Statistics 583, Problem Set 9 Solutions

Wellner; 6/1/2011

1. (a) Verify the variance part of Theorem 6.9, Wasserman, page 129.
   (b) Verify the proof of (6.12) - (6.14) in Theorem 6.11, Wasserman, page 129. A question about the proof: Is the $o(1/n)$ term on page 142, line -8, correct? Does this have any effect on (6.12)?
   (c) Can you find the limiting distribution of $n^{1/3}(\widehat{f}_n(x) - f(x))$ under hypotheses similar to those used in (a corrected version of) Theorem 6.11?

   Solution: (a) Since

$$\widehat{f}_n(x) = \sum_{j=1}^{m} \frac{\widehat{p}_j}{h} 1_{B_j}(x) \qquad \text{and} \qquad \mathrm{Var}(\widehat{p}_j) = \frac{p_j(1-p_j)}{n},$$

it follows that

$$\mathrm{Var}(\widehat{f}_n(x)) = \sum_{j=1}^{n} \frac{p_j(1-p_j}{nh^2} 1_{B_j}(x).$$

(b) First,

$$\begin{aligned} R(\widehat{f}_n, f) &= E\int (\widehat{f}_n(x) - f(x))^2 dx \\ &= \int \mathrm{Var}(\widehat{f}_n(x))dx + \int \mathrm{bias}_n^2(\widehat{f}_n(x))dx, \end{aligned}$$

so it suffices to compute these two contributions to the risk separately. First the integrated variance term: by the calculation in part (a) above,

$$\begin{aligned}
\int \mathrm{Var}(\widehat{f}_n(x))dx &= \int \sum_{j=1}^m \frac{p_j(1-p_j)}{nh^2} 1_{B_j}(x)dx \\
&= \frac{1}{nh^2}\sum_{j=1}^m p_j \int_{B_j}(x)dx - \frac{1}{nh^2}\sum_{j=1}^n p_j^2 \int_{B_j} dx \\
&= \frac{1}{nh} - \frac{1}{nh}\sum_{j=1}^m p_j^2 \\
&= \frac{1}{nh} - \frac{1}{nh}\sum_{j=1}^m h^2 f(x_j)^2 \qquad \text{since } \ p_j = hf(x_j) \ \text{ for some } x_j \in B_j \\
&= \frac{1}{nh} - \frac{1}{n}\sum_{j=1}^m hf(x_j)^2 \\
&= \frac{1}{nh} - \frac{1}{n}\sum_{j=1}^m \left(\int_{B_j} f^2(x)dx + o(h)\right) \\
&= \frac{1}{nh} - \frac{1}{n}\left(\int_0^1 f^2(x)dx + o(1)\right) = \frac{1}{nh} + O(n^{-1}).
\end{aligned}$$

To compute the bias term, we first use absolute continuity of f' to expand f as

$$f(y) = f(x) + (y-x)f'(x) + (y-x)^2 \int_0^1 f''(x+s(y-x))(1-s)ds$$

Therefore

$$\begin{aligned}
p_j &= \int_{B_j} f(y)dy \\
&= \int_{B_j}\left(f(x) + (y-x)f'(x) + (y-x)^2\int_0^1 f''(x+s(y-x))(1-s)ds\right)dy \\
&= hf(x) + f'(x)\frac{1}{2}(y-x)^2\Big|_{(j-1)/m}^{j/m} + O(h^{5/2}) \\
&= hf(x) + hf'(x)\left(h(j-1/2)-x\right) + O(h^{5/2})
\end{aligned}$$

since

$$\begin{aligned}
\left|\int_0^1 f''(x+s(y-x))(1-s)ds\right| &\le \left(\int_0^1 f''(x+s(y-x))^2 ds \int_0^1 (1-s)^2 ds\right)^{1/2} \\
&= \left(\frac{1}{3\sqrt{y-x}}\int_0^1 f''(v)^2 dv\right)^{1/2},
\end{aligned}$$

(assuming that the integral on the right side is finite) and hence

$$\begin{aligned}
&\left| \int_{B_j} (y-x)^2 \int_0^1 f''(x+s(y-x))(1-s)ds \right| \\
&\leq \int_{B_j} (y-x)^2 \frac{1}{3\sqrt{y-x}} \left(\int f''(v)^2 dv \right)^{1/2} dy \\
&\leq \frac{1}{3} \left(\int f''(v)^2 dv \right)^{1/2} \int_{B_j} \left(y - \frac{y-1}{m} \right)^{3/2} dy \\
&= \frac{2}{15} \left(\int f''(v)^2 dv \right)^{1/2} h^{5/2}.
\end{aligned}$$

