Statistics 522, Problem Set 5

Wellner; 2/4/2004

1. Let $\{X_{n,i}\}$ be a triangular array of random variables, independent with each row and satisfying: (i) $\sum_{i=1}^n P(|X_{n,i}| > \epsilon) \to 0$ for each $\epsilon > 0$,
(ii) $\sum_{i=1}^n Var(X_{n,i} 1_{[|X_{n,i}| \le \epsilon]}) \to 1$ for each $\epsilon > 0$.
Show that $\sum_{i=1}^n X_{n,i} - A_n \to_d Z \sim N(0,1)$ where $A_n \equiv \sum_1^n E(X_{n,i} 1_{[|X_{n,i}| \le 1]})$. Hint: Consider truncated variables $\eta_{n,i} \equiv X_{n,i} 1_{[|X_{n,i}| \le \epsilon_n]}$ and $\xi_{n,i} \equiv \eta_{n,i} - E(\eta_{n,i})$ for an some appropriate sequence ϵ_n.

   Solution: By (i) and (ii) we can choose a sequence $\epsilon_n \to 0$ so slowly that both

$$\sum_1^n P(|X_{n,i}| > \epsilon_n) \to 0 \qquad \text{and} \qquad \sum_{i=1}^n Var(X_{n,i} 1_{[|X_{n,i}| \le \epsilon_n]}) - 1 \to 0\,.$$

   Then the recentered and truncated random variables $\xi_{n,i}$ satisfy

$$\sum_1^n E(\xi_{n,i}) = 0, \qquad \sum_1^n Var(\xi_{n,i}) \to 1,$$

   and

$$\sum_1^n E|\xi_{n,i}|^3 \le 2\epsilon_n \sum_1^n E(\xi_{n,i}^2) \to 0 \cdot 1 = 0\,.$$

   Thus by our basic theorem for triangular arrays $\sum_1^n \xi_{n,i} \to_d Z \sim N(0,1)$. Now

$$P(\sum_1^n X_{n,i} \ne \sum_{i=1}^n \eta_{n,i}) \le \sum_1^n P(|X_{n,i}| \ge \epsilon_n) \to 0\,,$$

   while

$$\begin{aligned} |\sum_1^n E(X_{n,i} 1_{[|X_{n,i}| \le \epsilon_n]}) - \sum_1^n E(X_{n,i} 1_{[|X_{n,i}| \le 1]})| &\le \sum_1^n E|X_{n,i}| 1_{[\epsilon_n < |X_{n,i}| \le 1]} \\ &\le \sum_1^n P(|X_{n,i}| > \epsilon_n) \to 0\,. \end{aligned}$$

   Putting all this together we have

$$\begin{aligned} \sum_{i=1}^n X_{n,i} - A_n &= \sum_1^n \xi_{n,i} + \sum_1^n X_{n,i} 1_{[|X_{n,i}| > \epsilon_n]} \\ &\quad + \sum_1^n E(X_{n,i} 1_{[|X_{n,i}| \le \epsilon_n]}) - \sum_1^n E(X_{n,i} 1_{[|X_{n,i}| \le 1]}) \\ &\to_d Z + 0 + 0 = Z \sim N(0,1)\,. \end{aligned}$$

2. (Liapunov's $2+\delta$ CLT). Suppose that $\{X_{n,i}\}$ is a triangular array of row-wise independent random variables satisfying:
(i) $E(X_{n,i}) = 0$ for $i = 1, \ldots, n$;
(ii) $Var(X_{n,i}) \equiv \sigma^2_{n,i} < \infty$ for $i = 1, \ldots, n$;
(iii) $\sum_1^n E|X_{n,i}|^{2+\delta}/\sigma_n^{2+\delta} \to 0$ for some $\delta > 0$ where $\sigma_n^2 \equiv \sigma^2_{n,1} + \cdots + \sigma^2_{n,n}$. Show that $S_n/\sigma_n \to_d Z \sim N(0,1)$. [The classical version of this is with $\delta = 1$.]

Solution: It suffices to show that the Lindeberg condition holds. But

$$\begin{aligned} L_n(\epsilon) &= \frac{1}{\sigma_n^2}\sum_{i=1}^n E(X^2_{n,i} 1_{[|X_{n,i}|>\epsilon\sigma_n]}) \le \frac{1}{\sigma_n^2}\sum_{i=1}^n E(X^2_{n,i}\frac{|X_{n,i}|^\delta}{\sigma_n^\delta} 1_{[|X_{n,i}|>\epsilon\sigma_n]}) \\ &\le \frac{1}{\sigma_n^{2+\delta}}\sum_1^n E|X_{n,i}|^{2+\delta} \\ &\to 0\,. \end{aligned}$$

Thus it follows that $S_n/\sigma_n \to_d Z \sim N(0,1)$ by the Lindeberg-Feller CLT.

