Statistics 583, Problem Set 3 Solutions

Wellner; 4/20/2016

1. Suppose that U is a random variable with a Uniform$(0,1)$ distribution. For each integer $n \geq 1$ define $X_n \equiv c_n 1_{[0,1/n]}(U) + 1$ for some sequence of real numbers c_n.
(a) Compute $E(X_n)$ (for an arbitrary sequence c_n).
(b) Show that $X_n \rightarrow_{a.s.} 1 \equiv X$ for any real sequence c_n.
(c) Find a sequence c_n so that $E\{\underline{\lim} X_n\} < \underline{\lim} E\{X_n\} < \infty$, showing that strict inequality can occur in in Fatou's inequality.
(d) For $c_n = n/\log(n+1)$, show that $E(\lim X_n) = \lim E(X_n)$ and that $\{X_n\}$ is uniformly integrable, but that the domination hypothesis of the dominated convergence theorem fails.

Solution: (a) $E(X_n) = 1 + c_n E 1\{U \leq 1/n\} = 1 + n^{-1} c_n$.
(b) Note that $P(0 < U \leq 1) = 1$. Thus with $A \equiv \{\omega \in \Omega : \ 0 < U(\omega) \leq 1\}$, for all $\omega \in A$ we have $1/n < U(\omega)$ for $n > N_\omega \geq 1/U(\omega)$, and then the indicator function $1_{[0,1/n]}(U(\omega)) = 0$. Thus $X_n(\omega) = 1$ for all $n > N_\omega$; i.e. $X_n \rightarrow_{a.s.} 1$.
(c) If $c_n = n$, then $1 = E(1) = E(\lim_n X_n) = E(\liminf X_n)$ while $\lim_n E(X_n) = 2$. If $c_n = n^{3/2}$ then $1 = E(\liminf_n X_n)$ while $\liminf_n E(X_n) = \liminf_n\{1 + \sqrt{n}\} = +\infty$.
(d) When $c_n = n/\log(n+1)$ we have $E(X_n) = 1 + 1/\log(n+1) \rightarrow 1$, so $\limsup_n E(X_n) = 1 \leq E(\lim_n X_n)$ so that $\{X_n\}$ is uniformly integrable by Vitali's theorem. On the other hand, the smallest possible random variable Y satisfying $|X_n| \leq Y$ for all n is Y defined by

$$Y = \sum_{k=1}^{\infty} \frac{k}{\log(k+1)} 1_{(1/(k+1),1/k]}(U) \, .$$

But we compute

$$\begin{aligned} E(Y) &= \sum_{k=1}^{\infty} \frac{k}{\log(k+1)} \left\{ \frac{1}{k} - \frac{1}{(k+1)} \right\} \\ &= \sum_{k=1}^{\infty} \frac{1}{\log(k+1)} \cdot \frac{1}{k+1} = \infty \, . \end{aligned}$$

Thus there is no integrable dominating function Y. Thus the domination condition of the dominated convergence theorem is sufficient but not necessary. Uniform integrability is both necessary and sufficient.

2. Suppose that $\underline{N}_n \sim \text{Mult}_k(n, \underline{p}_0)$ where $\sum_{j=1}^k p_{0,j} = 1$. Let $\widehat{\underline{p}}_n \equiv \underline{N}_n/n$.
(a) Use the multivariate CLT to show that

$$\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_0) \to_d \underline{Y} \sim N_k(0, \text{diag}(\underline{p}_0) - \underline{p}_0\underline{p}_0^T).$$

(b) Use the result in (a) to prove that $C_n \equiv \sum_{j=1}^k (N_j - np_{0,j})^2/(np_{0,j}) \to_d \chi_{k-1}^2$.
(c) Use the result in (a) to prove that $H_n^2 \equiv 4n\sum_{j=1}^k (\widehat{p}_{n,j}^{1/2} - p_{0,j}^{1/2})^2 \to_d \chi_{k-1}^2$.

Solution: (a) We can write $\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_0) = \sqrt{n}\underline{\overline{Y}}$ where $\underline{Y}_i \equiv (Y_{1,i}, \ldots, Y_{k,i}) = (Z_{1,i} - p_{0,1}, \ldots, Z_{k,i} - p_{0,k})$ where $\underline{Z}_i$, $i \in \{1, \ldots, n\}$ are the vectors of indicators with $\underline{N}_n \stackrel{d}{=} \sum_{i=1}^n \underline{Z}_i \sim \text{Mult}_k(n, \underline{p}_0)$. The $\underline{Y}_i$'s are i.i.d. vectors in $\mathbb{R}^k$ with $E(\underline{Y}_1) = \underline{0}$, $E(\underline{Y}_1^T\underline{Y}_1) \le k < \infty$ and

$$E(\underline{Y}_1\underline{Y}_1^T) = \text{diag}(\underline{p}_0) - \underline{p}_0\underline{p}_0^T \equiv \Sigma.$$

