Statistics 581

Problem Set 5

Wellner; 10/28/2009

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

Due: Friday, November 6, 2009.

Reminder: Midterm exam, Wednesday, November 4, 2009

1. Verify the following claim made in our treatment of the asymptotic distribution of the sample correlation coefficient: if

$$(X, Y) \sim N_2\left(\underline{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

then

$$\begin{pmatrix} E(X^2Y^2) - \rho^2 & E(X^3Y) - \rho & E(XY^3) - \rho \\ E(X^3Y) - \rho & E(X^4) - 1 & E(X^2Y^2) - 1 \\ E(XY^3) - \rho & E(X^2Y^2) - 1 & E(Y^4) - 1 \end{pmatrix} = \begin{pmatrix} 1+\rho^2 & 2\rho & 2\rho \\ 2\rho & 2 & 2\rho^2 \\ 2\rho & 2\rho^2 & 2 \end{pmatrix}.$$

Hint: Compute conditionally and use Theorem 1.3.5, page 14, Chapter 1.

2. 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 $\nabla g = (g')^T$ exists at μ. Then the delta-method (or g' theorem) tells us that

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

(a) Show that we can strengthen (1) as follows: Suppose that $\nabla g = (g')^T$ 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 3(b) of problem set #3? If so, what is the resulting influence function?

3. (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 $\mathbb{G}_n$: $\mathbb{G}_n \rightarrow_{f.d.} \mathbb{G}$ where $\mathbb{G}_n$ and $\mathbb{G}$ are as defined on page 21 of the chapter 2 notes; i.e. for any $f_1, \ldots, f_k \in L_2(P)$, $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k)) \rightarrow_d (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))$.

4. Suppose that $X_1, \ldots, X_n$ are i.i.d. exponential(θ); i.e. with density $p_\theta(x) = \theta \exp(-\theta x) 1_{[0,\infty)}(x)$. Let $X_{(n)} = X_{n:n}$ be the largest order statistic of $X_1, \ldots, X_n$.
(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 of 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?)

5. **Optional bonus problem 1:** Ferguson, ACILST, problem 6, page 93.