36-705 Intermediate Statistics Homework #8 Solutions

November 17, 2016

Problem 1 [20 pts.]

$$
\begin{aligned}
\mathbb{E}\big[\widehat{p}(0)\big] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{-X_i}{h}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{h}K\left(\frac{X_i}{h}\right)\right] \\
&= \frac{1}{h}\int_0^1 K\Big(\frac{u}{h}\Big)p(u)du \\
&= \int_0^{1/h} K(t)p(ht)dt && \text{let } t = \frac{u}{h} \\
&= \int_0^{1/h} K(t)\left(p(0) + ht\cdot\partial_+ p(0) + \frac{h^2t^2}{2}\cdot\partial_+^2 p(0) + o(h^2)\right)dt \\
&= p(0)\int_0^{1/h} K(t)dt + h\cdot\partial_+ p(0)\int_0^{1/h} tK(t)dt + \frac{h^2}{2}\cdot\partial_+^2 p(0)\int_0^{1/h} t^2K(t)dt + \int_0^{1/h} o(h^2)dt \\
&= \frac{p(0)}{2} + o(1) \\
&\to \frac{p(0)}{2}.
\end{aligned}
$$

Thus,

$$
\mathbb{E}\big[\widehat{p}(0)\big] - p(0) = -\frac{p(0)}{2}.
$$

Problem 2 [40 pts.]

First we will show

$$R(k) = \mathbb{E}\left[\int (\widehat{p}(x) - p(x))^2 dx \right] = \sum_{j=1}^{k} \mathbb{V}(\widehat{\beta}_j) + \sum_{j=k+1}^{\infty} \beta_j^2. \tag{1}$$

Note that because $X_1, \ldots, X_n$ are i.i.d., the $\phi_j(X_1), \ldots, \phi_j(X_n)$ are i.i.d. for all $j \in \mathbb{N}$, and thus

$$\begin{aligned}
\mathbb{E}[\widehat{\beta}_j] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i) \right] \\
&= \mathbb{E}[\phi_j(x)] \\
&= \int_0^1 p(x)\phi_j(x)dx \\
&= \beta_j.
\end{aligned}$$

Now we compute $R(k)$ by each term in the expansion.

$$\begin{aligned}
\int_0^1 \widehat{p}^2(x)dx &= \int_0^1 \left(\sum_{j=1}^{k} \widehat{\beta}_j \phi_j(x) \right)^2 dx \\
&= \int_0^1 \sum_{h_1 + \cdots h_k = 2} \binom{2}{h_1, \ldots, h_k} \prod_{j=1}^{k} \widehat{\beta}_j^{h_j} \phi_j^{h_j}(x)dx \\
&\overset{(2)}{=} \int_0^1 \sum_{j=1}^{k} \widehat{\beta}_j^2 \phi_j^2(x)dx \\
&= \sum_{j=1}^{k} \widehat{\beta}_j^2 \int_0^1 \phi_j^2(x)dx \\
&= \sum_{j=1}^{k} \widehat{\beta}_j^2. \qquad (3)
\end{aligned}$$

Notice at (2) that all the cross terms disappeared because

$$2\widehat{\beta}_j \widehat{\beta}_k \int_0^1 \phi_j(x)\phi_k(x)dx = 0 \quad \text{for all } j \neq k.$$

Now for the second term,

$$\begin{aligned}
\int_0^1 p^2(x)dx &= \int_0^1 \left(\sum_{j=1}^{\infty} \beta_j \phi_j(x) \right)^2 dx \\
&= \sum_{j=1}^{\infty} \beta_j^2 \underbrace{\int_0^1 \phi_j^2(x)dx}_{=1} \\
&= \sum_{j=1}^{\infty} \beta_j^2 \\
&= \sum_{j=1}^{k} \beta_j^2 + \sum_{j=k+1}^{\infty} \beta_j^2. \qquad (4)
\end{aligned}$$

