford Ph.D. dissertation of K. M. Anderson (1982), which is not generally available, and the proofs in SW (1986) are incomplete. In particular, Theorem 4 and Exercise 2 on pages 475 - 476 are not proved there: see the last line of Section 11.4.7, page 480. The solution below relies on inequalities derived in Wellner (1977), *Ann. Statist.* **5**, 481-494.

Suppose that $n = 2m$, so that $T(\mathbb{F}_n) = X_{(m)}$. Since

$$\begin{aligned} Var(T(\mathbb{F}_n)) &= E(X_{(m)} - EX_{(m)})^2 \\ &= E(X_{(m)} - F^{-1}(1/2))^2 - (F^{-1}(1/2) - EX_{(m)})^2, \end{aligned}$$

it suffices to show that

$$\begin{aligned} &nE(X_{(m)} - F^{-1}(1/2))^2 \to \frac{1/4}{f^2(F^{-1}(1/2))}, \qquad \text{and} \\ &n^{1/2}\left\{EX_{(m)} - F^{-1}(1/2)\right\} \to 0. \end{aligned}$$

Since we know that $\sqrt{n}(X_{(m)} - F^{-1}(1/2)) \to_d N(0, 1/(4f^2(F^{-1}(1/2)))$, both of these will follow if we show that the sequence of random variables $\{V_n\} \equiv \{n^{1/2}|X_{(m)} - F^{-1}(1/2)|\}_{n\geq 1}$ is uniformly square integrable; i.e. if we show that

$$\limsup_{n\to\infty} E|V_n|^2 1_{[|V_n|\geq\lambda]} \to 0 \qquad \text{as } \lambda \to \infty.$$

This would follow easily from

$$\limsup_{n\to\infty} E|V_n|^s < \infty \tag{2}$$

for any $s > 2$. Now it follows from Wellner (1977) (see also Shorack and Wellner (1986), page 456) that, with $p_m = m/(2n+1)$,

$$E|\xi_{(m)} - p_m|^s \le C_s (p_m q_m / n)^{s/2} \le C_s (m/n^2)^{s/2} = C_s \left(\frac{1}{2n}\right)^{s/2},$$

where $C_s = 1 + 2 \cdot 10^s \Gamma(s+1)$. Equivalently,

$$E\{|n^{1/2}(\xi_{(m)} - p_m)|^s\} \le C_s (1/2)^{s/2}. \tag{3}$$

Now write $Q = F^{-1}$ and, with $\xi_1, \ldots, \xi_n$ i.i.d. Uniform$(0,1)$,

$$\begin{aligned} V_n &= \sqrt{n}(X_{(m)} - F^{-1}(1/2)) \stackrel{d}{=} \sqrt{n}(Q(\xi_{(m)}) - Q(1/2)) \\ &= Q'(\xi^*)\sqrt{n}(\xi_{(m)} - 1/2) \\ &\equiv Q'(\xi^*) U_n \end{aligned}$$

where $|\xi^* - 1/2| \le |\xi_{(m)} - 1/2|$ since the derivative Q' exists and is continuous in a neighborhood of $1/2$. Let $s > 2$, $0 < M < \infty$, and $\epsilon > 0$. Then we have

$$\begin{aligned} E|V_n|^s &= E|V_n|^s 1\{|V_n| \le M\} + E|V_n|^s 1\{|V_n| > M\} \\ &\le M^s + E|V_n|^s 1\{|V_n| > M\} \equiv I + II_n. \end{aligned}$$

Now we bound II_n in several steps:

$$\begin{aligned} II_n &= E|V_n|^s 1\{|V_n| > M\} \\ &\le E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| \le M\}\} + E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| > M\}\} \\ &\le \left(M \sup_{u:|u-1/2| \le M/\sqrt{n}} |Q'(u)|\right)^s + E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| > M\}\} \\ &\equiv III_n + IV_n \end{aligned}$$

since on $\{|U_n| \le M\}$,

$$|V_n| \le |Q'(\xi^*)||U_n| \le M|Q'(\xi^*)| \le M \sup_{u:|u-1/2| \le M/\sqrt{n}} |Q'(u)|.$$