Thus by the multivariate CLT it follows that

$$\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_0) = \sqrt{n}\underline{\overline{Y}} \to_d \underline{V} \sim N_k(0, \Sigma).$$

(b) With $D \equiv \text{diag}(1/\sqrt{p_0})$ we have

$$\underline{W}_n \equiv D\sqrt{n}(\widehat{\underline{p}}_n - \underline{p}_0) \to_d D\underline{V} \equiv \underline{W} \sim N_k\left(0, I - \sqrt{\underline{p}_0}\sqrt{\underline{p}_0^T}\right).$$

Now

$$C_n = \underline{W}_n^T\underline{W}_n = \|\underline{W}_n\|^2 \to_d \|\underline{W}\|^2.$$

It remains only to show that the distribution of $\|\underline{W}\|^2$ is χ_{k-1}^2. But if Γ is an orthogonal matrix with first row $\sqrt{\underline{p}_0}$, then $\|\underline{W}\|^2 = \|\Gamma\underline{W}\|^2$ where $\Gamma\underline{W} \sim N_k(0, \Sigma_0)$ where $\Sigma_0 = \Gamma\Sigma\Gamma^T$ has 0 in the upper left corner, the $(k-1) \times (k-1)$ identity matrix in the lower right corner, and 0's elsewhere. Thus $\underline{U} \equiv \Gamma\underline{W}$ has $\underline{U}^T\underline{U} = \sum_{j=2}^k U_j^2$ where the U_j's are i.i.d. $N(0,1)$. It follows that the limiting distribution of C_n is χ_{k-1}^2.
(c) Note that by the delta-method it follows from (a) and the notation in (b)

$$\underline{S}_n \equiv 2n^{1/2}(\sqrt{\widehat{\underline{p}}_n} - \sqrt{\underline{p}_0}) \to_d D\underline{V} \equiv \underline{W}$$

where $\underline{W} \sim N_k(0, I - \sqrt{\underline{p}_0}\sqrt{\underline{p}_0^T})$. Since $H_n = \|\underline{S}_n\|^2$ is a continuous function of $\underline{S}_n$ which converges in distribution, it follows by the continuous mapping theorem and the same argument as in (b) that

$$H_n = \|\underline{S}_n\|^2 \to_d \|\underline{W}\|^2 \sim \chi_{k-1}^2.$$

3. Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of random variables with distribution functions $\{F_n\}_{n=0}^{\infty}$ and that $X_n \to_d X_0$; i.e. $F_n(x) \to F_0(x)$ for all $x \in C(F_0) \equiv \{x \in \mathbb{R} : F_0 \text{ is continuous at } x\}$. Show that this holds if and only if $Eg(X_n) = \int_{\mathbb{R}} g dF_n \to \int_{\mathbb{R}} g dF_0 = Eg(X_0)$ for all bounded and continuous functions g. [Hint: bracket $1_{(-\infty,x]}(y)$ above and below by continuous functions.]

Solution. We first show that $Eg(X_n) \to Eg(X_0)$ implies that $F_n \to_d F$. Let $f_u \equiv f_{u,x,\epsilon}$ and $f_l \equiv f_{l,x,\epsilon}$ denote the (piecewise linear) continuous functions satisfying

$$1_{(-\infty,x-\epsilon]}(y) \le f_l(y) \le 1_{(-\infty,x]}(y) \equiv h_x(y) \le f_u(y) \le 1_{(-\infty,x+\epsilon]}(y) \tag{1}$$

for all $y \in \mathbb{R}$. Then, evaluating h_x, f_u, and f_l at X_n, using $F_n(x) = Eh_x(X_n)$, and taking expectations with respect to X_n we find that

$$Ef_l(X_n) \le F_n(x) \le Ef_u(X_n).$$

This yields

$$\begin{aligned} Ef_l(X_0) = \liminf_n Ef_l(X_n) &\le \liminf_n F_n(x) \\ &\le \limsup_n F_n(x) \le \limsup_n Ef_u(X_n) = Ef_u(X_0). \end{aligned}$$

Using the outer two inequalities in (1) we deduce that

$$F_0(x-\epsilon) \le Ef_l(X_0) = \liminf_n F_n(x) \le \limsup_n F_n(x) \le Ef_u(X_0) \le F_0(x+\epsilon).$$

Now if $x \in C(F_0)$ it follows by taking the limit as $\epsilon \searrow 0$ that

$$F_0(x) \le Ef_l(X_0) = \liminf_n F_n(x) \le \limsup_n F_n(x) \le Ef_u(X_0) \le F_0(x);$$

that is, $\lim_n F_n(x) = F_0(x)$ if $x \in C(F_0)$. To argue in the other direction, let g be a bounded and continuous function on $\mathbb{R}$. Note that if $X_n \to_d X_0$, then $g(X_n) \to_d g(X_0)$. By the Skorokhod theorem there exist $X_n^* \stackrel{d}{=} X_n$, $n \ge 0$, such that $X_n^* \to_{a.s.} X_0^*$, and then $g(X_n^*) \to_{a.s.} g(X_0^*)$. Then we have

$$Eg(X_n) = Eg(X_n^*) \to Eg(X_0^*) = Eg(X_0)$$

where the convergence follows from the dominated convergence theorem.