Now for the last term,

$$
\begin{aligned}
-2\int_0^1 \widehat{p}(x)p(x)dx &= -2\int_0^1 \widehat{p}(x)\left(\sum_{j=1}^k \beta_j\phi_j(x) + \sum_{j=k+1}^\infty \beta_j\phi_j(x)\right)dx \\
&= -2\int_0^1 \left(\sum_{j=1}^k \widehat{\beta}_j\phi_j(x)\right)^2 dx \\
&= -2\sum_{j=1}^k \widehat{\beta}_j\beta_j. \qquad (5)
\end{aligned}
$$

Now adding (3), (4), and (5) together and taking the expectation, we have

$$
\begin{aligned}
R(k) = \mathbb{E}\left[\int (\widehat{p}(x) - p(x))^2 dx\right] &= \mathbb{E}\left[\sum_{j=1}^k \widehat{\beta}_j^2 + \sum_{j=1}^k \beta_j^2 + \sum_{j=k+1}^\infty \beta_j^2 - 2\sum_{j=1}^k \widehat{\beta}_j\beta_j\right] \\
&= \mathbb{E}\left[\sum_{j=1}^k \widehat{\beta}_j^2\right] + \sum_{j=1}^k \beta_j^2 - 2\mathbb{E}\left[\sum_{j=1}^k \widehat{\beta}_j\beta_j\right] + \sum_{j=k+1}^\infty \beta_j^2 \\
&= \mathbb{E}\left[\sum_{j=1}^k \widehat{\beta}_j^2\right] + \sum_{j=1}^k \beta_j^2 - 2\sum_{j=1}^k \mathbb{E}[\widehat{\beta}_j]\beta_j + \sum_{j=k+1}^\infty \beta_j^2 \\
&= \mathbb{E}\left[\sum_{j=1}^k \widehat{\beta}_j^2\right] + \sum_{j=1}^k \beta_j^2 - 2\sum_{j=1}^k \beta_j^2 + \sum_{j=k+1}^\infty \beta_j^2 \\
&= \mathbb{E}\left[\sum_{j=1}^k \widehat{\beta}_j^2\right] - \sum_{j=1}^k \beta_j^2 + \sum_{j=k+1}^\infty \beta_j^2 \\
&= \sum_{j=1}^k \mathbb{E}[\widehat{\beta}_j^2] - \sum_{j=1}^k \mathbb{E}[\widehat{\beta}_j]^2 + \sum_{j=k+1}^\infty \beta_j^2 \\
&= \sum_{j=1}^k \mathbb{V}(\widehat{\beta}_j) + \sum_{j=k+1}^\infty \beta_j^2. \quad \square
\end{aligned}
$$

Now,

$$
\begin{aligned}
\sum_{j=1}^k \mathbb{V}(\widehat{\beta}_j) &= \sum_{j=1}^k \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n \phi_j(X_i)\right) \\
&= \sum_{j=1}^k \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\big(\phi_j(X_i)\big) \\
&\le \sum_{j=1}^k \frac{1}{n^2}\cdot nC_1 \\
&= \frac{C_1 k}{n}
\end{aligned}
$$

and by the Sobolev smoothness condition,

$$
\begin{aligned}
k^{2q}\sum_{j=k+1}^\infty \beta_j^2 &= \sum_{j=k+1}^\infty \beta_j^2 k^{2q} \\
&\le \sum_{j=k+1}^\infty \beta_j^2 j^{2q} \\
&\le C_2,
\end{aligned}
$$

for some $C_2 > 0$.
Thus,

$$R(k) \le \frac{C_1 k}{n} + \frac{C_2}{k^{2q}}.$$

Minimizing this bound over k leads to

$$k_* = \left(\frac{2qnC_2}{C_1}\right)^{\frac{1}{2q+1}}$$

and

$$\begin{aligned} R(k_*) &\le \frac{C_1}{n}\left(\frac{2qnC_2}{C_1}\right)^{\frac{1}{2q+1}} + C_2\left(\frac{C_1}{2qnC_2}\right)^{\frac{2q}{2q+1}} \\ &\le C_3 n^{-\frac{2q}{2q+1}} + C_4 n^{-\frac{2q}{2q+1}}, \end{aligned}$$

for some $C_3, C_4 > 0$. Therefore,

$$R(k_*) = O\left(n^{-\frac{2q}{2q+1}}\right).$$

Problem 3 [20 pts.]