3. Construct an example with i.i.d. random variables $X_1, X_2, \ldots$ for which the Lindeberg condition holds but for which Liapunov's $2+\delta$ condition fails for each $0 < \delta \le 1$.

Solution: For i.i.d. random variables $X_1, X_2, \ldots$ the Lindeberg condition becomes

$$L_n(\epsilon) = E(X_1^2 1_{[|X_1|>\epsilon\sqrt{n}]}) \to 0$$

as $n \to \infty$, and this holds if $E(X_1^2) < \infty$. On the other hand the Liapunov $2+\delta$ condition becomes $n^{-1-\delta/2}E|X_1|^{2+\delta} \to 0$. Thus it suffices to find a distribution for which $E(X_1^2) < \infty$ but $E|X_1|^{2+\delta} = \infty$ for every $\delta > 0$. Suppose that X has density function

$$f(x) = \frac{c}{(|x|^3(\log(|x|))^2} 1_{[e,\infty)}(|x|)$$

where the constant c is chosen so that f integrates to 1. Note that f is symmetric about zero, $E|X| < \infty$, and $E(X) = 0$ Then via the change of variable $y = \log(x)$, $dy = x^{-1}dx$,

$$\begin{aligned} E(X^2) &= 2c\int_e^\infty \frac{1}{x(\log x)^2}dx \\ &= 2c\int_1^\infty y^{-2}dy = 2c < \infty\,, \end{aligned}$$

while, on the other hand, for any $\delta > 0$

$$E|X|^{2+\delta} = 2c\int_e^\infty \frac{1}{x^{1-\delta}(\log x)^2}dx = \infty\,.$$

Note that

$$1=\int_{-\infty}^{\infty}f(x)dx=2c\int_{e}^{\infty}\frac{1}{x^3(\log x)^2}dx=2c\int_1^{\infty}y^{-2}e^{-2y}dy$$

if

$$\frac{1}{c}=2\int_1^{\infty}y^{-2}e^{-2y}dy\doteq .0750685....$$

4. Suppose that U_n is a sequence of independent random variables with $P(U_n=\pm cn)=1/(2n^2)$, $P(U_n=0)=1-1/n^2$ for some $c>0$. Let $\{Y_n\}_{n=1}^{\infty}$ be independent random variables, independent of the U_n, with mean 0 and variance 1 so that $\sqrt{n}\overline{Y}_n\to_d Z\sim N(0,1)$. Consider the independent random variables $X_n=Y_n+U_n$.
(i) Show that with $S_n=X_1+\cdots X_n$ and $\sigma_n^2=\sum_1^n Var(X_i)$, $S_n/\sigma_n\to_d aZ\sim N(0,a^2)$ with $a^2=1/(1+c^2)$.
(ii) Show that the Lindeberg condition fails.

Proof. (i) Note that $Var(X_k)=Var(Y_k)+Var(U_k)=1+c^2$, and hence $\sigma_n^2=(1+c^2)n$. Furthermore, $\sum_1^{\infty}P(|U_k|>\epsilon)=\sum_1^{\infty}k^{-2}<\infty$, so by the Borel-Cantelli lemma $P(|U_k|>\epsilon \text{ i.o.})=0$. Thus $\sqrt{n}\overline{U}_n\to_{a.s.}0$. Thus by Slutsky's theorem

$$\begin{aligned}\frac{S_n}{\sigma_n} &= \frac{\sqrt{n}\overline{Y}_n}{\sqrt{1+c^2}}+\frac{\sqrt{n}\overline{U}_n}{\sqrt{1+c^2}}\\ &\to_d \frac{Z}{\sqrt{1+c^2}}+0\sim N(0,(1+c^2)^{-1})\equiv N(0,a^2)\,.\end{aligned}$$

(ii) On the other hand, Lindeberg's condition fails. To see this, note that

$$|U_k|=|Y_k+U_k-Y_k|\le|Y_k+U_k|+|Y_k|$$

so that

$$[|U_k|>2\epsilon\sigma_n]\cap[|Y_k|\le\epsilon\sigma_n]\subset[|Y_k+U_k|>\epsilon\sigma_n]\,,$$

