

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 2, 2005.

Reminder: Midterm exam, Wednesday, November 9, 2005.

1. Suppose that X_1, \dots, X_n are independent with common mean μ , but with variances $\sigma_1^2, \dots, \sigma_n^2$ respectively.

(a) Show that \bar{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 $\epsilon_1, \dots, \epsilon_n$ are i.i.d. with some distribution function F with $E(\epsilon_1) = 0$ and $Var(\epsilon_1) = 1 < \infty$. Show that if

$$(1) \quad \max_{1 \leq i \leq n} \sigma_i^2 / \sum_{i=1}^n \sigma_i^2 \rightarrow 0$$

then with $\bar{\sigma}_n^2 \equiv n^{-1} \sum_{i=1}^n \sigma_i^2$,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{\bar{\sigma}_n^2}} \rightarrow_d N(0, 1).$$

Hence show that if both (1) and

$$(2) \quad \bar{\sigma}_n^2 \rightarrow \text{“something”} \equiv \sigma_0^2,$$

then

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

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

2. Suppose that $X_i \sim \text{Bernoulli}(p_{n,i})$, $i = 1, \dots, n$ are independent.

(a) Show that if

$$(3) \quad \sum_{i=1}^n p_{n,i}(1 - p_{n,i}) \rightarrow \infty,$$

then

$$\frac{\sqrt{n}(\bar{X}_n - \bar{p}_n)}{\sqrt{n^{-1} \sum_{i=1}^n p_{n,i}(1 - p_{n,i})}} \rightarrow_d N(0, 1).$$

Give one example $\{p_{n,i}\}_{i \geq 1}$ for which (3) holds and another example for which it fails.

(b) Compare the condition (3) to the condition for “Poisson approximation” given by $\sum_{i=1}^n p_{n,i}^2 \rightarrow 0$ given by Le Cam’s inequality (see Ferguson, ACILST, problem 5, page 18). Can both conditions hold?

3. Ferguson, ACILST, problem 3, page 93 (modified slightly): suppose that X_1, \dots, X_n are i.i.d. F with continuous and positive density f in neighborhoods of $F^{-1}(1/4)$ 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) For a symmetric distribution F , $r = r(F)$ equals the center of symmetry. In this case what is the asymptotic efficiency of R_n relative to the median as an estimator of the center of symmetry?
- (c) Determine the efficiency in (b) explicitly when F is the Cauchy distribution with center of symmetry μ and scale σ .
4. 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

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

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

5. Suppose that X_1, \dots, X_n are i.i.d. exponential(θ); i.e. with density $p_\theta(x) = \theta e^{-\theta x} 1_{[0, \infty)}(x)$. Let $X_{(n)} = X_{n:n}$ be the largest order statistic of X_1, \dots, X_n .
- (a) Find constants c_n so that $Y_n = X_{(n)} - c_n \rightarrow_d Y$ for some random variable Y and find the limiting distribution 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** Suppose that X_1, \dots, 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. $\text{Uniform}(0, 1)$ 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 $F \neq F_0$.

7. 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.]