The Converse CLT

Wellner; 3/3/99

Proposition 1. Suppose that $X_1, \ldots, X_n$ are i.i.d., and let $S_n \equiv n^{-1/2} \sum_{i=1}^n X_i$. If $S_n = O_p(1)$, then $E(X_1^2) < \infty$ and $E(X_1) = 0$.

Our proof of Proposition 1 will rely on the following three lemmas.

Lemma 1. (Symmetrization.) For independent rv's $X_1, \ldots, X_n$ and $\epsilon_1, \ldots, \epsilon_n$ i.i.d. Rademacher rv's independent of the X_i's,

$$P(|n^{-1/2} \sum_{i=1}^n \epsilon_i X_i| > 2t) \le 2 \sup_n P(|n^{-1/2} \sum_{i=1}^n X_i| > t)\,. \tag{1}$$

Proof of Lemma 1. By conditioning on the Rademacher's we see that

$$\begin{aligned}
P\left(n^{-1/2} |\sum_{i=1}^n \epsilon_i X_i| > 2t\right) &\le P\left(n^{-1/2} |\sum_{i:\epsilon_i = 1} \epsilon_i X_i| + n^{-1/2} |\sum_{i:\epsilon_i=-1} \epsilon_i X_i| > 2t\right) \\
&\le E_\epsilon P_X \left(n^{-1/2} |\sum_{i:\epsilon_i=1} X_i| > t\right) \\
&\qquad + E_\epsilon P_X \left(n^{-1/2} |\sum_{i:\epsilon_i=-1} X_i| > t\right) \\
&\le 2 \sup_{k \le n} P\left(n^{-1/2} |\sum_{i=1}^k X_i| > t\right) \\
&\le 2 \sup_{k \le n} P\left(k^{-1/2} |\sum_{i=1}^k X_i| > t\right) \\
&\le 2 \sup_{1 \le k < \infty} P\left(k^{-1/2} |\sum_{i=1}^k X_i| > t\right),
\end{aligned}$$

i.e. (1) holds. $\square$

Lemma 2. (Khinchine's inequalities.) There exist constants A_p, B_p, such that, for $a = (a_1, \ldots, a_n) \in R^n$, and $p \geq 1$,

$$A_p \left\{ \sum_{i=1}^n a_i^2 \right\}^{p/2} \leq E|\sum_{i=1}^n a_i \epsilon|^p \leq B_p \left\{ \sum_{i=1}^n a_i^2 \right\}^{p/2} .$$

Recall that we proved this for $p = 1$ and found that $A_1 = 1/\sqrt{3}$ and $B_1 = 1$ work.

Lemma 3. (Paley-Zygmund inequality.) Suppose that Y is a non-negative random variable with mean EY and second moment $E(Y^2) = \|Y\|_2^2$. Then

$$P(Y > t) \geq \left(\frac{(EY - t)^+}{\|Y\|_2} \right)^2 . \tag{2}$$

Proof of Lemma 3.

$$\begin{aligned} E(Y) &= E(Y 1_{[Y \leq t]}) + E(Y 1_{[Y > t]}) \\ &\leq t + \sqrt{E(Y^2) P(Y > t)} \end{aligned}$$

by the Cauchy-Schwarz inequality. Rearranging this inequality yields (2). □

Proof of Proposition 1. The following proof is from Giné and Zinn (1994) Lemma 1 yields

$$\sup_n P(|n^{-1/2} \sum_{i=1}^n \epsilon_i X_i| > 2t) \leq 2 \sup_n P(|n^{-1/2} \sum_{i=1}^n X_i| > t) .$$

Thus tightness of $\{S_n\}$ implies that

$$\{n^{-1/2} \sum_{i=1}^n \epsilon_i X_i\} \qquad \text{is tight} .$$

By Khinchine's inequality (Lemma 2), regarding the X_i's as fixed (conditioning on the X_i's), we find that

$$E_\epsilon |n^{-1/2} \sum_{i=1}^n \epsilon_i X_i| \geq A_1 \left(n^{-1} \sum_{i=1}^n X_i^2 \right)^{1/2} \equiv c[S_n] .$$