4. (a) Show that the key condition of the Liapunov CLT implies the Lindeberg condition.
(b) Use the Cramér - Wold device together with the Lindeberg-Feller CLT to prove that for the linear regression model discussed in class on 4/11-13, the

estimators $\widehat{\alpha}_n = \overline{Y}_n$ and $\widehat{\beta}_n = \sum_{i=1}^n Y_i(x_i - \overline{x}_n)/\sum_1^n (x_i - \overline{x}_n)^2$ are jointly asymptotically normal.

Solution: (a) Fix $\epsilon > 0$ and suppose that the condition of Liapunov's theorem holds: i.e. $\gamma_n/\sigma_n^3 \to 0$ where $\gamma_n \equiv \sum_1^n \gamma_{ni}$. Then

$$\begin{aligned} LF_n(\epsilon) &\equiv \frac{1}{\sigma_n^2}\sum_{i=1}^n E\left\{X_{ni}^2 1\{|X_{ni}| > \epsilon\sigma_n\}\right\} \\ &\le \frac{1}{\sigma_n^2}\sum_{i=1}^n E\left\{|X_{ni}|^2 \frac{|X_{ni}|}{\sigma_n \epsilon} 1\{|X_{ni}| > \epsilon\sigma_n\}\right\} \\ &\le \frac{1}{\epsilon\sigma_n^3}\sum_{i=1}^n E\left\{|X_{ni}|^3\right\} = \frac{\gamma_n}{\epsilon\sigma_n^3} \\ &\to 0. \end{aligned}$$

Thus the Lindeberg condition holds.
(b) Let $\underline{a} = (a_1, a_2) \in \mathbb{R}^2$. We want to show that

$$\underline{a}^T (XX^T)^{1/2}(\underline{\widehat{\beta}}_n - \underline{\beta}) \to_d N_1(0, \sigma^2\|\underline{a}\|^2).$$

But we know that

$$\begin{aligned} \underline{a}^T (XX^T)^{1/2}(\underline{\widehat{\beta}}_n - \underline{\beta}) &= \underline{a}^T \begin{pmatrix} \sqrt{n}\overline{\epsilon}_n \\ \frac{\sum_{i=1}^n (x_i - \overline{x}_n)\epsilon_i}{\sqrt{\sum_1^n (x_i - \overline{x}_n)^2}} \end{pmatrix} \\ &= \sum_{i=1}^n \left(a_1 n^{-1/2} + \frac{a_2(x_i - \overline{x}_n)}{\sqrt{\sum_1^n (x_i - \overline{x}_n)^2}} \right) \epsilon_i \\ &\equiv \sum_{i=1}^n c_{ni}\epsilon_i \equiv \sum_{i=1}^n X_{ni} \equiv S_n \end{aligned}$$

where $E(X_{ni}) = 0$, $\sigma_{ni}^2 = Var(X_{ni}) = \sigma^2 c_{ni}^2$, and hence

$$\sigma_n^2 = \sum_{i=1}^n \sigma_{ni}^2 = \sigma^2 \sum_{i=1}^n c_{ni}^2 = \sigma^2(a_1^2 + a_2^2).$$

Note that by our hypothesis on the x_i's it follows that

$$\begin{aligned} \frac{\max_{i \le n} c_{ni}^2}{\sigma_n^2} &\le \frac{2\{a_1^2 n^{-1} + a_2^2 \max_{1 \le i \le n}(x_i - \overline{x}_n)^2 / \sum_{i=1}^n (x_i - \overline{x}_n)^2\}}{\sigma^2(a_1^2 + a_2^2)} \\ &\to 0; \end{aligned} \tag{2}$$

here we used the inequality $(a+b)^2 \le 2(a^2+b^2)$ for all real $a, b \in \mathbb{R}$. To show that $S_n/\sigma_n \to Z \sim N(0,1)$ we need to verify the Lindeberg condition. To do this, fix $\delta > 0$. Then