— Thus

$$\begin{aligned}
\text{bias}_n(\widehat{f}(x)) &\equiv b(x) = E\widehat{f}_n(x) - f(x) = \frac{p_j}{h} - f(x) \\
&= \frac{1}{h} \left(hf(x) + hf'(x) \left(h(j-1/2) - x \right) + O(h^{5/2}) \right) - f(x) \\
&= f'(x) \left(h(j-1/2) - x \right) + O(h^{3/2}),
\end{aligned}$$

and this yields

$$\begin{aligned}
\int_{B_j} b^2(x)dx &= \int_{B_j} f'(x)^2 \left(h(j-1/2) - x \right)^2 dx + O(h^4) \\
&= f'(\tilde{x}_j)^2 \int_{B_j} \left(h(j-1/2) - x \right)^2 dx + O(h^4) \\
&= f'(\tilde{x}_j)^2 \frac{h^3}{12} + O(h^4)
\end{aligned}$$

for some $\tilde{x}_j \in B_j$ and where we have computed

$$\int_{B_j} \left(h(j-1/2) - x \right)^2 dx = -\frac{1}{3} \left(h(j-1/2) - x \right) \Big|_{(j-1)/m}^{j/m} = \frac{h^3}{12}.$$

Therefore

$$\begin{aligned}
\int_0^1 b^2(x)dx &= \sum_{j=1}^m \int_{B_j} b^2(x)dx \\
&= \sum_{j=1}^m \left(f'(\tilde{x}_j)^2 \frac{h^3}{12} + O(h^4) \right) \\
&= \frac{h^2}{12} \sum_{j=1}^m h(f'(\tilde{x}_j))^2 + O(h^3) \\
&= \frac{h^2}{12} \int_0^1 f'(x)^2 dx + o(h^2).
\end{aligned}$$

Putting these pieces together yields

$$R(\widehat{f}_n, f) = \frac{1}{nh} + O(n^{-1}) + \frac{h^2}{12}\int_0^1 f'(x)^2 dx + o(h^2).$$

Thus it seems that the term $o(n^{-1})$ in Wasserman's (6.12) should be $O(n^{-1})$, and the assumption should be that $\int f''(v)^2 dv < \infty$. For a very careful treatment of this problem, see Freedman and Diaconis, "On the histogram as a density estimator: L_2 theory", *Z. Wahrschein. verw. Geb.* **57** (1981), 453 - 476.

(c) If we take the bandwidth $h = h_n = Cn^{-1/3}$, then

$$\begin{aligned} n^{1/3}(\widehat{f}_n(x) - f(x)) &= n^{1/3}(\widehat{f}_n(x) - E\widehat{f}_n(x) + n^{1/3}(E\widehat{f}_n(x) - f(x)) \\ &= n^{1/3}(\widehat{f}_n(x) - E\widehat{f}_n(x)) + n^{1/3} b_n(x) \end{aligned}$$

where, from our calculations in (b),

$$\begin{aligned} n^{1/3} b_n(x) &= f'(x) n^{1/3}\left(h_n(j - 1/2) - x\right) + O(h_n^2) \\ & \begin{cases} \le f'(x)n^{1/3}(h_n/2) + O(h_n^2) = f'(x)C/2 + O(h_n^2), \\ \ge f'(x)n^{1/3}(-h_n/2) + O(h_n^2) = -f'(x)C/2 + O(h_n^2) \end{cases} \end{aligned}$$

Thus it follows that

$$-f'(x)C/2 \le \liminf_n n^{1/3} b_n(x) \le \limsup_n n^{1/3} b_n(x) \le f'(x)C/2.$$

On the other hand for the random term $n^{1/3}(\widehat{f}_n(x) - E\widehat{f}_n(x))$ we have

$$\begin{aligned} n^{1/3}(\widehat{f}_n(x) - E\widehat{f}_n(x)) &= \frac{n^{1/3}}{h_n}(\widehat{p}_j - p_j) 1_{B_j}(x) \\ &= n^{2/3} C(\widehat{p}_j - p_j) 1_{B_j}(x) \\ &= \frac{1}{C}\sum_{i=1}^n n^{-1/3}(1_{B_j}(X_i) - p_j) \\ &\equiv \frac{1}{C}\sum_{i=1}^n X_{ni}. \end{aligned}$$

