Statistics 581

Problem Set 5 Solutions, Revised

Wellner; 11/06/2009, 11/13/09

1. Verify the following claim made in our treatment of the asymptotic distribution of the sample correlation coefficient: if
$$(X, Y) \sim N_2\left(\underline{0}, \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right)\right),$$
then
$$\left(\begin{array}{ccc} E(X^2Y^2) - \rho^2 & E(X^3Y) - \rho & E(XY^3) - \rho \\ E(X^3Y) - \rho & E(X^4) - 1 & E(X^2Y^2) - 1 \\ E(XY^3) - \rho & E(X^2Y^2) - 1 & E(Y^4) - 1 \end{array}\right) = \left(\begin{array}{ccc} 1+\rho^2 & 2\rho & 2\rho \\ 2\rho & 2 & 2\rho^2 \\ 2\rho & 2\rho^2 & 2 \end{array}\right).$$

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

   Solution: First note that $(Y|X) \sim N(\rho X, 1-\rho^2)$. Thus we compute
$$\begin{aligned} E(X^2Y^2) &= E(E(X^2Y^2|X)) = E\{X^2 E(Y^2|X)\} \\ &= E\{X^2\left[Var(Y|X) + E(Y|X)^2\right]\} \\ &= E\{X^2\left[1-\rho^2+\rho^2X^2\right]\} \\ &= 1-\rho^2+\rho^2E(X^4) = 1-\rho^2+3\rho^2 = 1+2\rho^2. \end{aligned}$$

   Thus the claimed entries for the $(1,1)$, $(3,2)$ and $(2,3)$ entries of the matrix hold. Furthermore,
$$\begin{aligned} E(XY^3) &= EE(XY^3|X) = E\{XE(Y^3|X)\} \\ &= E\{XE\left[(Y-\rho X+\rho X)^3|X\right]\} \\ &= E\{X\left[(Y-\rho X)^3 + 3(Y-\rho X)^2(\rho X) + 3(Y-\rho X)(\rho X)^2 + (\rho X)^3|X\right]\} \\ &= E\{X\left[0 + 3(\rho X)(1-\rho^2) + 3(\rho X)^2\cdot 0 + \rho^3X^3\right]\} \\ &= E\{3(1-\rho^2)\rho X^2 + \rho^3X^4\} = 3\rho. \end{aligned}$$

   Thus the result for the $(3,1)$ and $(1,3)$ entries holds, and by symmetry this yields the result for the $(2,1)$ and $(1,2)$ entries. Note that the result holds for the $(2,2)$ and $(3,3)$ entries since $E(X^4) = E(Y^4) = 3$.

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 $E(X_1^TX_1) < \infty$ so that $\Sigma = E(X_1-\mu)(X_1-\mu)^T$ is well-defined. Thus
$$Z_n \equiv \sqrt{n}(\overline{X}_n - \mu) \rightarrow_d Z \sim N_k(0, \Sigma).$$

   Suppose that $g : R^k \rightarrow R$ is a function, and suppose that $\nabla g = (g')^T$ exists at μ. Then the delta-method (or g' theorem) tells us that
$$\sqrt{n}(g(\overline{X}_n) - g(\mu)) \rightarrow_d \nabla g(\mu)^T Z \sim N(0, \nabla g(\mu)^T \Sigma \nabla g(\mu)). \tag{1}$$

(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

$$\psi(x) = \nabla g(\mu)^T (x - \mu) \tag{2}$$

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?

Solution: By Taylor's theorem, for some Y_n^* satifying $|Y_n^* - \mu| \le |\overline{X}_n - \mu| \to_p 0$ it follows that

$$\begin{aligned}\sqrt{n}(g(\overline{X}_n) - g(\mu)) &= \nabla g(Y_n^*)\sqrt{n}(\overline{X}_n - \mu) \\ &= \nabla g(\mu)\sqrt{n}(\overline{X}_n - \mu) \\ &\qquad + \{\nabla g(Y_n^*) - \nabla g(\mu)\}\sqrt{n}(\overline{X}_n - \mu) \\ &= \nabla g(\mu)\sqrt{n}(\overline{X}_n - \mu) + o_p(1)\end{aligned}$$

since $\nabla g(Y_n^*) \to_p \nabla g(\mu)$ by continuity of ∇g at μ and since $\sqrt{n}(\overline{X}_n - \mu) = O_p(1)$. Now note that

$$\nabla g(\mu)\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g(\mu)(X_i - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i)$$

with ψ as in (2).

