Statistics 581

Problem Set 5

Wellner; 10/25/2017

Reading: Course Notes, Chapter 2, pages 30-40;
Ferguson, ACLST, Chapters 13 and 14, pages 87 - 100;
vdVaart, Asymp. Stat., Sections 21.1-21.2, pages 304-310.

Due: Wednesday, November 1, 2017.

Reminder: Midterm exam, Friday, November 3, 2017

1. Suppose that $X_i = \mu + \sigma_i \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with some distribution function F with $E(\epsilon_1) = 0$ and $Var(\epsilon_1) = 1 < \infty$. Consider estimators of μ of the form $T_n \equiv T_n(w) = \sum_{i=1}^n w_{ni} X_i$ where $w = w_n = (w_{n1}, \ldots, w_{nn})$ is a vector of weights with $\sum_{i=1}^n w_{ni} = 1$.
(a) Show that all the estimators $T_n(w)$ are unbiased, and that the choice of weights which minimizes $Var(T_n(w))$ is

$$w_{ni}^{opt} = \frac{1/\sigma_i^2}{\sum_{j=1}^n (1/\sigma_j^2)} \qquad \text{for } \ i = 1, \ldots, n\,. \tag{1}$$

(b) Compute $Var(T_n(w^{opt}))$ and show that $T_n(w^{opt})$ is a consistent estimator of μ if $\sum_1^n (1/\sigma_j^2) \to \infty$.
(c) Now suppose that $X_i = \mu + \sigma_i \epsilon_i$ where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. with some distribution function F with $E(\epsilon_1) = 0$ and $Var(\epsilon_1) = 1 < \infty$ as in 2(b) above. Show that

$$\sqrt{\sum_1^n (1/\sigma_i^2)}\,\big(T_n(w^{opt}) - \mu\big) \to_d N(0,1)$$

if

$$\frac{\max_{1 \le i \le n}(1/\sigma_i^2)}{\sum_{j=1}^n (1/\sigma_j^2)} \to 0\,.$$

(d) Compute $Var[T_n(w^{opt})]/Var[\overline{X}_n]$ in the case $\sigma_i^2 = A i^r$ for $r = .25,\ .50,\ .75,\ 1$ and $n = 5, 10, 20, 50, 100$, and ∞.

2. Suppose that $X_1, \ldots, X_n$ are i.i.d. with density given by $f(x) = r x^{-(r+1)} 1_{[1,\infty)(x)}$ for some $r > 1$.
(a) Compute the distribution function F of the X_i's.
(b) Compute and plot the inverse distribution function F^{-1} corresponding to F.
(c) For what values of $s > 0$ is $E|X_1|^s < \infty$?
(d) Find the distribution function of $M_n \equiv \max_{1 \le i \le n} X_i$.
(e) For what values of s is $E|M_n|^s < \infty$?
(f) Find a sequence of constants b_n so that $M_n/b_n \ \to_d$ and find the limiting distribution. [Hint: see Ferguson, ACLST, Theorem 14, page 95.]

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. random vectors with values in R^k with $E(X_1) = \mu$ and $E(X_1^T X_1) < \infty$ so that $\Sigma = E(X_1 - \mu)(X_1 - \mu)^T$ is well-defined. Thus

$$Z_n \equiv \sqrt{n}(\overline{X}_n - \mu) \to_d Z \sim N_k(0, \Sigma) .$$

Suppose that $g : R^k \to R$ is a function, and suppose that $\nabla g = (g')^T$ exists at μ. Then the delta-method (or g' theorem) tells us that

$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \to_d \nabla g(\mu)^T Z \sim N(0, \nabla g(\mu)^T \Sigma \nabla g(\mu)) . \tag{2}$$

(a) Show that we can strengthen (2) as follows: Suppose that $\nabla g = (g')^T$ is continuous at μ. Then $\sqrt{n}(g(\overline{X}_n) - g(\mu))$ is *asymptotically linear* at μ:

$$\begin{aligned} \sqrt{n}(g(\overline{X}_n) - g(\mu)) &= \nabla g(\mu)^T \sqrt{n}(\overline{X}_n - \mu) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i) + o_p(1) \end{aligned}$$

where

$$\psi(x) = \nabla g(\mu)^T (x - \mu)$$

which is called the *influence function* of $g(\overline{X}_n)$ as an estimator of $g(\mu)$, has mean $E\psi(X_i) = 0$ and $Var(\psi(X_i)) = \nabla g(\mu)^T \Sigma \nabla g(\mu)$.
(b) Does the result of (a) apply to the situation considered in problem 3(b) of problem set #4? If so, what is the resulting influence function?
(c) Does the result of (a) apply to Example 2.3.6 (the correlation coefficient) on pages 18-19 of the course notes?