Now $\{X_{ni} : \ i = 1, \ldots, n\}$ satisfy $E(X_{ni}) = 0$,

$$\begin{aligned} \sigma_{ni}^2 = Var(X_{ni}) &= n^{-2/3} p_j(1 - p_j) = n^{-2/3} f(x_j^*)h - n^{-2/3}(f(x_j^*)h)^2 \\ &= Cf(x_j^*)n^{-1} - C^2 f(x_j^*)^2 n^{-4/3}, \end{aligned}$$

so that

$$\sum_{i=1}^{n} \sigma_{ni}^2 = Cf(x_j^*) - O(h_n) \to Cf(x)$$

assuming that f is continuous at x. Furthermore,

$$\frac{1}{\sigma_n^2} \sum_{i=1}^{n} E|X_{ni}|^2 1\{|X_{ni}| > \epsilon \sigma_n\} \to 0$$

for every $\epsilon > 0$ since

$$|X_{ni}| \le n^{-1/3} \le \epsilon C f(x)/2$$

for all n sufficiently large if $f(x) > 0$. Thus we conclude from the Lindeberg-Feller CLT that

$$\frac{\sum_{i=1}^{n} n^{-1/3}(1_{B_j}(X_i) - p_j)}{\sigma_n} \to_d N(0,1),$$

and hence that

$$\sum_{i=1}^{n} n^{-1/3}(1_{B_j}(X_i) - p_j) \to_d N(0, Cf(x)),$$
$$n^{1/3}(\widehat{f}_n(x) - E\widehat{f}_n(x)) \to_d \frac{1}{C} N(0, Cf(x)) = N(0, f(x)/C).$$

Therefore although we cannot conclude that $n^{1/3}(\widehat{f}_n(x) - f(x)) \to_d$, it does follow that

$$n^{1/3}(\widehat{f}_n(x) - f(x)) \quad = \quad n^{1/3}(\widehat{f}_n(x) - E\widehat{f}_n(x)) + n^{1/3} b_n(x)$$

where the random term converges in distribution to $N(0, f(x)/C)$ and the bias term is bounded between $-f'(x)C/2$ and $f'(x)C/2$.

If we choose the bandwidth to be just slightly smaller than order $n^{-1/3}$, say $h_n = C(n \log n)^{-1/3}$ then the bias term is small relative to $\sqrt{nh_n} = \sqrt{nn^{-1/3}(\log n)^{-1/3}} = n^{1/3}(\log n)^{-1/6}$ (since $\sqrt{nh_n} b_n(x) = O(n^{1/3}(\log n)^{-1/6} n^{-1/3}(\log n)^{-1/3}) = O((\log n)^{-1/2})$), and the random term still satisfies

$$\sqrt{nh_n}(\widehat{f}_n(x) - E\widehat{f}_n(x)) \to_d N(0, Cf(x)),$$

so we can conclude that

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \to_d N(0, f(x)/C)$$

with this choice of the bandwidth.

An interesting comparison is to investigate the behavior of a kernel estimator with a "boxcar kernel", $K(x) = 2^{-1} 1_{[-1,1]}(x)$ in this same situation with just one continuous derivative f': then with $h_n = Cn^{-1/3}$ I find that

$$n^{1/3}(\widehat{f}_n(x) - f(x)) \to_d N(Cf'(x)/2, f(x) \int K^2(z)dz/C).$$

Note that as C increases the asymptotic bias grows, but the asymptotic variance decreases.

2. (a) Consider the kernel density estimator defined in (6.26), Wasserman, page 132. Show that if the density f and the kernel k satisfy the hypotheses of Wasserman's theorem 6.28, page 133, and $h = h_n$ satisfies the hypotheses of Theorem 6.27, then for fixed $x \in \mathbb{R}$,

$$\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \to_d N\left(0, f(x) \int k^2(x)dx\right).$$

(b) Under what restriction on h_n does it follow (from (a) together with further analysis of the bias) that

$$\sqrt{nh_n}(\hat{f}_n(x) - f(x)) \to_d N\left(0, f(x) \int k^2(x)dx\right)?$$

(c) If $h_n = cn^{-1/5}$ and the hypotheses of (a) hold, find the limiting distribution of $\sqrt{nh_n}(\hat{f}_n(x) - f(x))$.
(d) Under the same assumptions as in (c), find the limiting distribution of $\sqrt{nh_n}(\sqrt{\hat{f}_n(x)} - \sqrt{f(x)})$.
(e) Suppose that $x, y \in \mathbb{R}$ with $x < y$. Find the joint limiting distribution of $(\sqrt{nh_n}(\hat{f}_n(x) - f(x)), \sqrt{nh_n}(\hat{f}_n(y) - f(y))$ under the assumptions in (b) and (c).