Since the distribution of X_i^*, for $1, 2, \ldots, n$ given $X_1, \ldots, X_n$ is uniform on the discrete set of n observations, by definition of conditional expectation, we have:

$$\begin{aligned}
\mathbb{E}(\overline{X}_n^*|X_1, X_2, \ldots, X_n) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^*|X_1, \ldots, X_n) \\
&= \mathbb{E}(X_1^*|X_1, \ldots, X_n) \\
&= \overline{X}_n.
\end{aligned}$$

By the law of total expectation,

$$\begin{aligned}
\mathbb{E}(\overline{X}_n^*) &= \mathbb{E}\big(\mathbb{E}(\overline{X}_n^*|X_1, X_2, \ldots, X_n)\big) \\
&= \mathbb{E}(\overline{X}_n) \\
&= \mu.
\end{aligned}$$

As before since given the data X_i^* is uniformly distributed on $X_1, \ldots, X_n$ we have:

$$\begin{aligned}
\mathbb{V}(\overline{X}_n^*|X_1, \ldots, X_n) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i^*|X_1, \ldots, X_n) \\
&= \frac{1}{n}\mathbb{V}(X_1^*|X_1, \ldots, X_n) \\
&= \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\overline{X}_n\right)^2\right) \\
&= \frac{S_n^2}{n},
\end{aligned}$$

where the last equality follows from part (a) and the fact that

$$\mathbb{E}((X_1^*)^2|X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n}X_i^2.$$

By the law of total variance,

$$\begin{aligned}
\mathbb{V}(\overline{X}_n^*) &= \mathbb{V}(\mathbb{E}(\overline{X}_n^*|X_1, X_2, \ldots, X_n)) + \mathbb{E}(\mathbb{V}(\overline{X}_n^*|X_1, \ldots, X_n)) \\
&= \mathbb{V}(\overline{X}_n) + \frac{1}{n}\mathbb{E}(S_n^2) \\
&= \frac{\sigma^2}{n} + \frac{(n-1)\sigma^2}{n^2} \\
&= \frac{2\sigma^2}{n} - \frac{\sigma^2}{n^2}.
\end{aligned}$$

Problem 4 [20 pts.]

(a) We have that $X_1, ..., X_n \sim \text{Bernoulli}(p)$, such that $X_i \in \{0, 1\}$. Therefore the empirical distribution $\hat{F}_n(x)$ will put mass $\hat{p}$ on 1, and $1 - \hat{p}$ on 0, i.e. $X_1^*, ..., X_n^* | X_1, ..., X_n \sim \text{Bernoulli}(\hat{p})$. Therefore, since $n\hat{p}^* = \sum_{i=1}^n X_i^*$, we have $n\hat{p}^* | X_1, ..., X_n \sim \text{Binomial}(n, \hat{p})$.

(b)

$$\mathbb{V}(\hat{p}^* | X_1, \ldots, X_n) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^n X_i^* | X_1, \ldots, X_n\right) = \frac{1}{n}\mathbb{V}(X_1^* | X_1, \ldots, X_n) = \frac{1}{n}\hat{p}(1 - \hat{p})$$

(c) Since $\hat{p}$ is the MLE of p we know that:

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}} \rightsquigarrow N(0, 1)$$

Similarly, given $X_1, \ldots, X_n$ the MLE of $\hat{p}$ for the sample $X_1^*, \ldots, X_n^*$ is $\hat{p}^*$. Thus

$$\frac{\sqrt{n}(\hat{p}^* - \hat{p})}{\sqrt{\hat{p}(1-\hat{p})}} | X_1, ..., X_n \rightsquigarrow N(0, 1)$$

Notice that in the Bernoulli case, the nonparametric bootstrap is equivalent to the parametric bootstrap.