Homework 8

Name: Shashank Singh[1]
36-705 Intermediate Statistics
Due: Thursday, November 6, 2014

1. (a) By definition of C_n, $\mathbb{P}[\theta_0 \notin C_n | \theta = \theta_0] = \alpha$, and hence the test has type I error α. ∎

   (b) By definition of C_n,

$$\mathbb{P}_\theta[\theta \in C_n] = \mathbb{P}_\theta[\phi(\theta, X_1, \dots, X_n) = 0] = 1 - \mathbb{P}_\theta[\phi(\theta, X_1, \dots, X_n) = 1] = 1 - \alpha. \quad \blacksquare$$

2. (a) For $j \in \{1, \dots, m\}$, define $p_j := \int_{B_j} p(x)\,dx$. Then, each $n_j \sim$Binomial(n, p_j), and so, for $x \in B_j$,

$$\mathbb{E}[\hat{p}(x)] = \frac{\mathbb{E}[n_j]}{nh} = \frac{np_j}{nh} = \boxed{\frac{p_j}{h},}$$

   and

$$\mathbb{V}[\hat{p}(x)] = \frac{\mathbb{V}[n_j]}{n^2h^2} = \frac{np_j(1-p_j)}{n^2h^2} = \boxed{\frac{p_j(1-p_j)}{nh^2}.}$$

   (b) For any x and y in the same bin, since $|x - y| \leq h$, by the Lipschitz condition, $p(y) = p(x) + c$, for some $|c| \leq Lh$. Integrating a constant over a interval of measure gives h, $p_j = h(p(x) + c)$. Hence, $|\mathbb{E}[\hat{p}(x)] - p(x)| = |p(x) + c - p(x)| \leq Lh$, and also, for $x \in B_j$,

$$\begin{aligned}\mathbb{V}[\hat{p}(x)] = \frac{p_j(1-p_j)}{nh^2} &\leq \frac{h(p(x) + Lh)(1 - p(x) + Lh)}{nh^2} \\ &= \frac{p(x) - p(x)^2}{nh} + \frac{p(x)L + L - Lp(x) + L^2h}{n} \leq 2\frac{p(x)}{nh},\end{aligned}$$

   for large n (since $h \leq 1$). Decomposing MSE into squared bias and variance,

$$\begin{aligned}R_n(h) &= \int_0^1 (\mathbb{E}[\hat{p}(x) - p(x)])^2 + \mathbb{V}[\hat{p}(x)]\,dx \\ &\leq \int_0^1 L^2h^2 + 2\frac{p(x)}{nh}\,dx = \boxed{L^2h^2 + \frac{2}{nh}.}\end{aligned} \tag{1}$$

   (c) Note that (1) differentiable and convex in h. Hence,

$$0 = \frac{d}{dh} L^2h^2 + \frac{2}{nh}\bigg|_{h=h_n} = 2L^2h_n - \frac{2}{nh_n^2},$$

   so $h_n = (L^2n)^{-1/3}$. Plugging this into (1) gives $R_n(h_n) \leq 3(L/n)^{2/3} \in O(n^{-2/3})$. ∎

3. (a) Applying the Lipschitz condition followed by the Glivenko-Cantelli Theorem,

$$|\theta_n - \theta| \leq L \sup_x |F_n(x) - F(x)| \to 0$$

   almost surely, and hence $\theta_n \to \theta$ in probability. ∎

[1]sss1@andrew.cmu.edu

(b) Consider Bernoulli CDFs

$$F(x) = \begin{cases} 0 & : x < 0 \\ 1 & : 0 \leq x \end{cases} \qquad \text{and} \qquad F_n(x) = \begin{cases} 0 & : x < 0 \\ 1 - 1/n & : 0 \leq x < 1 \\ 1 & : 1 \leq x \end{cases}$$

for $n \in \mathbb{N}$. Then, $T(F) = 0$ and $T(F_n) = 1$, but $\sup_x |F(x) - F_n(x)| = 1/n$, and so

$$\frac{|T(F) - T(F_n)|}{\sup_x |F(x) - F_n(x)|} = n \to \infty \quad \text{as } n \to \infty. \quad \blacksquare$$

4. (a) By definition of the bootstrap distribution, each

$$\mathbb{E}[X_i^* | X_1, \ldots, X_n] = \sum_{j=1}^{n} \frac{X_i}{n} = \overline{X}_n.$$

Hence,

$$\mathbb{E}[\overline{X}_n^* | X_1, \ldots, X_n] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i^*] = \frac{1}{n} \sum_{i=1}^{n} \overline{X}_n = \overline{X}_n. \quad \blacksquare$$

(b) By definition of the bootstrap distribution, each

$$\mathbb{V}[X_i^* | X_1, \ldots, X_n] = \frac{(X_i - \overline{X}_n)^2}{n} = s^2.$$

Hence, since, $X_1^*, \ldots, X_n^*$ are conditionally independent given $X_1, \ldots, X_n$,

$$\mathbb{V}[\overline{X}_n^* | X_1, \ldots, X_n] = \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{V}[X_i^* | X_1, \ldots, X_n] = \frac{1}{n^2} \sum_{i=1}^{n} s^2 = s^2/n. \quad \blacksquare$$