Solution: (a) First write

$$\begin{aligned}
&\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \\
&= \sqrt{nh_n}\left(\frac{1}{nh_n}\sum_{i=1}^n k((x - X_i)/h_n) - \frac{1}{h_n}Ek((x - X_1)/h_n)\right) \\
&= \sum_{i=1}^n \frac{1}{\sqrt{nh_n}}(k((x - X_i)/h_n) - Ek((x - X_1)/h_n)) \\
&\equiv \sum_{i=1}^n X_{n,i}
\end{aligned}$$

where the $X_{n,i}$'s are independent and identically distributed for each n, but with a distribution depending on n. Thus we will use the Lindeberg-Feller CLT. By easy calculations, $E(X_{n,i}) = 0$, and

$$\sigma_{n,i}^2 = Var(X_{n,i}) = \frac{1}{nh_n} \int \left(k((x-y)/h_n) - \int k((x-v)/h_n) f(v) dv \right)^2 f(y) dy,$$

so

$$\begin{aligned} \sigma_n^2 &= \sum_{i=1}^n \sigma_{n,i}^2 \\ &= \frac{1}{h_n} \int \left(k((x-y)/h_n) - \int k((x-v)/h_n) f(v) dv \right)^2 f(y) dy \\ &= \int k(z)^2 f(x - zh_n) dz - h_n \left(\int k(z) f(x - zh_n) dz \right)^2 \\ &\to f(x) \int k(z)^2 dz \end{aligned}$$

by the dominated convergence theorem (if, for example, f is bounded). It remains only to verify the Lindeberg condition:

$$\frac{1}{\sigma_n^2} \sum_{i=1}^n E\{|X_{n,i}|^2 1\{|X_{n,i}| > \epsilon \sigma_n\} \to 0$$

as $n \to \infty$ for every $\epsilon > 0$. But

$$\begin{aligned} X_{ni} &\equiv \frac{1}{\sqrt{nh_n}} \left\{ k\left(\frac{x - X_i}{h_n} \right) - \int k\left(\frac{x-v}{h_n} \right) f(v) dv \right\} \\ &= \frac{1}{\sqrt{nh_n}} k\left(\frac{x - X_i}{h_n} \right) - \frac{h_n}{\sqrt{nh_n}} \int k(z) f(x - zh_n) dz \\ &\equiv Y_{ni} - \sqrt{\frac{h_n}{n}} (f(x) + o(1)). \end{aligned}$$

Thus

$$\begin{aligned} &|X_{ni}| \le |Y_{ni}| + \sqrt{\frac{h_n}{n}} (f(x) + o(1)), \\ &|X_{ni}|^2 \le 2\{Y_{ni}^2 + \frac{h_n}{n} (f(x) + o(1))^2\}, \quad \text{and} \\ &\{|X_{ni}| > \epsilon \sigma_n\} \subset \{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\} \cup \{\sqrt{h_n/n}(f(x) + o(1)) > (1/2)\epsilon\sigma_n\} \\ &\qquad = \{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\} \end{aligned}$$

for n sufficiently large since the second event on the right is empty for large n. It follows that

$$
\begin{aligned}
&E|X_{n,i}|^2 1\{|X_{n,i}| > \epsilon\sigma_n\} \\
&\quad \le \; 2E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\sigma_n/2\} + 2\frac{h_n}{n}(f(x)+o(1))^2 E1\{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\} \\
&\quad \le \; 2E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\sigma_n/2\} + \frac{2\frac{h_n}{n}(f(x)+o(1))^2}{((1/2)\epsilon\sigma_n)^2} E|Y_{n,i}|^2 1\{|Y_{ni}| \ge (1/2)\epsilon\sigma_n\}\} \\
&\quad = \; (2+o(1))E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\sigma_n/2\},
\end{aligned}
$$

and hence it suffices to show that

$$\sum_{i=1}^n E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\} \to 0 \qquad \text{for all } \; \epsilon > 0.$$

