Statistics 581

Problem Set 5 Solutions

Wellner; 10/30/2008

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.

Solution: First note that $(Y|X) \sim N(\rho X, 1-\rho^2)$. Thus we compute

$$\begin{aligned} E(X^2Y^2) &= E(E(X^2Y^2|X)) = E\{X^2E(Y^2|X)\} \\ &= E\{X^2\left[Var(Y|X)+E(Y|X)^2\right]\} \\ &= E\{X^2\left[1-\rho^2+\rho^2X^2\right]\} \\ &= 1-\rho^2+\rho^2E(X^4) = 1-\rho^2+3\rho^2 = 1+2\rho^2. \end{aligned}$$

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. For a random variable X taking values in $(0,\infty)$ define the *entropy* H of X by $H(X) \equiv E(X\log X) - E(X)\log(E(X))$.
(a) Show that $H(X) \geq 0$.
(b) Compute $H(X)$ and $H(X)/E(X)$ for $X \sim \text{Exp}(\lambda)$.
(c) Suppose that $X, X_1, \ldots, X_n$ are i.i.d. Propose an estimator of $H(X)$ and give conditions under which your estimator, say $\hat{H}_n$, satisfies $\hat{H}_n \to_p H(X)$.
(d) Give conditions under which the estimator you proposed in (c) satisfies $\sqrt{n}(\hat{H}_n - H(X)) \to_d N(0,V^2)$ and find V^2.

Solution: (a) Let $g(x) = x \log x$. Then $g'(x) = 1 + \log x$ and $g''(x) = 1/x > 0$. Thus g is convex, and by Jensen's inequality

$$E\{X \log X\} = Eg(X) \geq g(E(X)) = E(X)\log(E(X)).$$

This implies that $H(X) = E\{X \log X\} - E(X)\log(E(X)) \geq 0$.
(b) When $X \sim \text{Exp}(\lambda)$, we compute $E(X) = 1/\lambda$ and

$$\begin{aligned}
E\{X \log X\} &= \int_0^\infty x \log x \lambda e^{-\lambda x} dx \\
&= \int_0^\infty \frac{y}{\lambda} \log(y/\lambda) e^{-y} dy \quad \text{by letting} \quad y = \lambda x \\
&= \frac{1}{\lambda}\left\{\int_0^\infty y \log y \cdot e^{-y} dy - \log \lambda \int_0^\infty y e^{-y} dy\right\} \\
&= \frac{1}{\lambda}\{1 - \gamma - \log \lambda\}
\end{aligned}$$

where $\gamma \approx 0.577216...$ is Euler's constant,

$$\gamma = \lim_{n\to\infty}\left\{\sum_{j=1}^n \frac{1}{j} - \log n\right\} = -\int_0^\infty (\log y) e^{-y} dy.$$

Thus for $X \sim \text{Exp}(\lambda)$,

$$\begin{aligned}
&H(X) = \frac{1}{\lambda}\{1 - \gamma - \log\lambda\} - \frac{1}{\lambda}\log(1/\lambda) = \frac{1-\gamma}{\lambda}, \\
&H(X)/E(X) = 1 - \gamma.
\end{aligned}$$

(c) If $X, X_1, \ldots, X_n$ are i.i.d. with $E|X \log X| < \infty$, then, by the strong law of large numbers applied to both $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ and to $\overline{X \log X}_n = n^{-1}\sum_{i=1}^n (X_i \log X_i)$ (noting that $E|X \log X| < \infty$ implies $E|X| < \infty$), it follows that

$$\begin{aligned}
\widehat{H}_n &\equiv \frac{1}{n}\sum_{i=1}^n X_i \log(X_i) - \overline{X}_n \log(\overline{X}_n) \\
&\to_{a.s.} E(X_1 \log X_1) - E(X_1)\log(X_1) = H(X).
\end{aligned}$$

(d) Now suppose that $E(X \log X)^2 < \infty$. (Note that this implies that $E(X^2) < \infty$.) Then by the multivariate CLT it follows that

$$\sqrt{n}\begin{pmatrix} \overline{X \log X}_n - E(X \log X) \\ \overline{X}_n - E(X) \end{pmatrix} \to_d Z \sim N_2(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} E\{(X \log X)^2\} - [E(X \log X)]^2 & E\{X^2 \log X\} - E(X)E(X \log X) \\ E\{X^2 \log X\} - E(X)E(X \log X) & Var(X) \end{pmatrix}.$$

Now $\widehat{H}_n = h(\overline{X \log X}_n, \overline{X}_n)$ where $h(u, v) = u - v \log v$ has

$$\frac{\partial}{\partial u} h(u, v) = 1, \qquad \frac{\partial}{\partial v} h(u, v) = -(1 + \log v).$$

Thus by the delta-method with $\nabla h = (1, -(1+\log\mu))$, $\mu = E(X)$,

$$\sqrt{n}(\widehat{H}_n - H(X)) \to_d \nabla h \cdot Z \sim N(0, \nabla h^T \Sigma \nabla h).$$

