Homework 8 - November 17, 2016. Questions.

Intermediate Statistics (Carnegie Mellon University)

Homework 8
Due Thursday Nov 17 by 3:00

1. Let $X_1, \ldots, X_n \sim p$ where $X_i \in [0,1]$. Let $\widehat{p}$ be the kernel density estimator with bandwidth h where $h \to 0$ as $n \to \infty$. In class, we showed that the bias of $\widehat{p}(x)$ is $O(h^2)$ for any $x \in (0,1)$. Show that the bias of $\widehat{p}(0)$ is C for some $C > 0$. Hence, the bias is larger at the boundary.

2. (Density Estimation Using Orthogonal Series.) Let $X_1, \ldots, X_n \sim p$ where p is a density on $[0,1]$. Let $\phi_1, \phi_2, \ldots,$ be an orthonormal series of functions. This means that
$$\int \phi_j^2(x)dx = 1 \ \text{ for all } j, \qquad \int \phi_j(x)\phi_k(x)dx = 0 \ \text{ for } j \neq k.$$
Assuming that $\int p^2(x)dx < \infty$, we can write $p(x) = \sum_{j=1}^\infty \beta_j \phi_j(x)$ where $\beta_j = \int_0^1 \phi_j(x)p(x)dx$. Suppose that p satisfies the following smoothness condition:
$$\sum_{j=1}^\infty \beta_j^2 j^{2q} < \infty$$
for some $q > 1/2$. (This is called a Sobolev smoothness condition.) Define the following estimator:
$$\widehat{p}(x) = \sum_{j=1}^k \widehat{\beta}_j \phi_j(x)$$
where
$$\widehat{\beta}_j = \frac{1}{n}\sum_{i=1}^n \phi_j(X_i).$$
Show that
$$R(k) = \mathbb{E}\Big[\int_0^1 (\widehat{p}(x) - p(x))^2 dx\Big] \leq \frac{C_1 k}{n} + \frac{C_2}{k^{2q}}.$$
What is the optimal value k_* of k? What is $R(k_*)$?

3. Let $X_1, \ldots, X_n \sim P$ and let $\mu = \mathbb{E}(X_i)$ and $\sigma^2 = V(X_i)$. Let $X_1^*, \ldots, X_n^*$ denote a bootstrap sample and let $\overline{X}_n^* = \frac{1}{n}\sum_{i=1}^n X_i^*$. Find: $\mathbb{E}(\overline{X}_n^* | X_1, \ldots, X_n)$, $\mathbb{E}(\overline{X}_n^*)$, $\text{Var}(\overline{X}_n^* | X_1, \ldots, X_n)$ and $\text{Var}(\overline{X}_n^*)$.

4. Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. Let $\widehat{p} = \frac{1}{n}\sum_{i=1}^n X_i$. Let $X_1^*, \ldots, X_n^*$ denote a bootstrap sample and let $\widehat{p}^* = \frac{1}{n}\sum_{i=1}^n X_i^*$.

   (a) What is the exact distribution of $n\widehat{p}^*$, conditional on $X_1, \ldots, X_n$?

   (b) Find an explicit expression for the bootstrap variance. That is, find $\mathsf{Var}(\widehat{p}^* | X_1, \ldots, X_n)$.

   (c) What is the asymptotic distribution of $\sqrt{n}(\widehat{p} - p)$? What is the asymptotic distribution of $\sqrt{n}(\widehat{p}^* - \widehat{p}) \mid X_1, \ldots, X_n$?