But

$$
\begin{aligned}
&\sum_{i=1}^n E|Y_{n,i}|^2 1\{|Y_{n,i}| > \epsilon\} \\
&\quad = \sum_{i=1}^n E\frac{k((x-X_i)/h_n)^2}{nh_n} 1\{k((x-X_i)/h_n) > \epsilon\sqrt{nh_n}\} \\
&\quad = \frac{1}{h_n}\int k((x-y)/h_n)^2 1\{k((x-y)/h_n) > \epsilon\sqrt{nh_n}\} f(y)dy \\
&\quad = \int k^2(z) f(x-zh_n) 1\{k(z) > \epsilon\sqrt{nh_n}\} dz \\
&\quad \to 0
\end{aligned}
$$

by the dominated convergence theorem if $\|f\|_\infty < \infty$. Thus by the Lindeberg-Feller CLT it follows that $\sum_1^n X_{n,i}/\sigma_n \to N(0,1)$. Since $\sigma_n^2 \to f(x)\int k^2(z)dz$, the conclusion follows.

(b) Note that the bias is

$$
\begin{aligned}
E\hat{f}_n(x) - f(x) &= \int k(z) f(x - zh_n) dz - f(x) = \int k(z)\{f(x-zh_n) - f(x)\} dz \\
&= \int k(z) \frac{1}{2} f''(z_x^*) z^2 h_n^2 dz
\end{aligned}
$$

where $|z_x^* - x| \le |x - zh_n - x| = |z|h_n$. Thus if $f''(x)$ is bounded and $\sqrt{nh_n}h_n^2 \to 0$, the bias term is negligible and

$$
\begin{aligned}
\sqrt{nh_n}(\hat{f}_n(x) - f(x)) &= \sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + \sqrt{nh_n}(E\hat{f}_n(x) - f(x)) \\
&\to_d N(0, f(x)\int k^2(z)dz) + 0 = N(0, f(x)\int k^2(z)dz).
\end{aligned}
$$

Note that $\sqrt{nh_n}h_n^2 = (nh_n^5)^{1/2} \to 0$ if and only if $nh_n^5 \to 0$; i.e. if $h_n = o(n^{-1/5})$. Thus $h_n = cn^{-1/4}$ yields negligible bias when f'' exists (and is sufficiently bounded).
(c) When $h_n = cn^{-1/5}$, then the first display in (b) yields

$$\begin{aligned}\sqrt{nh_n}(E\hat{f}_n(x) - f(x)) &= \sqrt{c}n^{2/5}c^2n^{-2/5}\frac{1}{2}\int z^2k(z)f''(z_x^*)dz \\ &\to \frac{c^{5/2}}{2}f''(x)\int z^2k(z)dz \equiv B(c,f,k,x)\end{aligned}$$

Thus it follows that

$$\begin{aligned}\sqrt{nh_n}(\hat{f}_n(x) - f(x)) &= \sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + \sqrt{nh_n}(E\hat{f}_n(x) - f(x)) \\ &\to_d N(0, f(x)\int k^2(z)dz) + B(c,f,k) \\ &= N(B(c,f,k,x), f(x)\int k^2(z)dz).\end{aligned}$$

(d) Under the assumptions in (c), it follows from the delta-method with $g(v) = v^{1/2}$ that for any x such that $f(x) > 0$ we have

$$\begin{aligned}\sqrt{nh_n}(\sqrt{\hat{f}_n(x)} - \sqrt{f(x)}) &\to_d \frac{1}{2}f(x)^{-1/2}N\left(B(c,f,k), f(x)\int k^2(z)dz\right) \\ &= N\left(\frac{B(c,f,k,x)}{2\sqrt{f(x)}}, 4^{-1}\int k^2(z)dz\right).\end{aligned}$$