Notes: It was not part of the problem, but when $X \sim$ Exponential(λ), I find that

$$\Sigma = \frac{1}{\lambda^2}\begin{pmatrix} 3(\log\lambda)^2 + 6(\gamma-2)\log\lambda + d & 2-\gamma-\log\lambda \\ 2-\gamma-\log\lambda & 1 \end{pmatrix},$$

where $d \equiv \pi^2 + 3\gamma^2 + 3 - 12\gamma$, and hence

$$\nabla h^T \Sigma \nabla h = \lambda^{-2}((\gamma-1)^2 + (\pi^2-9)/3) \approx (0.468615...)\lambda^{-2}.$$

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

Solution: (a) $\mathbb{U}_n \to_{f.d.} \mathbb{U}$. 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

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

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 $(\mathbb{U}_n(t_1), \ldots, \mathbb{U}_n(t_k))' \to_d (\mathbb{U}(t_1), \ldots, \mathbb{U}(t_k))'$. But since this holds for every k and every choice of $t_1, \ldots, t_k$, it follows that $\mathbb{U}_n \to_{f.d.} \mathbb{U}$.
(b) $\mathbb{G}_n \to_{f.d.} \mathbb{G}$. 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

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

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 $(\mathbb{G}_n(f_1),\ldots,\mathbb{G}_n(f_k))' \to_d (\mathbb{G}(f_1),\ldots,\mathbb{G}(f_k))'$. But since this holds for every k and every choice of $f_1,\ldots,f_k \in L_2(P)$, it follows that $\mathbb{G}_n \to_{f.d.} \mathbb{G}$.

4. Ferguson, ACILST, problem 6, page 93.

Solution: (a) Now $F_\theta(x) = x^\theta$ for $0 \le x \le 1$, so $F_\theta^{-1}(t) = t^{1/\theta}$ for $0 \le t \le 1$, and hence $m(\theta) = F_\theta^{-1}(1/2) = (1/2)^{1/\theta}$. Since

$$f_\theta(m(\theta)) = f_\theta(F_\theta^{-1}(1/2)) = \theta(1/2)^{(\theta-1)/\theta} > 0,$$

it follows from Theorem 2.6.2 that

$$\begin{aligned}
\sqrt{n}(M_n - m(\theta)) &= \sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F_\theta^{-1}(1/2)) \\
&\to_d -\frac{1}{f_\theta(F_\theta^{-1}(1/2))}\mathbb{U}(1/2) \sim N\left(0, \frac{1/4}{f_\theta^2(F_\theta^{-1}(1/2))}\right) \\
&= N\left(0, \frac{1/4}{\theta^2(1/2)^{2(\theta-1)/\theta}}\right) = N(0, \theta^{-2}2^{-2/\theta}).
\end{aligned}$$

(b) Since $M_n = \mathbb{F}_n^{-1}(1/2) \to_p F_\theta^{-1}(1/2) = (1/2)^{1/\theta}$, it follows from the continuous mapping theorem that

$$\hat{\theta}_n \equiv \frac{\log(1/2)}{\log M_n} \to_p \frac{\log(1/2)}{\log\{(1/2)^{1/\theta}\}} = \frac{\log(1/2)}{(1/\theta)\log(1/2)} = \theta.$$

(c) Let $g(x) = \log(1/2)/\log(x)$. Then

$$g'(x) = \frac{-\log(1/2)}{[\log(x)]^2}\frac{1}{x},$$

and

$$g'(m(\theta)) = g'((1/2)^{1/\theta}) = \frac{1}{(1/2)^{1/\theta}} \cdot \frac{-\log(1/2)}{[\log(1/2)^{1/\theta}]^2} = \theta^2 2^{1/\theta}/\log 2.$$

Thus it follows from the delta-method (or g'-theorem) that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta) &= \sqrt{n}(g(M_n) - g(m(\theta))) \\
&\to_d g'(m(\theta))N(0, \theta^{-2}2^{-2/\theta}) = N(0, \theta^2/(\log 2)^2).
\end{aligned}$$

5. 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 $\nabla g = \dot{g}$ 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 $\nabla g = \dot{g}$ 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 1(b) of problem set #3? If so, what is the resulting influence function?
(c) Does the result of (a) apply to the situation in Problem 4 above? If so, what is the resulting influence function?

Solution: (a) By Taylor's theorem, for some Y_n^* satisfying $|Y_n^* - \mu| \le |\overline{X}_n - \mu| \to_p 0$ 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).

(b) In problem 1(b) of problem set # 3, $X_i \sim$ Poiss(λ) are i.i.d., and $g(\lambda) \equiv p_2(\lambda) \equiv P_\lambda(X_1 = 2) = \lambda^2 e^{-\lambda}/2$, while $E(X_1) = \lambda \equiv \mu$. Thus the result of part (a) applies with $g : \mathbb{R} \to R$, and in this case

$$\begin{aligned}\sqrt{n}(\hat{p}_2 - p_2(\lambda)) &= \sqrt{n}(g(\overline{X}_n) - g(\lambda)) = g'(\lambda\sqrt{n}(\overline{X}_n - \lambda) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\lambda(X_i) + o_p(1)\end{aligned}$$

where $g'(\lambda) = \lambda(2 - \lambda)e^{-\lambda}/2$ and $\psi_\lambda(x) = g'(\lambda)(x - \lambda)$.

