

Statistics 581
Problem Set 5
Wellner; 10/31/2001

Reading: Ferguson, ACLST, Chapters 13 and 14, pages 87 - 100;
Wellner Notes, Chapter 2, sections 3 and 4.

Due: Wednesday, November 7, 2001.

Reminder: Midterm exam, Monday, November 5, 2001.

1. Suppose that $\underline{N}_n = (N_{11}, N_{12}, N_{21}, N_{22}) \sim \text{Mult}_4(n, \underline{p})$ where $\underline{p} = (p_{11}, p_{12}, p_{21}, p_{22})$ where $\sum_{i=1}^2 \sum_{j=1}^2 p_{ij} = 1$. (Thus \underline{N}_n is the sum of n independent $\text{Mult}_4(1, \underline{p})$ random vectors $\{\underline{Y}_i\}_{i=1}^n$.) Since there are really just three independently varying parameters for this problem, it is often useful to re-express the cell probabilities in terms of two marginal probabilities, say $p_{1\cdot} = p_{11} + p_{12}$ and $p_{\cdot 1} = p_{11} + p_{21}$, and ψ , the log of the odds-ratio, defined by

$$(0.1) \quad \psi \equiv \log \frac{p_{21}/p_{22}}{p_{11}/p_{12}} = \log \frac{p_{12}p_{21}}{p_{11}p_{22}}.$$

You may use the fact that $\psi = 0$ if and only if independence holds for the 2×2 table (i.e. $p_{ij} = p_{i\cdot}p_{\cdot j}$ for $i, j = 1, 2$).

(a) Suggest an estimator of ψ , say $\hat{\psi}$.

(b) Show that the estimator you proposed in (a) is asymptotically normal and compute the asymptotic variance of your estimator.

2. Suppose that $Y_i = \alpha + \theta'(x_i - \bar{x}) + \epsilon_i$, $i = 1, \dots, n$, where $\epsilon_i \sim (0, \sigma^2)$ are i.i.d. and the x_i 's are known vectors in R^k . Equivalently, $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ where

$$X^T = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 - \bar{x} & x_2 - \bar{x} & \cdots & x_n - \bar{x} \end{pmatrix}$$

so that X is an $n \times (k+1)$ matrix. Let $\hat{\underline{\beta}}$ be the least squares estimator of $\underline{\beta} = (\alpha, \theta)'$; i.e. $\hat{\underline{\beta}} = (X^T X)^{-1} X^T \underline{Y}$. Suppose that $n^{-1}(X^T X) \rightarrow D$ where D is positive definite.

(a) What additional condition(s) do you need to impose to prove that

$$\sqrt{n}(\hat{\underline{\beta}}_n - \underline{\beta}) \rightarrow_d N_{k+1}(0, \text{"something"})?$$

(b) Find “something” in part (a).

3. Suppose that X_1, \dots, X_n are i.i.d. F on R . Let $\mathbb{F}_n(x) = n^{-1} \sum_{i=1}^n 1_{(-\infty, x]}(X_i)$ be the empirical d.f. of the sample. Let $\alpha \in (0, 1)$. The goal of the following problem is to find a number c so that

$$(0.2) \quad P_F\{c\mathbb{F}_n(x) \leq F(x) \text{ for all } -\infty < x < \infty\} = 1 - \alpha,$$

i.e. so that $c\mathbb{F}_n(x)$ is a lower $1 - \alpha$ confidence bound for F . Let

$$A_n(c, F) = \{\mathbb{F}_n(x) \leq F(x)/c \text{ for all } -\infty < x < \infty\}.$$

(a) Show that $P_F(A_n(c, F)) = P(B_n(c))$ where, for \mathbb{G}_n the empirical distribution function of U_1, \dots, U_n i.i.d. Uniform(0, 1) random variables,

$$B_n(c) = \{\mathbb{G}_n(x) \leq x/c \text{ for all } 0 < x \leq 1\}.$$

(b) Re-express the event $B_n(c)$ in terms of the order statistics $0 \leq U_{(1)} \leq \dots \leq U_{(n)} \leq 1$ of the Uniform(0, 1) sample. [Hint: draw a picture first!]

(c) Compute $P(B_n(c))$ using the re-expression of the event $B_n(c)$ you found in (b) and the joint density of the uniform order statistics for $n = 1$, $n = 2$, and $n = 3$.

(d) Extend the calculations in (b) to a general n .

(e) Find c explicitly as a function of α and give the resulting lower confidence bound.

4. 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 $F \neq F_0$.

5. **Optional bonus problem:** This is a continuation of problem 1 above. One standard test of independence in the 2×2 table is the test based on a Pearson-type chi-square statistic.

(a) Write down the chi-square statistic Q_n for this problem, state its asymptotic distribution under the null hypothesis, and explain briefly why the claimed result holds.

(b) Suppose that the alternative hypothesis holds. Show that under the alternative hypothesis $n^{-1}Q_n \rightarrow_p$ some constant q and compute q as explicitly as possible.

(c) Can you find the asymptotic distribution of Q_n under local alternatives of the form $\psi_n = tn^{-1/2}$?

6. **Optional bonus problem:** This is a continuation of problem 4 above.

(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 3(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.]