Note that the variance now depends only on the kernel and not on f.
(e) To find the joint limiting distribution of $(\sqrt{nh_n}(\hat{f}_n(x) - f(x)), \sqrt{nh_n}(\hat{f}_n(y) - f(y))$ we will use the Cramér - Wold device: let $(a,b) \in \mathbb{R}^2$. We will first show that

$$\begin{aligned}&a\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + b\sqrt{nh_n}(\hat{f}_n(y) - E\hat{f}_n(y)) \\ &\quad\to_d N\left(0, a^2f(x)\int k^2(z)dz + b^2f(y)\int k^2(z)dz\right)\end{aligned} \tag{1}$$

for each fixed $(a,b) \in \mathbb{R}^2$. This implies that

$$\begin{pmatrix}\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \\ \sqrt{nh_n}(\hat{f}_n(y) - E\hat{f}_n(y))\end{pmatrix} \to_d N_2\left(0, \begin{pmatrix} f(x) & 0 \\ 0 & f(y)\end{pmatrix}\int k^2(z)dz\right) \tag{2}$$

and will yield the desired results after analyzing the relevant bias terms. But

$$
\begin{aligned}
& a\sqrt{nh_n}(\hat{f}_n(x) - E\hat{f}_n(x)) + b\sqrt{nh_n}(\hat{f}_n(y) - E\hat{f}_n(y)) \\
= \;& \frac{1}{nh_n}\sum_{i=1}^{n}\left\{a\left\{k\left(\frac{x-X_i}{h_n}\right) - \int k\left(\frac{x-v}{h_n}\right) f(v)dv\right\}\right. \\
& \left. + \, b\left\{k\left(\frac{y-X_i}{h_n}\right) - \int k\left(\frac{y-v}{h_n}\right) f(v)dv\right\}\right\} \\
\equiv \;& \sum_{i=1}^{n} \tilde{X}_{n,i}
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{X}_{n,i} \;\equiv\; & \frac{1}{nh_n}\left\{a\left\{k\left(\frac{x-X_i}{h_n}\right) - \int k\left(\frac{x-v}{h_n}\right) f(v)dv\right\}\right. \\
& \left. + \, b\left\{k\left(\frac{y-X_i}{h_n}\right) - \int k\left(\frac{y-v}{h_n}\right) f(v)dv\right\}\right\}
\end{aligned}
$$

We compute $E\tilde{X}_{n,i} = 0$ as in (a), but now

$$
\begin{aligned}
& \sigma_{ni}^2 = Var(\tilde{X}_{n,i}) \\
= \;& (nh_n)^{-1}\left\{a^2 nVar\{k((x-X_1)/h_n) + b^2 Var((y-X_1)/h_n)\right. \\
& \left. + \, 2abCov(k((x-X_1)/h_n), k((y-X_1)/h_n))\right\}
\end{aligned}
$$

so that

$$
\begin{aligned}
\sigma_n^2 \;=\; & \sum_{i=1}^{n} \sigma_{ni}^2 \\
= \;& h_n^{-1}\left\{a^2 \int \left(k((x-w)/h_n) - \int k((x-v)/h_n) f(v)dv\right)^2 f(w)dw\right. \\
& + \, b^2 \int \left(k((y-w)/h_n) - \int k((y-v)/h_n) f(v)dv\right)^2 f(w)dw \\
& + 2ab \int \left(k((x-w)/h_n) - \int k((x-v)/h_n) f(v)dv\right) \\
& \left. \left(k((y-w)/h_n) - \int k((y-v)/h_n) f(v)dv\right) f(w)dw\right\} \\
\to \;& a^2 f(x) \int k^2(z)dz + b^2 f(y) \int k^2(z)dz + 0
\end{aligned}
$$

where the argument for the first two terms is exactly as in (a), and the third term converges to 0 since

$$\begin{aligned}
& h_n^{-1}\int\left(k((x-w)/h_n)-\int k((x-v)/h_n)f(v)dv\right) \\
& \qquad \left(k((y-w)/h_n)-\int k((y-v)/h_n)f(v)dv\right)f(w)dw \\
= \;& h_n^{-1}\int\left(k((x-w)/h_n)-h_n^2(f(x)+o(1))\right) \\
& \qquad \left(k((y-w)/h_n)-h_n^2(f(y)+o(1))\right)f(w)dw \\
= \;& \int k\left(z-\frac{y-x}{h_n}\right)k(z)f(y-zh_n)dz+O(h_n) \\
\to \;& \int k(-\infty)k(z)dzf(y)=0
\end{aligned}$$

by the dominated convergence theorem if f and k are bounded using $(y-x)/h_n \to \infty$ since $y-x>0$ and $h_n \to 0$. Since $|\tilde{X}_{n,i}| \le |a||X_{n,i}(x)|+|b||X_{n,i}(y)|$, verification of the Lindeberg condition proceeds as in (a). Thus by the Lindeberg-Feller CLT, (1) holds, and this in turn yields (2).