$$
\begin{aligned}
LF_n(\delta) &\equiv \frac{1}{\sigma_n^2}\sum_{i=1}^n E\left\{|X_{ni}|^2 1\{|X_{ni}| > \delta\sigma_n\}\right\} \\
&= \frac{1}{\sigma_n^2}\sum_{i=1}^n E\left\{c_{ni}^2\epsilon_i^2 1\{|c_{ni}||\epsilon_i| > \delta\sigma_n\}\right\} \\
&\le \frac{1}{\sigma_n^2}\sum_{i=1}^n E\left\{c_{ni}^2|\epsilon_i|^2 1_{[|\epsilon_i|>\delta/(\max_{i\le n}|c_{ni}|/\sigma_n)]}\right\} \\
&= \frac{1}{\sigma^2}E\left\{\epsilon_1^2 1_{[|\epsilon_1|>\delta/((\max_{i\le n}|c_{ni}|/\sigma_n)]}\right\} \\
&\to 0 \ \text{ by the dominated convergence theorem since } \ E(\epsilon_1^2) < \infty,
\end{aligned}
$$

and (2).

5. Suppose that Y is a random variable with $E(Y^2) < \infty$.
(a) Show that

$$Var(Y) = E\{Var(Y|X)\} + Var\{E(Y|X)\};$$

i.e.

$$E(Y-EY)^2 = E\{E[(Y-E(Y|X))^2|X]\} + E\{[E(Y|X)-E(Y)]^2\}.$$

(b) Interpret (a) geometrically. [Hint: the space of all (centered) square integrable random variables is a Hilbert space.]
(c) Suppose that $Y \sim \chi_n^2(\delta)$. Compute $E(Y)$ and $Var(Y)$. [Hint: here it helps to use the following conditional representation of the random variable Y having the noncentral chisquare distribution with n degrees of freedom and noncentrality parameter δ: $(Y|K=k) \sim \chi^2_{2k+n}$ where $K \sim \text{Poisson}(\delta/2)$. See e.g. JonW Chapter 1 notes, pages 15-16. Then use $E(Y) = E\{E(Y|K)\}$ and (a).]
(d) Show that

$$\frac{\chi_n^2(\delta) - (n+\delta)}{\sqrt{2n+4\delta}} \to_d N(0,1)$$

as either $n \to \infty$ or $\delta \to \infty$.

Solution: (a) We compute directly:

$$
\begin{aligned}
Var(Y) &= E[Y-E(Y)]^2 = E[Y-E(Y|X)+E(Y|X)-E(Y)]^2 \\
&= E[Y-E(Y|X)]^2 + 2E[(Y-E(Y|X))][E(Y|X)-E(Y)] \\
&\qquad + E[E(Y|X)-E(Y)]^2 \\
&= E\{E\{[Y-E(Y|X)]^2|X\}\} + 0 + Var[E(Y|X)] \\
&= E\{Var[Y|X]\} + Var[E(Y|X)]
\end{aligned}
$$

since, by computing conditionally,

$$
\begin{aligned}
&E[(Y - E(Y|X)][E(Y|X) - E(Y)] \\
&\quad = E\{E\{[(Y - E(Y|X)][E(Y|X) - E(Y)]|X\}\} \\
&\quad = E\{[E(Y|X) - E(Y)]E\{[Y - E(Y|X)]|X\}\} \\
&\quad = E\{[E(Y|X) - E(Y)]\{E(Y|X) - E(Y|X)\}\} \\
&\quad = E\{[E(Y|X) - E(Y)] \cdot 0\} \\
&\quad = 0\,.
\end{aligned}
$$

(b) A geometric interpretation of (a) is that $Y - E(Y|X)$ is orthogonal to $E(Y|X) - E(Y)$ in $L_2(\Omega, \mathcal{A}, P) = L_2(P)$, thus the identity in (a) can be intepreted as a "pythagorean theorem". Also note that $Y - E(Y|X)$ is orthogonal to any function $g(X)$: much as in the last part of (a)

$$
\begin{aligned}
&E\{[(Y - E(Y|X)]g(X)\} \\
&\quad = E\{E\{[(Y - E(Y|X)]g(X)|X\}\} \\
&\quad = E\{g(X)E\{[Y - E(Y|X)]|X\}\} \\
&\quad = E\{g(X)\{E(Y|X) - E(Y|X)\}\} \\
&\quad = E\{g(X) \cdot 0\} \\
&\quad = 0\,.
\end{aligned}
$$

(c) Now $(Y|K) \sim \chi^2_{2K+n}$ where $K \sim \text{Poisson}(\delta/2)$, so

$$E(Y) = E\{E(Y|K)\} = E\{2K + n\} = n + 2(\delta/2) = n + \delta\,.$$

Furthermore, using part (a) we get

$$
\begin{aligned}
Var(Y) &= E\{Var(Y|K)\} + Var\{E(Y|K)\} \\
&= E\{2(2K + n)\} + Var\{2K + n\} \\
&= 4(\delta/2) + 2n + 4(\delta/2) \\
&= 2n + 4\delta\,.
\end{aligned}
$$