and

$$|U_k|^2=|Y_k+U_k-Y_k|^2\le 2(|Y_k+U_k|^2+|Y_k|^2)$$

and hence

$$|Y_k+U_k|^2\ge\frac{1}{2}|U_k|^2-|Y_k|^2\,.$$

Using these inequalities together gives

$$\begin{aligned}E(X_k^2 1_{[|X_k|>\epsilon\sigma_n]}) &= E\left\{(Y_k+U_k)^2 1_{|Y_k+U_k|>\epsilon\sigma_n]}\right\}\\ &\ge E\left\{\left(\frac{1}{2}U_k^2-Y_k^2\right)1_{[|U_k|>2\epsilon\sigma_n]}1_{[|Y_k|\le\epsilon\sigma_n]}\right\}\\ &= \frac{1}{2}E(U_k^2 1_{[|U_k|>2\epsilon\sigma_n]})P(|Y_k|\le\epsilon\sigma_n)\\ &\quad -E(Y_k^2 1_{[|Y_k|\le\epsilon\sigma_n]})P(|U_k|>2\epsilon\sigma_n) \qquad (0.1)\end{aligned}$$

where we have used independence of Y_k and U_k in the last equality. Here, since the Y_k's are i.i.d. with $E(Y_k) = 0$, $Var(Y_k) = 1$, and $\sigma_n = \sqrt{n}/a$,

$$P(|Y_k| \le \epsilon\sigma_n) \to 1, \qquad \text{and} \qquad 1 \ge E(Y_k^2 1_{[|Y_k| \le \epsilon\sigma_n]}) \to 1$$

as $n \to \infty$. Thus for n sufficiently large these are both $\ge 1/2$ (uniformly in k) and hence the right side of (0.1) is larger than

$$\frac{1}{4}E(U_k^2 1_{[|U_k|>2\epsilon\sigma_n]}) - \ P(|U_k| > 2\epsilon\sigma_n)\,.$$

Thus we see, with $a^2 \equiv 1/(1+c^2)$ and $\sigma_n^2 = n/a^2$, that

$$\begin{aligned}
L_n(\epsilon) &= \frac{a^2}{n}\sum_{k=1}^n E(X_k^2 1_{[|X_k|>\epsilon\sigma_n]}) \\
&\ge \frac{a^2/4}{n}\sum_{k=1}^n \frac{c^2k^2}{k^2} 1_{[ck>2\epsilon\sigma_n]} - \frac{a^2}{n}\sum_{k=1}^n P(|U_k| > 2\epsilon\sigma_n) \\
&\to \frac{a^2c^2}{4} - 0 > 0
\end{aligned}$$

since

$$\frac{a^2}{n}\sum_{k=1}^n P(|U_k| > 2\epsilon\sigma_n) \ = \ \frac{a^2}{n}\sum_{k=1}^n \frac{1}{k^2}1_{[ck>2\epsilon\sigma_n]} \le \frac{a^2}{n}\sum_{k=1}^\infty k^{-2} \to 0$$

as $n \to \infty$.

5. Consider the triangular array of row-wise independent random variables $\{X_{n,i}\}$ with $X_{n,1} \sim N(0, pn)$ for some $p \in (0,1)$, $X_{n,j} = 0$ for $2 \le j \le \lfloor pn \rfloor$, and $X_{n,j} \sim N(0,1)$ for $pn < j \le n$. Show that $S_n/\sigma_n \to_d Z \sim N(0,1)$ while Lindeberg's condition fails and $\max_{1\le k\le n} \sigma_{n,k}^2/\sigma_n^2 \to p > 0$.

Solution: Here

$$\sigma_n^2 = \sigma_{n,1}^2 + 0 + \cdots + 0 + \sigma_{n,[np]+1}^2 + \cdots + \sigma_{n,n}^2 = pn + 1\cdot(n - [np]) = n + np - [np]$$

so that $\sigma_n^2/n \to 1$, and

$$\begin{aligned}
\frac{S_n}{\sigma_n} &\stackrel{d}{=} \frac{\sqrt{n}}{\sigma_n}\frac{N(0,pn) + N(0, n - [np])}{\sqrt{n}} \\
&\stackrel{d}{=} \frac{\sqrt{n}}{\sigma_n}\frac{N(0, n + np - [np])}{\sqrt{n}} \\
&\stackrel{d}{=} \frac{\sqrt{n}}{\sigma_n}N(0, 1 + p - n^{-1}[np]) \\
&\to_d 1\cdot N(0,1)\,.
\end{aligned}$$

Here

$$\max_{1\le k\le n} \sigma_{n,k}^2/\sigma_n^2 = \sigma_{n,1}^2/\sigma_n^2 = pn/(n+np-[np]) \to p$$

while

$$\begin{aligned} L_n(\epsilon) &= \frac{1}{n+np-[np]}\left\{E(pnZ^2 1_{[|Z|>\epsilon\sigma_n/\sqrt{np}]}) + (n-[np])E(Z^2 1_{[Z|>\epsilon\sigma_n])}\right\} \\ &\to pE(Z^2 1_{[|Z|>\epsilon/\sqrt{p}]}) > 0\,. \end{aligned}$$