3. (a) Wasserman, problem 6.9.3, page 143.
(b) Does (6.35) on Wasserman's page 136 hold? Give a formula that is more precise.

Solution: (a) First the case of kernel density estimates: here

$$\begin{aligned}
J(h) &= \int \widehat{f}_n(x)^2dx-2\int \widehat{f}_n(x)f(x)dx, \qquad \text{and} \\
\widehat{J}(h) &= \int \widehat{f}_n(x)^2dx-\frac{2}{n}\sum_{i=1}^n \widehat{f}_{(-i)}(X_i),
\end{aligned}$$

so to show that $E\{\widehat{J}(h)\}=E\{J(h)\}$, it suffices to show that

$$E\left\{\frac{1}{n}\sum_{i=1}^n \widehat{f}_{(-i)}(X_i)\right\}=E\left\{\int \widehat{f}_n(x)f(x)dx\right\}. \tag{3}$$

Now the right side in the last display is

$$\begin{aligned}
E\left\{\int \widehat{f}_n(x)f(x)dx\right\} &= \int E\{\widehat{f}_n(x)\}f(x)dx \\
&= \int\int \frac{1}{h}k\left(\frac{x-y}{h}\right)f(y)dyf(x)dx,
\end{aligned}$$

while the left side is, by conditioning on X_i,

$$
\begin{aligned}
E\left\{\frac{1}{n}\sum_{i=1}^{n}\widehat{f}_{(-i)}(X_i)\right\} &= \frac{1}{n}\sum_{i=1}^{n}E\left\{E(\widehat{f}_{(-i)}(X_i)|X_i)\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}E\{\int\frac{1}{h}k\left(\frac{X_i-x}{h}\right)f(x)dx\} \\
&= \iint\frac{1}{h}k\left(\frac{y-x}{h}\right)f(y)dyf(x)dx \\
&= \iint\frac{1}{h}k\left(\frac{x-y}{h}\right)f(y)dyf(x)dx
\end{aligned}
$$

by Fubini's theorem. Comparing the last two displays yields the claim.
In the case of histogram estimators, it again suffices to show that (3) holds. But

$$
\begin{aligned}
E\left\{\int\widehat{f}_n(x)f(x)dx\right\} &= \int E\{\widehat{f}_n(x)\}f(x)dx = \int\sum_{j=1}^{m}\frac{p_j}{h}1_{B_j}(x)f(x)dx \\
&= \sum_{j=1}^{m}\frac{p_j^2}{h},
\end{aligned}
$$

and, on the other hand,

$$
\begin{aligned}
E\left\{\frac{1}{n}\sum_{i=1}^{n}\widehat{f}_{(-i)}(X_i)\right\} &= E\left\{E\{n^{-1}\sum_{i=1}^{n}\widehat{f}_{(-i)}(X_i)|X_i\}\right\} \\
&= E\{n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{p_j}{h}1_{B_j}(X_i)\} \\
&= \sum_{j=1}^{m}\frac{p_j}{h}\int_{B_j}f(x)dx = \sum_{j=1}^{m}\frac{p_j^2}{h},
\end{aligned}
$$

and hence we conclude that (again) $E\{\widehat{J}(h)\} = E\{J(h)\}$.
(b) The claimed identity is

$$
\widehat{J}(h) = \frac{1}{n^2h}\sum_i\sum_j k^*\left(\frac{X_i-X_j}{h}\right) + \frac{2}{nh}k(0) + O(n^{-2})
$$

where $k^*(x) = k^{(2)}(x) - 2k(x)$ and $k^{(2)}(z) = \int k(z-y)k(y)dy$. To check this, write

$$
\widehat{f}_n^2(x) = \frac{1}{n^2h^2}\sum_i\sum_j k\left(\frac{x-X_i}{h}\right)k\left(\frac{x-X_j}{h}\right)
$$