Now we bound IV_n in two steps: for n so large that $\epsilon\sqrt{n} > M$,

$$
\begin{aligned}
IV_n &= E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| > M\}\} \\
&= E\{|V_n|^s 1\{|V_n| > M\} 1\{M < |U_n| \le \epsilon\sqrt{n}\}\} + E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| > \epsilon\sqrt{n}\}\} \\
&\le \left(\sup_{u:|u-1/2|\le\epsilon} |Q'(u)| \right)^s E|U_n|^s 1\{|U_n| > M\} + E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| > \epsilon\sqrt{n}\}\} \\
&\le \left(\sup_{u:|u-1/2|\le\epsilon} |Q'(u)| \right)^s E|U_n|^s + E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| > \epsilon\sqrt{n}\}\} \\
&\le \left(\sup_{u:|u-1/2|\le\epsilon} |Q'(u)| \right)^s \cdot C_s (1/2)^{s/2} + \ E\{|V_n|^s 1\{|V_n| > M\} 1\{|U_n| > \epsilon\sqrt{n}\}\} \\
&\equiv V_n + VI_n.
\end{aligned}
$$

by using Wellner's moment bound (**??**). Thus it remains to bound VI_n. Since $E|X|^r < \infty$, we have $P(|X| \ge x) \le E|X|^r / x^r$, so

$$
x^r \{F(-x) + (1 - F(x)\} \le x^r \{F(-x) + (1 - F(x-))\} \le E|X|^r \equiv K.
$$

It follows that $|x|^r F(x) \le K$ and $|x|^r (1 - F(x)) \le K$, or equivalently,

$$
|Q(u)| \le K_r \{u(1-u)\}^{-1/r}, \qquad 0 < u < 1.
$$

Since $|V_n| \le \sqrt{n}(|Q(\xi_{(m)})| + |Q(1/2)|)$, it follows by the C_r−inequality that

$$
|V_n|^s \le 2^{s-1} n^{s/2} \left\{ K_r^s [\xi_{(m)}(1 - \xi_{(m)})]^{-s/r} + |Q(1/2)|^s \right\},
$$

and hence

$$
\begin{aligned}
VI_n &\le 2^{s-1} n^{s/2} K_r^s E \left\{ [\xi_{(m)}(1 - \xi_{(m)})]^{-s/r} 1\{|\xi_{(m)} - 1/2| > \epsilon\} \right\} \\
&\qquad + 2^{s-1} Q(1/2)^s n^{s/2} P(|\xi_{(m)} - 1/2| > \epsilon) \\
&\equiv VI_n(a) + VI_n(b).
\end{aligned}
$$

Now the second term here, $VI_n(b)$, can be handled easily using an exponential bound of Wellner (1977):

$$
P(\sqrt{n}|\xi_{(m)} - p_m| \ge \sqrt{p_m q_m} \lambda) \le 2\exp(-\lambda/10)
$$

where $p_m \equiv m/(n+1) = m/(2m+1)$; see Wellner (1977) and display (7) on page 454 of SW (1986) – where (7) there should read as in the last display here. This implies that

$$
P(|\xi_{(m)} - 1/2| > \epsilon) \le 2\exp(-2\epsilon\sqrt{n}/5).
$$

Thus we find that

$$VI_n(b) \le 2^s |Q(1/2)|^s n^{s/2} \exp(-2\epsilon\sqrt{n}/5) \to 0 \qquad \text{as } \ n \to \infty.$$

It remains to handle $VI_n(a)$. The best approach here is to return to the finite sample calculations and to bound much as in the proof of the pointwise convergence argument outlined below.

$$\begin{aligned}
&n^{s/2} E\left\{[\xi_{(m)}(1-\xi_{(m)})]^{-s/r} 1\{|\xi_{(m)} - 1/2| > \epsilon\}\right\} \\
&= n^{s/2} \int_{|u-1/2|>\epsilon} u^{-s/r}(1-u)^{-s/r} u^{m-1}(1-u)^m du \cdot \frac{n!}{(m-1)!(n-m)!} \\
&= n^{s/2} \int_{|u-1/2|>\epsilon} u^{m-1-s/r}(1-u)^{m-1-s/r}(1-u) du \cdot \frac{1}{\sqrt{2\pi}} 2^n (1+o(1))\sqrt{n} \\
&= n^{(s+1)/2} \frac{1}{\sqrt{2\pi}}(1+o(1)) \int_{|u-1/2|>\epsilon} 2^n u^{m-1-s/r}(1-u)^{m-1-s/r}(1-u) du \\
&= n^{(s+1)/2} \frac{1}{\sqrt{2\pi}}(1+o(1)) \int_{|u-1/2|>\epsilon} (2u)^{m-s/r}(2(1-u))^{m-1-s/r}(1-u) du \cdot 2^{-2m+2s/r} 2^n \\
&= 2^{2s/r} n^{(s+1)/2} \frac{1}{\sqrt{2\pi}}(1+o(1)) \int_{|u-1/2|>\epsilon} (4u(1-u))^{m-1-s/r}(1-u) du
\end{aligned}$$

