Statistics 581

Problem Set 5

Wellner; 10/25/2006

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

Due: Wednesday, November 1, 2006.

1. 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

$$(1) \qquad \sqrt{n}(g(\overline{X}_n) - g(\mu)) \to_d \nabla g(\mu)^T Z \sim N(0, \nabla g(\mu)^T \Sigma \nabla g(\mu)) .$$

(a) Show that we can strengthen (1) 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 $Var(\psi(X_i)) = \nabla g(\mu)^T \Sigma \nabla g(\mu)$.

(b) Does the result of (a) apply to the situation considered in problem 2(a) of problem set #4? If not, formulate another result of the same type as in (a) which does apply, and use it to find the influence function of $S_n^2/\overline{X}_n$.

2. (a) Write out a proof of (10) on page 16 of the Chapter 2 notes.
(b) Write out a proof of the corresponding fact 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 $f_1, \ldots, f_k \in L_2(P)$, $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k)) \to_d (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))$.

3. Ferguson, ACILST, problem 4, page 93 (modified slightly): suppose that $X_1, \ldots, X_n$ are i.i.d. F with continuous and positive density f in neighborhoods of $F^{-1}(1/4)$, $F^{-1}(1/2)$, and $F^{-1}(3/4)$.
(a) Find the asymptotic distribution of the mid-quartile range $R_n \equiv (X_{(3n/4)} + X_{(n/4)})/2$; i.e. find the asymptotic distribution of $\sqrt{n}(R_n - r)$ where $r = (F^{-1}(3/4) + F^{-1}(1/4))/2$.
(b) Find the asymptotic distributions of the median.
(c) For a general distribution function F, the mid-quartile range and median estimate different parameters, the population mid-quartile range and the population

median respectively, but in the case of a distribution function F that is symmetric about some point μ (so $1 - F(x + \mu) = F(x - \mu)$), they both estimate the point of symmetry, μ. Compute the asymptotic relative efficiency of the mid-quartile range relative to the median when: (i) F is Cauchy(μ, σ); (ii) F is Uniform$(0, 2\mu)$.

4. 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 $F \neq F_0$.

5. **Optional bonus problem:** This is a continuation of the previous problem, and should be though 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.]