so that, with $k^{(2)}(z) = \int k(z-y)k(y)dy$,

$$\begin{aligned}
\int \widehat{f}_n^2(x)dx &= \frac{1}{n^2h^2}\sum_i\sum_j \int k\left(\frac{x-X_i}{h}\right)k\left(\frac{x-X_j}{h}\right)dx \\
&= \frac{1}{n^2h^2}\sum_i\sum_j \int k\left(\frac{X_j+hy-X_i}{h}\right)k(y)hdy \\
&\quad \text{by the change of variables } \ (x-X_j)/h = y, \quad x = X_j + hy \\
&= \frac{1}{n^2h}\sum_i\sum_j \int k\left(y - \frac{X_i-X_j}{h}\right)k(y)dy \\
&= \frac{1}{n^2h}\sum_i\sum_j k^{(2)}\left(\frac{X_i-X_j}{h}\right) \\
&\quad \text{if we assume that } \quad k(z) = k(-z).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \widehat{f}_{(-i)}(X_i) &= \frac{1}{n}\sum_{i=1}^n\sum_{j\neq i}\frac{1}{(n-1)h}k\left(\frac{X_i-X_j}{h}\right) \\
&= \frac{1}{n(n-1)h}\sum_i\sum_j k\left(\frac{X_i-X_j}{h}\right) - \frac{1}{n-1}k(0) \\
&= \frac{1}{n^2h}\sum_i\sum_j k\left(\frac{X_i-X_j}{h}\right) - \frac{1}{n}k(0) + O_p(n^{-1})
\end{aligned}$$

Putting these pieces together yields the claimed identity, but with a remainder term which is $O_p(n^{-1})$ rather than $O_p(n^{-2})$. Thus it seems that the computational formula should really read as follows:

$$\widehat{J}(h) = \frac{1}{n^2h}\sum_i\sum_j k^{(2)}\left(\frac{X_i-X_j}{h}\right) - \frac{2}{n(n-1)h}\sum_{i=1}^n\sum_{j\neq i} k\left(\frac{X_i-X_j}{h}\right).$$

4. Suppose that $X_1, \ldots, X_n$ are i.i.d. $p \in \mathcal{P}$ where $\mathcal{P}$ is some class of densities on some subset $\mathcal{X} \subset \mathbb{R}^d$. (Thus $p(x) \geq 0$ for all $x \in \mathcal{X}$ and $\int_{\mathcal{X}} p(x)dx = 1$.) Let $\mathcal{M}$ denote the set of all functions p on $\mathcal{X}$ with $p(x) \geq 0$ (without imposing $\int_{\mathcal{X}} p(x)dx = 1$). Show that maximizing

$$\begin{aligned}
\widetilde{l}_n(p) &\equiv \mathbb{P}_n(\log f(\cdot)) - \left(\int_{\mathcal{X}} p(y)dy - 1\right) \\
&= n^{-1}\sum_{i=1}^n \log p(X_i) - \left(\int_{\mathcal{X}} p(y)dy - 1\right)
\end{aligned}$$

over $\mathcal{M}$ yields the maximum of

$$l_n(p) \equiv \mathbb{P}_n(\log f(\cdot))$$

over $\mathcal{P}$. More exactly, show that $\widetilde{l}_n(p/c) \geq l_n(p)$ where $c = \int_{\mathcal{X}} p(x)dx$ with equality if and only if $c = 1$. [This is the rationale for the first display in Wasserman's section 6.4, page 137.]

Solution: The first part of this statement seems OK, but in the second display it should read

$$l_n(p) \equiv \mathbb{P}_n(\log p(\cdot)).$$

The second part should probably be replaced by: "More exactly, show that for arbitrary $f \in \mathcal{M}$ we have $\tilde{l}_n(f) \leq l_n(f/c)$ where $c = \int_{\mathcal{X}} f(x)dx$ with equality if and only if $c = 1$." Here is a proof of this corrected statement: for $f \in \mathcal{M}$

$$\begin{aligned}
\tilde{l}_n(f) &= \mathbb{P}_n \log f - \int f(x)dx + 1 \\
&= \mathbb{P}_n \log(f/c) + \log c - c + 1 \\
&\leq \tilde{l}_n(f/c) \quad \text{since} \quad \log(c) - c + 1 \leq 0 \quad \text{and} \int (f(x)/c)dx = 1 \\
&= l_n(f/c).
\end{aligned}$$

It follows that

$$\tilde{l}_n(\hat{f}_n) = \sup_{f \in \mathcal{M}} \tilde{l}_n(f) \leq \sup_{p \in \mathcal{P}} l_n(p) = l_n(\hat{p}_n)$$

with equality when $\hat{f}_n = \hat{p}_n \in \mathcal{P} \subset \mathcal{M}$.