In fact, the hypothesis of continuity of ∇g can be dropped: consider a new function $h(x) = g(x) - \nabla g(\mu)x$. Then $\nabla h(\mu) = \nabla g(\mu) - \nabla g(\mu) = 0$, and we can write

$$\begin{aligned}\sqrt{n}(g(\overline{X}_n) - g(\mu) - \nabla g(\mu)(\overline{X}_n - \mu)) &= \sqrt{n}(h(\overline{X}_n) - h(\mu)) \\ &\to_d \nabla h(\mu) Z = 0 \cdot Z = 0\end{aligned} \tag{3}$$

by the delta-method applied to the function h. Since convergence in distribution to a constant implies convergence in probability to the same constant, we conclude from (3) that the left side of (3) converges in probability to 0. But this is just the claimed asymptotic linearity with $\psi(x) = \nabla g(\mu)(x - \mu)$.

(b) The result in (a) does not quite apply since

$$Z_n \equiv \sqrt{n}(\overline{X}_n - \mu, S_n^2 - \sigma^2)'$$

is not exactly an average of i.i.d. random vectors. But the key features of the proof in (a) do carry through since $n^{-1/2} Z_n \to_p 0$ and $Z_n = n^{-1/2} \sum_{i=1}^{n} \underline{Y}_i + o_p(1)$ where $\underline{Y}_i = (X_i - \mu, (X_i - \mu)^2 - \sigma^2)'$ are i.i.d. with mean 0 and finite second moment under the assumptions of problem 2(a) of problem set #4. Thus the conclusion continues to hold.

Thus with $g(u, v) = u/\sqrt{v}$

$$\begin{aligned}
\sqrt{n}\left(\frac{\overline{X}_n}{S_n} - \frac{\mu}{\sigma}\right) &= \sqrt{n}(g(\overline{X}_n, S_n^2) - g(\mu, \sigma^2)) \\
&= \nabla g(\mu, \sigma^2)\sqrt{n}(\underline{Y}_n - \underline{\mu}_Y) + o_p(1) \\
&= (-1/\sigma, -\mu/(2\sigma^3))\sqrt{n}(\underline{Y}_n - \underline{\mu}_Y) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{X_i - \mu}{\sigma} - \frac{\mu}{2\sigma}\left(\frac{(X_i - \mu)^2}{\sigma^2} - 1\right)\right\} + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(X_i) + o_p(1)
\end{aligned}$$

where

$$\psi(x) = \frac{x - \mu}{\sigma} - \frac{\mu}{2\sigma}\left(\frac{(x - \mu)^2}{\sigma^2} - 1\right)$$

has $E\psi(X_i) = 0$ and

$$\begin{aligned}
Var(\psi(X_i)) &= 1 - \frac{\mu}{\sigma}\gamma_1 + \frac{1}{4}\left(\frac{\mu}{\sigma}\right)^2\left(\frac{\mu_4}{\sigma^4} - 1\right) \\
&= 1 - r\gamma_1 + \frac{1}{4}r^2(2 + \gamma_2)
\end{aligned}$$

as in the solution of problem 3.3(b).

Here is the analogous calculation to accompany problem 4.1(a): Thus with $g(u, v) = v/u$

$$\begin{aligned}
\sqrt{n}\left(\frac{S_n^2}{\overline{X}_n} - \frac{\sigma^2}{\mu}\right) &= \nabla g(\mu, \sigma^2)\sqrt{n}(\underline{Y}_n - \underline{\mu}_Y) + o_p(1) \\
&= \frac{1}{\mu}(-\sigma^2/\mu, 1)\sqrt{n}(\underline{Y}_n - \underline{\mu}_Y) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\mu}\left\{(X_i - \mu)^2 - \sigma^2 - \frac{\sigma^2}{\mu}(X_i - \mu)\right\} + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(X_i) + o_p(1)
\end{aligned}$$

where

$$\begin{aligned}
\psi(x) &= \frac{1}{\mu}\{(x - \mu)^2 - \sigma^2 - (\sigma^2/\mu)(x - \mu)\} \\
&= \frac{\sigma^2}{\mu}\left\{\left(\frac{x - \mu}{\sigma}\right)^2 - 1 - \frac{\sigma}{\mu}\left(\frac{x - \mu}{\sigma}\right)\right\}
\end{aligned}$$

has $E\psi(X_i) = 0$ and

$$Var(\psi(X_i)) = \frac{\sigma^4}{\mu^2}\left\{2 + \gamma_2 - 2\frac{\sigma\gamma_1}{\mu} + \frac{\sigma^2}{\mu^2}\right\} = V^2$$

as in the solution of problem 4.1(a).

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 $f_1, \ldots, f_k \in L_2(P)$, $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k)) \rightarrow_d (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))$.

