

Statistics 581
Problem Set 5
Wellner; 10/22/2014

Reading: Ferguson, ACLST, Chapters 13 and 14, pages 87 - 100;
Wellner Notes, Chapter 2, sections 5 - 7.

Due: Wednesday, October 29, 2014.

1. Suppose that X_1, \dots, X_n are i.i.d. $\text{Cauchy}(0, 1)$; so the density of each X_i with respect to Lebesgue measure on R is $f(x) = \pi^{-1}(1 + x^2)^{-1}$, $x \in R$.
 - (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 $r > 0$ is $E|X_1|^r < \infty$?
 - (d) Find the distribution function of $M_n \equiv \max_{1 \leq i \leq n} X_i$.
 - (e) For what values of r is $E|M_n|^r < \infty$?
 - (f) Find a sequence of constants b_n so that $M_n/b_n \rightarrow_d$ and find the limiting distribution. [Hint: see Ferguson, ACLST, Theorem 14, page 95.]
2. Suppose that X_1, \dots, 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

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

Suppose that $g : R^k \rightarrow R$ is a function, and suppose that $\nabla g = \dot{g}$ exists at μ . Then the delta-method (or g' theorem) tells us that

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

(a) Show that we can strengthen (1) as follows: Suppose that $\nabla g = \dot{g}$ is continuous at μ . Then $\sqrt{n}(g(\bar{X}_n) - g(\mu))$ is *asymptotically linear* at μ :

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

where

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

which is called the *influence function* of $g(\bar{X}_n)$ 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(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/\bar{X}_n .

3. (a) Write out a direct proof of (17) on page 27 of the Chapter 2 notes, using the multivariate CLT, and compare with the result of using (10) and (16).
(b) Write out a proof of the corresponding fact concerning the general empirical process \mathbb{G}_n : $\mathbb{G}_n \rightarrow_{f.d.} \mathbb{G}$ where \mathbb{G}_n and \mathbb{G} are as defined on pages 30-31 of the chapter 2 notes; i.e. for any $f_1, \dots, f_k \in L_2(P)$, $(\mathbb{G}_n(f_1), \dots, \mathbb{G}_n(f_k)) \rightarrow_d (\mathbb{G}(f_1), \dots, \mathbb{G}(f_k))$ where \mathbb{G} is a P -Brownian bridge process with mean 0 and Covariance $E\{\mathbb{G}(f)\mathbb{G}(g)\} = P(fg) - P(f)P(g)$ as in (39) on page 32.

4. Suppose that X_1, \dots, 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 if the null hypothesis H holds (so that F_0 is the true distribution function of the X_i 's), then

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

where \mathbb{G}_n is the empirical d.f. of n i.i.d. $\text{Uniform}(0, 1)$ rv's.

- (b) Show that when the null hypothesis is true,

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

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

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

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 thought of in analogy with our development for the Pearson chi-square statistic.

- (a) Suppose that $F = F_n$ satisfies $\sqrt{n}(F_n(x) - F_0(x)) \rightarrow g(x)$ in $L_2(F_0)$; i.e.

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

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

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

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.]