where

$$4u(1-u) = 1 - 4(u-1/2)^2 \le \exp(-4(u-1/2)^2).$$

Therefore, for n sufficiently large,

$$\begin{aligned}
&n^{s/2} E\left\{[\xi_{(m)}(1-\xi_{(m)})]^{-s/r} 1\{|\xi_{(m)} - 1/2| > \epsilon\}\right\} \\
&\le 2^{2s/r} n^{(s+1)/2} \frac{2}{\sqrt{2\pi}} \int_{|u-1/2|>\epsilon} \exp(-4(u-1/2)^2(m-1-s/r)) du \\
&= 2^{2s/r} n^{(s+1)/2} \frac{2}{\sqrt{2\pi}} \int_{|z|>\epsilon} \exp(-4z^2(m-1-s/r)) dz \\
&= 2^{2s/r} n^{(s+1)/2} \frac{2}{\sqrt{2\pi}} \int_{|z|>\epsilon} \exp(-4z^2(m-1-s/r)) dz \\
&= 2^{2s/r+1} n^{(s+1)/2} \sigma_n P(\sigma_n |Z| > \epsilon), \qquad \sigma_n^2 \equiv \frac{1}{8(m-1-s/r)}, \\
&\le 2^{2s/r+2} n^{(s+1)/2} \sigma_n \frac{1}{\epsilon/\sigma_n} \phi(\epsilon/\sigma_n) \qquad \text{by Mills ratio inequality} \\
&= 2^{2s/r+2} n^{(s+1)/2} \sigma_n^2 \frac{1}{\epsilon} \frac{1}{\sqrt{2\pi}} \exp(-4\epsilon^2(m-1-s/r)) \\
&\to 0 \qquad \text{as } \ n \to \infty.
\end{aligned}$$

Combining the pieces yields

$$\begin{aligned} E|V_n|^s &\le M^s + M^s \left(\sup_{u:|u-1/2|\le M/\sqrt{n}} |Q'(u)|^s \right) + C_s 2^{-s/2} \left(\sup_{u:|u-1/2|\le \epsilon} |Q'(u)|^s \right) \\ &\quad + 2^s |Q(1/2)|^s n^{s/2} P(|\xi_{(m)} - 1/2| > \epsilon) \\ &\quad + 2^{s-1} K_r^s 2^{2s/r+s} \frac{1}{\sqrt{2\pi}} \frac{1}{\epsilon} \frac{n^{(s+1)/2}}{8(m-1-s/r)} \exp(-4\epsilon^2(m-1-s/r)), \end{aligned}$$

so that

$$\limsup_{n\to\infty} E|V_n|^s \le M^s \left(1 + |Q'(1/2)|^s\right) + C_s 2^{-s/2} \left(\sup_{u:|u-1/2|\le \epsilon} |Q'(u)|^s \right) < \infty;$$

i.e. (**??**) holds.

Some of the intuition for the proof above has relied on the following development concerning convergence of the densities of $\sqrt{n}(X_{(m)} - F^{-1}(1/2))$ to $N(0, 1/(4f^2(F^{-1}(1/2))))$. Note that $X_{(n/2)} = X_{(m)}$ has density

$$\frac{n!}{(m-1)!(n-m)!} F(y)^{m-1} f(y)(1-F(y))^{n-m},$$

and hence $\sqrt{n}(X_{(m)} - F^{-1}(1/2))$ has density

$$n^{-1/2} \frac{n!}{(m-1)!(n-m)!} F(F^{-1}(1/2) + xn^{-1/2})^{m-1} f(F^{-1}(1/2) + xn^{-1/2})(1 - F(F^{-1}(1/2) + xn^{-1/2}))^{n-}$$

Thus it follows that

$$\begin{aligned} &E\left(\sqrt{n}(X_{(m)} - F^{-1}(1/2))\right)^2 \\ &= \int_{-\infty}^{\infty} x^2 n^{-1/2} \frac{n!}{(m-1)!(n-m)!} F(F^{-1}(1/2) + xn^{-1/2})^{m-1} f(F^{-1}(1/2) + xn^{-1/2}) \\ &\qquad \cdot (1 - F(F^{-1}(1/2) + xn^{-1/2}))^{n-m}\, dx \\ &= \left(\frac{1}{\sqrt{2\pi}} + o(1) \right) 2^n \int_{-\infty}^{\infty} x^2 F(F^{-1}(1/2) + xn^{-1/2})^{m-1} f(F^{-1}(1/2) + xn^{-1/2}) \\ &\qquad (1 - F(F^{-1}(1/2) + xn^{-1/2}))^{n-m}\, dx \qquad (4) \end{aligned}$$

