

Problem Set #4

1. Let $F_n, n \geq 1$ be a random sequence of distribution functions on the real line, and let F be a nonrandom distribution function on the real line. Suppose $F_n(x) \xrightarrow{P} F(x)$ for all continuity points x of F . Suppose further that F is continuous at $F^{-1}(1 - \alpha) = \inf\{x \in \mathbf{R} : F(x) \geq 1 - \alpha\}$. Show that $F_n^{-1}(1 - \alpha) = \inf\{x \in \mathbf{R} : F_n(x) \geq 1 - \alpha\} \xrightarrow{P} F^{-1}(1 - \alpha)$. Use this result to complete the proof of part (ii) of Theorem 1.1 in the lecture notes on subsampling.
2. Suppose $X_i, i = 1, \dots, n$ are i.i.d. Bernoulli(p) random variables with $p \in (0, 1)$. Let

$$\hat{p}_n = \frac{1}{n} \sum_{1 \leq i \leq n} X_i .$$

- (a) Show that the closed interval I_n with endpoints

$$\hat{p}_n \pm z_{1-\alpha} \left(\frac{\hat{p}_n(1 - \hat{p}_n)}{n} \right)^{1/2}$$

satisfies

$$\Pr_p\{p \in I_n\} \rightarrow 1 - 2\alpha$$

for all $p \in (0, 1)$.

- (b) Brown, Cai and Dasgupta (2001) demonstrate that the actual finite sample coverage probability of I_n may be quite poor, even for very large values of n . Show that

$$\inf_{p \in (0,1)} \Pr_p\{p \in I_n\} = 0$$

for every n .

- (c) Let $J_n(x, p)$ denote the distribution of the root

$$R_n = \sqrt{n}(\hat{p}_n - p) .$$

Show that

$$B_n = \{p \in (0, 1) : J_n^{-1}(\alpha, \hat{p}_n) \leq \sqrt{n}(\hat{p}_n - p) \leq J_n^{-1}(1 - \alpha, \hat{p}_n)\}$$

satisfies

$$\Pr_p\{p \in B_n\} \rightarrow 1 - 2\alpha$$

for all $p \in (0, 1)$. Do we still have that

$$\inf_{p \in (0,1)} \Pr_p\{p \in B_n\} = 0$$

for every n ?

- (d) Construct a set $C_n = C_n(X_1, \dots, X_n)$ (not equal to $[0, 1]$) such that

$$\Pr_p\{p \in C_n\} \geq 1 - \alpha$$

for all n and p .

3. Suppose $X_i, i = 1, \dots, n$ are i.i.d. with distribution F on \mathbf{R} and $U_i, i = 1, \dots, n$ are i.i.d. $U(0, 1)$. Let $\hat{F}_n(x)$ and $\hat{U}_n(x)$ denote the empirical c.d.f.s of the $X_i, i = 1, \dots, n$ and $U_i, i = 1, \dots, n$, respectively. Denote by $c_n(1 - \alpha)$ the $1 - \alpha$ quantile

$$\sqrt{n} \sup_{x \in [0,1]} |\hat{U}_n(x) - u| .$$

- (a) Is it true that

$$\Pr_F\{\sqrt{n} \sup_{x \in \mathbf{R}} |\hat{F}_n(x) - F(x)| \leq c_n(1 - \alpha)\} \geq 1 - \alpha$$

for all F ? Explain.

- (b) Let H and G be two c.d.f.s on the interval $[-1, 1]$. Show that if

$$\sup_{x \in [-1,1]} |H(x) - G(x)| \leq \epsilon ,$$

then

$$|\mu(H) - \mu(G)| \leq 2\epsilon .$$

- (c) Use parts (a) and (b) to show that the closed interval I_n with endpoints

$$\bar{X}_n \pm 2 \frac{c_n(1 - \alpha)}{\sqrt{n}}$$

satisfies

$$\Pr_F\{\mu(F) \in I_n\} \geq 1 - \alpha .$$

(d) Explain why part (c) does not contradict Bahadur and Savage (1956).

4. Suppose $Y_i, i = 1, \dots, B$ are exchangeable real-valued random variables, i.e., their joint distribution is invariant under permutations. Let

$$\tilde{q} = \frac{1}{B} \left(1 + \sum_{1 \leq i \leq B-1} I\{Y_i \geq Y_B\} \right) .$$

(a) Show that $\Pr\{\tilde{q} \leq u\} \leq u$ for all $0 \leq u \leq 1$. (Hint: Condition on the order statistics.)

(b) Show that \tilde{p} defined in the lecture notes on randomization tests satisfies $P\{\tilde{p} \leq u\} \leq u$ for all $0 \leq u \leq 1$ and $P \in \mathbf{P}_0$.

5. Suppose one observes data $X \sim P \in \Omega$ and wishes to test null hypotheses $H_i : P \in \omega_i, i = 1, \dots, s$. Let $\hat{p}_i, i = 1, \dots, s$ be mutually independent p -values for testing $H_i, i = 1, \dots, s$. In particular, they satisfy

$$\Pr_P\{\hat{p}_i \leq u\} \leq u \text{ for all } u \in (0, 1) \text{ and } P \in \omega_i .$$

(a) Show that the single-step testing method with cutoff $c = c(\alpha, s) = 1 - (1 - \alpha)^{1/s}$ controls the familywise error rate at level α .

(b) Is it possible to improve upon the choice of c above?

(c) Compare c from (a) with the Bonferonni cutoff $c = \alpha/s$. In particular, show that

$$\lim_{s \rightarrow \infty} \frac{c(\alpha, s)}{\alpha/s} = \frac{-\log(1 - \alpha)}{\alpha} .$$

(Hint: Use L'hospital's Rule.)

(d) For $\alpha = .05$, approximately how much larger is $c(\alpha, s)$ than α/s for large s ? Is the improvement substantial?

6. For $1 \leq j < k \leq 3$, let $H_{j,k} : \mu_j = \mu_k$. Let

$$I = \{(j, k) : 1 \leq j < k \leq 3 \text{ and } H_{j,k} \text{ true}\}$$

is true. Suppose there exists a p -value $\hat{p}_{j,k}$ for each of these three null hypotheses that satisfies

$$\Pr_P\{\hat{p}_{j,k} \leq u\} \leq u \text{ for all } u \in (0, 1) \text{ and } (j, k) \in I .$$

- (a) Is it possible for $H_{1,2}$ and $H_{2,3}$ to be true, but for $H_{1,3}$ to be false? Is it possible for $|I| = 2$? What are the possible values for $|I|$?
- (b) Let $c_1 = \alpha/3$ and $c_2 = c_3 = \alpha$. Show that it is possible to use the logical restrictions among the null hypotheses to improve upon the Holm procedure. Specifically, show that the stepdown testing procedure with this choice of critical values controls the *FWER* at level α . (Hint: Consider the different possible values for $|I|$ one at a time.)