Statistics 581

Problem Set 5

Wellner; 10/30/2002

Reading: Ferguson, ACLST, Chapters 13 and 14, pages 87 - 100;
Wellner Notes, Chapter 2, sections 4 - 6.

Due: Wednesday, November 6, 2002.

Reminder: Midterm exam, Friday, November 8, 2002.

1. Suppose that $X_i \sim \text{Bernoulli}(p_i)$, $i = 1, \ldots, n$ are independent. Show that if

$$\sum_{i=1}^{n} p_i(1-p_i) \to \infty\,, \tag{0.1}$$

then

$$\frac{\sqrt{n}(\overline{X}_n - \overline{p}_n)}{\sqrt{n^{-1}\sum_{i=1}^{n} p_i(1-p_i)}} \to_d N(0,1)\,.$$

Give one example $\{p_i\}_{i\ge 1}$ for which (0.1) holds and another example for which it fails.

2. Suppose that $X_1, \ldots, X_n$ are independent with common mean μ, but with variances $\sigma_1^2, \ldots, \sigma_n^2$ respectively.
(a) Show that $\overline{X}_n$ is a consistent estimator of μ if $\sum_{i=1}^{n} \sigma_i^2 = o(n^2)$.
(b) 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 $Var(\epsilon_1) = 1 < \infty$. Show that if

$$\max_{1\le i\le n} \sigma_i^2 \Big/ \sum_{i=1}^{n} \sigma_i^2 \to 0 \tag{0.2}$$

then with $\overline{\sigma}_n^2 \equiv n^{-1}\sum_1^n \sigma_i^2$,

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\overline{\sigma}_n} \to_d N(0,1)\,.$$

Hence show that if both (0.2) and

$$\overline{\sigma_n^2} \to \text{“something”} \equiv \sigma_0^2\,, \tag{0.3}$$

then

$$\sqrt{n}(\overline{X}_n - \mu) \to_d N(0, \sigma_0^2) .$$

(c) Show that (0.2) holds but that (0.3) fails if $\sigma_i^2 = Ai^r$ with $r < 1$. Hence show that in this case $n^{(1-r)/2}(\overline{X}_n - \mu) = O_p(1)$.

3. Suppose that $X_1, \ldots, X_n$ are independent with common mean μ, but with variances $\sigma_1^2, \ldots, \sigma_n^2$ respectively, exactly as in problem 2 above. 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

$$w_{ni}^{opt} = \frac{1/\sigma_i^2}{\sum_{j=1}^n (1/\sigma_j^2)} \qquad \text{for } \; i = 1, \ldots, n . \tag{0.4}$$

(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

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

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 $\sigma_i^2 = Ai^r$ for $r = .25, \; .50, \; .75, \; 1$ and $n = 5, 10, 20, 50, 100$, and ∞.

4. 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 $\nabla g = \dot{g}$ exists at μ. Then the delta-method (or g' theorem) tells us that

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

Show that we can strengthen this as follows: Suppose that $\nabla g = \dot{g}$ 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 $Var(\psi(X_i)) = \nabla g(\mu)^T \Sigma \nabla g(\mu)$.

5. **Optional bonus problem** Suppose that $X_1, \ldots, X_n$ are i.i.d. with continuous distribution function F. Let F_0 be a fixed, specified 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

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

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 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 $F \neq F_0$.

6. **Optional bonus problem:** This is a continuation of the previous problem. (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.]