by Stirling's formula:

$$n! \sim \sqrt{2\pi n}(n/e)^n, \qquad m! = (n-m)! = \sqrt{2\pi m}(m/e)^m = \sqrt{2\pi(n/2)}((n/2)/e)^{n/2},$$

and hence

$$\begin{aligned}
\frac{n!}{(m-1)!(n-m)!} &= m\frac{n!}{(m!)^2} = m\frac{\sqrt{2\pi n}(n/e)^(1+o(1))}{\left(\sqrt{2\pi n/2}((n/2)/e)^{n/2}(1+o(1))\right)^2} \\
&= \frac{n}{2}\frac{1}{\sqrt{2\pi}}\frac{(n/e)^n}{((n/2)/e)n}\cdot\frac{\sqrt{n}}{n/2}\cdot(1+o(1)) \\
&= \frac{1}{\sqrt{2\pi}}2^n\sqrt{n}(1+o(1)).
\end{aligned}$$

Now note that

$$F(F^{-1}(1/2)+xn^{-1/2}) \to F(F^{-1}(1/2)) = 1/2$$

and, moreover,

$$g_n(x) \equiv \sqrt{n}(F(F^{-1}(1/2)+xn^{-1/2})-1/2) \to f(F^{-1}(1/2))x \equiv g(x)$$

for each fixed x. Thus we can rewrite the integrand of the right side of (**??**) as

$$\begin{aligned}
&2^n x^2 F(F^{-1}(1/2)+xn^{-1/2})^{-1} f(F^{-1}(1/2)+xn^{-1/2}) F(F^{-1}(1/2)+xn^{-1/2})^m \\
&\quad\cdot(1-F(F^{-1}(1/2)+xn^{-1/2}))^{n-m} \\
&= x^2 F(F^{-1}(1/2)+xn^{-1/2})^{-1} f(F^{-1}(1/2)+xn^{-1/2}) \\
&\quad\cdot\{4u_n(x)(1-u_n(x))\}^m \qquad (5)
\end{aligned}$$

where

$$u_n(x) \equiv F(F^{-1}(1/2)+xn^{-1/2}) = 1/2 + n^{-1/2}g_n(x)$$

Note that the function $r(u) \equiv 4u(1-u)$ is maximized at $u = 1/2$ with $r(1/2) = 1$ and

$$r(u) = 1 - 4(u-1/2)^2.$$

Hence the term $\{4u_n(x)(1-u_n(x))\}^m$ can be written as

$$\begin{aligned}
\left\{1-4(u_n(x)-1/2)^2\right\}^m &= \left\{1-\frac{4g_n^2(x)}{n}\right\}^m = \left\{1-\frac{4g_n^2(x)}{n}\right\}^{n/2} \\
&\to \exp(-4g^2(x)/2) = \exp(-2g^2(x)) \\
&= \exp\left(-\frac{x^2/2}{\frac{1/4}{f^2(F^{-1}(1/2))}}\right) = \exp\left(-\frac{x^2}{2\sigma^2}\right)
\end{aligned}$$

with $\sigma^2 \equiv (1/4)/f^2(F^{-1}(1/2))$. Since

$$\begin{aligned} F(F^{-1}(1/2) + xn^{-1/2})^{-1} f(F^{-1}(1/2) + xn^{-1/2}) &\rightarrow F(F^{-1}(1/2))^{-1} f(F^{-1}(1/2)) \\ &= 2f(F^{-1}(1/2)) = 1/\sigma, \end{aligned}$$

It follows that we have

$$\begin{aligned} &E\left(\sqrt{n}(X_{(m)} - F^{-1}(1/2))\right)^2 \\ &= \left(\frac{1}{\sqrt{2\pi}} + o(1)\right) \int_{-\infty}^{\infty} x^2 F(F^{-1}(1/2) + xn^{-1/2})^{-1} f(F^{-1}(1/2) + xn^{-1/2}) \\ &\qquad \{4u_n(x)(1 - u_n(x))\}^m \, dx \qquad (6) \\ &\rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 \exp(-x^2/(2\sigma^2)) dx = \sigma^2 \qquad (7) \end{aligned}$$

if the interchange of limit and integral can be justified. To do this we need to either find an integrable dominating function and apply the dominated convergence theorem, or use a uniform integrability argument.