Solution: (a) $\mathbb{U}_n \rightarrow_{f.d.} \mathbb{U}$. To see this, let $0 < t_1 < t_2 < \cdots < t_k < 1$. Then define random vectors $\underline{Y}_i$ by

$$\underline{Y}_i = (1_{[0,t_1]}(\xi_i) - t_1, \ldots, 1_{[0,t_k]}(\xi_i) - t_k),$$

for $i = 1, \ldots, n$. Note that $E\underline{Y}_1 = 0$ and

$$\begin{aligned} E\underline{Y}_1\underline{Y}_1' &= \begin{pmatrix} t_1(1-t_1) & t_1 - t_1t_2 & \cdots & t_1 - t_1t_k \\ t_1 - t_1t_2 & t_2(1-t_2) & \cdots & t_2 - t_2t_k \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ t_1 - t_1t_k & t_2 - t_2t_k & \cdots & t_k(1-t_k) \end{pmatrix} \\ &= (t_i \wedge t_j - t_it_j)_{i,j=1}^k \equiv \Sigma. \end{aligned}$$

Thus it follows from the multivariate central limit theorem that

$$(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))' = \sqrt{n}\overline{\underline{Y}}_n \rightarrow_d N_k(0, \Sigma).$$

But for a Brownian bridge process $\mathbb{U}$, $(\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k))' \sim N_k(0, \Sigma)$, so we have shown that $(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))' \rightarrow_d (\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k))'$. But since this holds for every k and every choice of $t_1, \ldots, t_k$, it follows that $\mathbb{U}_n \rightarrow_{f.d.} \mathbb{U}$.
(b) $\mathbb{G}_n \rightarrow_{f.d.} \mathbb{G}$. To see this, let $f_1, \ldots, f_k \in L_2(P)$. Then define random vectors $\underline{Y}_i$ by

$$\underline{Y}_i = (f_1(X_i) - Pf_1, \ldots, f_k(X_i) - Pf_k)$$

for $i = 1, \ldots, n$. Note that $E\underline{Y}_i = 0$ and

$$\begin{aligned} E\underline{Y}_1\underline{Y}_1' &= \begin{pmatrix} P(f_1^2) - (P(f_1)^2 & P(f_1f_2) - Pf_1Pf_2 & \cdots & P(f_1f_k) - Pf_1Pf_k \\ P(f_1f_2) - Pf_1Pf_2 & P(f_2^2) - (Pf_2)^2 & \cdots & P(f_2f_k) - Pf_2Pf_k \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ P(f_1f_k) - Pf_1Pf_k & P(f_2f_k) - Pf_2Pf_k & \cdots & P(f_k^2) - (Pf_k)^2 \end{pmatrix} \\ &= (P(f_if_j) - Pf_iPf_j)_{i,j=1}^k \equiv \Sigma. \end{aligned}$$

Thus it follows from the multivariate central limit theorem that

$$(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k))' = \sqrt{n}\overline{\underline{Y}}_n \rightarrow_d N_k(0, \Sigma).$$

But for a $P-$Brownian bridge process $\mathbb{G}_P$, $(\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))' \sim N_k(0, \Sigma)$, so we have shown that $(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k))' \rightarrow_d (\mathbb{G}(f_1), \ldots, \mathbb{G}(f_k))'$. But since this holds for every k and every choice of $f_1, \ldots, f_k \in L_2(P)$, it follows that $\mathbb{G}_n \rightarrow_{f.d.} \mathbb{G}$.

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?)

Solution: Let $c_n = \theta^{-1} \log n = F_\theta^{-1}(1 - 1/n)$. Then

$$\begin{aligned} F_n(y) = P(Y_n \le y) &= P(X_{(n)} - c_n \le y) = P(X_{(n)} \le y + c_n) \\ &= P(X_j \le y + c_n \text{ for all } 1 \le j \le n) \\ &= P(X_1 \le y + c_n)^n = (1 - \exp(-\theta(y + c_n)))^n \\ &= \left(1 - \frac{e^{-\theta y}}{n}\right)^n \rightarrow \exp(-e^{-\theta y}) \\ &\equiv F_Y(y); \end{aligned}$$

this is an extreme - value distribution of the the "double-exponential" or "Gumbel" type; see part (c) of Theorem 14, Ferguson, ACILST page 95.
(b) The density of Y_n is found easily by differentiating in the previous display. The result is that

$$f_n(y) = (1 - n^{-1} \exp(-\theta y))^{n-1} \exp(-\theta y) \rightarrow \exp(-e^{-\theta y}) \exp(-\theta y) = f_Y(y) \equiv f(y).$$

(c) Since the densities f_n converge pointwise to the limiting density, we conclude by Scheffé's theorem that with P_n begin the probability measure on $\mathbb{R}$ corresponding to F_n and P the corresponding probability measure on $\mathbb{R}$ corresponding to F_Y,

$$d_{TV}(P_n, P) = \frac{1}{2} \int |f_n(y) - f(y)| dy \rightarrow 0.$$