(c) In problem 4 above, the estimator $\hat{\theta}_n$ is expressed as $\hat{\theta}_n = g(M_n)$ where $g(x) = \log(1/2)/\log(x)$ and $M_n = \mathbb{F}_n^{-1}(1/2)$ is the sample median, so the development in

(a) above does not apply (directly). On the other hand, by the asymptotic linearity of the sample median expressed in Theorem 2.6.3, it follows that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta) &= g'(m(\theta))\sqrt{n}(M_n - m(\theta)) + o_p(1) \\
&= -g'(m(\theta))\frac{1}{f_\theta(F_\theta^{-1}(1/2))}\sqrt{n}\left(\mathbb{F}_n(F_\theta^{-1}(1/2)) - 1/2\right) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\theta(X_i) + o_p(1)
\end{aligned}$$

where

$$\psi_\theta(x) = -\frac{g'(m(\theta))}{f_\theta(F_\theta^{-1}(1/2))}(1_{(-\infty, F_\theta^{-1}(1/2)]}(x) - 1/2).$$

Thus $\hat{\theta}_n$ is asymptotically linear with influence function ψ_θ.

6. Suppose that $X_i \sim$ Bernoulli(p_i), $i = 1, \ldots, n$ are independent. Show that if

$$\sum_{i=1}^n p_i(1-p_i) \to \infty\,, \tag{3}$$

then

$$\frac{\sqrt{n}(\overline{X}_n - \overline{p}_n)}{\sqrt{n^{-1}\sum_{i=1}^n p_i(1-p_i)}} \to_d N(0,1)\,.$$

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

Solution: (a) With $Y_{ni} \equiv X_{n,i} - p_{n,i}$, $i = 1, \ldots, n, \ldots$ we have $E(Y_{ni}) = 0$, $\sigma_{ni}^2 = Var(Y_{ni}) = p_{n,i}(1-p_{n,i})$, and

$$\begin{aligned}
\gamma_{ni} &= E|Y_{ni}|^3 = E|X_{n,i} - p_{n,i}|^3 = |1-p_{n,i}|^3 p_{n,i} + |0 - p_{n,i}|^3(1-p_{n,i}) \\
&\leq p_{n,i}(1-p_{n,i})\{(1-p_{n,i})^2 + p_{n,i}^2\} \leq 2p_{n,i}(1-p_{n,i})\,,
\end{aligned}$$

so that $\sigma_n^2 = \sum_{i=1}^n p_{n,i}(1-p_{n,i})$ and $\gamma_n \leq 2\sum_{i=1}^n p_{n,i}(1-p_{n,i})$. hence

$$\frac{\gamma_n}{\sigma_n^3} \leq \frac{2}{\{\sum_{i=1}^n p_{n,i}(1-p_{n,i})\}^{1/2}} \to 0$$

if $\sum_1^n p_{n,i}(1-p_{n,i}) \to \infty$. Hence it follows from the Liapunov CLT that

$$\frac{\sum_{i=1}^n (X_{n,i} - p_{n,i})}{\sqrt{\sum_1^n p_{n,i}(1-p_{n,i})}} \to_d N(0,1)\,,$$

and this is equivalent to the stated conclusion.
If $p_{n,i} = 1/i^r$ with $r > 1$, then the assumption fails:

$$\sum_{i=1}^n p_{n,i}(1-p_{n,i}) = \sum_{i=1}^n i^{-r} - \sum_{i=1}^n i^{-2r} \to \sum_{i=1}^\infty i^{-r} - \sum_{i=1}^\infty i^{-2r} < \infty\,.$$

On the other hand, if $p_{n,i} = 1/i$, then it holds:

$$\sum_{i=1}^{n} p_{n,i}(1 - p_{n,i}) = \sum_{i=1}^{n} i^{-1} - \sum_{i=1}^{n} i^{-2} \to \infty - \sum_{i=1}^{\infty} i^{-2} = \infty\,.$$

(b) Suppose that $p_{n,i} = i^\alpha/n$, $i = 1, \ldots, n$ for some $\alpha > 0$. Then

$$\sum_{i=1}^{n} p_{n,i}^2 = \frac{1}{n^2} \sum_{i=1}^{n} i^{2\alpha} \sim \frac{1}{2\alpha + 1} n^{2\alpha - 1} \to 0$$

if $\alpha < 1/2$. On the other hand for $p_{n,i} = i^\alpha/n$ and $0 < \alpha < 1/2$

$$\begin{aligned}
\sum_{i=1}^{n} p_{n,i}(1 - p_{n,i}) &= \sum_{i=1}^{n} p_{n,i} - \sum_{i=1}^{n} p_{n,i}^2 \\
&= n^{-1} \sum_{i=1}^{n} i^\alpha - o(1) \\
&\sim \frac{1}{\alpha + 1} n^\alpha - o(1) \to \infty,
\end{aligned}$$

and thus both conditions hold for these $p_{n,i}$.