Thus by the Paley-Zygmund inequality (Lemma 3) applied with $Y = |n^{-1/2}\sum_{i=1}^n \epsilon_i X_i|$ and the X_i's held fixed (conditioning on the X_i's)

$$\begin{aligned} P_\epsilon(|n^{-1/2}\sum_{i=1}^n \epsilon_i X_i| > t) &\geq \left(\frac{(EY-t)^+}{(E(Y^2))^{1/2}}\right)^2 \\ &\geq \left(\frac{(c[S_n]-t)^+}{[S_n]}\right)^2 \\ &= c^2\left(1 - \frac{t}{c[S_n]}\right)^2 \\ &\geq \frac{c^2}{4} 1_{[[S_n]>2t/c]} \,. \end{aligned}$$

Taking expectations across this inequality with respect to the X_i's yields

$$P(|n^{-1/2}\sum_{i=1}^n \epsilon_i X_i| > t) \geq \frac{c^2}{4} P([S_n] > 2t/c) \,.$$

It follows that the sequence $\{[S_n]\}$ is tight. Now for fixed $M \in (0,\infty)$

$$\frac{1}{n}\sum_{i=1}^n X_i^2 1_{[X_i^2 \leq M]} \to_{a.s.} E(X_1^2 1_{[X_1^2\leq M]}) \qquad \text{as} \quad n \to \infty \,.$$

Thus in particular this convergence holds in probability and in distribution. Therefore, by the Portmanteau theorem 11.7.4 (f),

$$\begin{aligned} 1_{[E(X_1^2 1_{[X_1^2\leq M]})>t]} &\leq \liminf_{n\to\infty} P(\frac{1}{n}\sum_{i=1}^n X_i^2 1_{[X_i^2\leq M]} > t) \\ &\leq \sup_n P(\frac{1}{n}\sum_{i=1}^n X_i^2 1_{[X_i^2\leq M]} > t) \,, \end{aligned}$$

so it follows that

$$\begin{aligned} \sup_{M>0} 1_{[E(X_1^2 1_{[X_1^2\leq M]})>t]} &\leq \sup_{M>0}\sup_n P(\frac{1}{n}\sum_{i=1}^n X_i^2 1_{[X_i^2\leq M]} > t) \\ &\leq \sup_n P(\frac{1}{n}\sum_{i=1}^n X_i^2 > t) \\ &= \sup_n P([S_n]^2 > t) \,. \end{aligned}$$

By the tightness of $\{[S_n]\}$, we can make the right side of the last display as small as we please; in particular there exists a number $t_0 < \infty$ such that the right side is less than $1/2$. But this implies that for this t_0 the indicator on the left side of the inequality must be zero, uniformly in M; i.e.

$$\sup_{M>0} E(X_1^2 1_{[X_1^2 \leq M]}) \leq t_0\,.$$

But the last supremum is just $E(X_1^2)$, and hence we have $E(X_1^2) \leq t_0 < \infty$.

To complete the proof, note that $E(X_1^2) < \infty$ implies that $E|X_1| < \infty$, and hence by the strong law of larg numbers we have

$$n^{-1} \sum_{i=1}^{n} X_i \rightarrow_{a.s.} E(X_1)\,.$$

But the hypothesis $n^{-1/2} \sum_{i=1}^{n} X_i = O_p(1)$ implies that

$$n^{-1} \sum_{i=1}^{n} X_i \rightarrow_p 0\,,$$

Combining these two displays yields $E(X_1) = 0$. $\square$

Giné and Zinn (1994) use similar methods to establish the corresponding theorem for U-statistics.

Theorem. (Giné and Zinn, 1994). If the sequence $\{n^{m/2} U_n(h)\}_{n=1}^{\infty}$ is tight (stochastically bounded), then $Eh^2(X_1, \ldots, X_m) < \infty$ and $Eh(X_1, x_2, \ldots, x_m) = 0$ for almost every $(x_2, \ldots, x_m) \in \mathcal{X}^{m-1}$.

References: Giné, E. and Zinn, J. (1994). A remark on convergence in distribution of U-statistics. *Ann. Probability* **22**, 117 - 125.