4. (a) Write out a proof of (17) on page 27 of the Chapter 2 notes. Compare the result with what you get by combining (11) and (16).
(b) Write out a proof of the corresponding fact (38) concerning the general empirical process $\mathbb{G}_n$: $\mathbb{G}_n \to_{f.d.} \mathbb{G}$ where $\mathbb{G}_n$ and $\mathbb{G}$ are as defined on page 21 of the chapter 2 notes; i.e. for any $f_1, \ldots, f_k \in L_2(P)$, $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k)) \to_d (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))$.

5. Suppose that $X_1, \ldots, X_n$ are i.i.d. exponential(θ); i.e. with density $p_\theta(x) = \theta \exp(-\theta x) 1_{[0,\infty)}(x)$. Let $X_{(n)} = X_{n:n}$ be the largest order statistic of $X_1, \ldots, X_n$.
(a) Find constants c_n so that $Y_n = X_{(n)} - c_n \to_d Y$ for some random variable Y and find the limiting distribution of F_Y.
(b) Compute the density of Y_n and show that it converges to the density f_Y of Y.
(c) What can you conclude from the result of (b) and Scheffé's theorem (chap. 2 notes, prop. 1.14, page 9)?

6. **Optional bonus problem 1:** Van der Vaart, problem 3.5, page 34.

7. **Optional bonus problem 2:** Suppose that $X_1, \ldots, X_n$ are i.i.d. with continuous distribution function F. Let F_0 be a fixed, specified (continuous) distribution function. Suppose we want to test $H : \ F = F_0$ versus $K : \ F \neq F_0$. Consider the *Cramér - von Mises statistic* given by

$$C_n^2 \equiv \int_{-\infty}^{\infty} n(\mathbb{F}_n(x) - F_0(x))^2 dF_0(x) .$$

(a) Show that when the null hypothesis H is true

$$C_n^2 =_d \int_0^1 n(\mathbb{G}_n(t) - t)^2 dt \,,$$

where $\mathbb{G}_n$ is the empirical d.f. of n i.i.d. Uniform$(0,1)$ rv's.
(b) Show that when the null hypothesis H is true,

$$C_n^2 \to_d \int_0^1 \mathbb{U}(t)^2 dt$$

where $\mathbb{U}$ is a standard Brownian bridge process.
[Hint: Use the fact that $\mathbb{U}_n \Rightarrow \mathbb{U}$ in $(D[0,1], \|\cdot\|_\infty)$ and the continuous mapping theorem.]
(c) Suppose that the null hypothesis fails. Thus $F \neq F_0$. Show that in this case

$$n^{-1} C_n^2 \to_{a.s.} \int_{-\infty}^{\infty} (F(x) - F_0(x))^2 dF_0(x) > 0 \,,$$

and hence the test based on C_n^2 is consistent for all $F \neq F_0$.

8. **Optional bonus problem 3:** This is a continuation of the previous problem, and should be thought of in analogy with our development for the Pearson chi-square statistic.
(a) Suppose that $F = F_n$ satisfies $\sqrt{n}(F_n(x) - F_0(x)) \to g(x)$ in $L_2(F_0)$; i.e.

$$\int [\sqrt{n}(F_n(x) - F_0(x)) - g(x)]^2 dF_0(x) \to 0 \,.$$

Describe the limiting distribution of C_n^2 under the local alternatives F_n in terms of a Brownian bridge process $\mathbb{U}$ and g.
(b) Let c^2 denote the constant on the right side in Problem 5(c) above. In the set-up of that problem, show that when $F \neq F_0$ it follows that

$$\sqrt{n}(n^{-1} C_n^2 - c^2) \to_d N(0, V^2)$$

and find V^2.
[Hint: Use $\sqrt{n}(\mathbb{F}_n - F) =_d \mathbb{U}_n(F)$, $\mathbb{U}_n \Rightarrow \mathbb{U}$, and the continuous mapping theorem.]