Central limit theorem for indicator random variables indexed by primes
Moumanti Podder

Suppose, for every \(n \in \mathbb{N}\), we define the set
\[ \mathcal{P}(n) = \{1 \leq p \leq n : p \text{ a prime}\}. \]
We now consider \(X_p : p \in \mathcal{P}(n)\), that are mutually independent indicator random variables with
\[ P[X_p = 1] = \frac{1}{p}. \] (0.1)
We first compute the mean \(\mu_p\) of \(X_p\).
\[ \mu_p = E[X_p] = \frac{1}{p}. \] (0.2)
Suppose \(\sigma_p^2 = \text{Var}[X_p]\). Then
\[ \sigma_p^2 = \text{Var}[X_p] = \frac{1}{p} - \frac{1}{p^2}. \] (0.3)
Consider
\[ Y_p = X_p - \mu_p, \] (0.4)
and define
\[ Z_n = \sum_{p \in \mathcal{P}(n)} Y_p. \] (0.5)
Suppose \(s_n^2\) denotes the variance of \(Z_n\), then
\[ s_n^2 = \text{Var}[Z_n] = \sum_{p \in \mathcal{P}(n)} \sigma_p^2 = \sum_{p \in \mathcal{P}(n)} \frac{1}{p} - \sum_{p \in \mathcal{P}(n)} \frac{1}{p^2}. \] (0.6)

We shall now try to use the Lyapunov Central Limit Theorem to examine the distributional convergence of \(Z_n\).

**Theorem 0.1.** Suppose \(X_1, X_2, \ldots\) are independent random variables such that \(E[X_n] = \mu_n\) and \(\text{Var}[X_n] = \sigma_n^2\). Define
\[ Y_n = X_n - \mu_n, \]
\[ T_n = \sum_{i=1}^{n} Y_i, \]
\[ s_n^2 = \text{Var}[T_n] = \sum_{i=1}^{n} \sigma_i^2. \]

If there exists \(\delta > 0\) such that
\[ \lim_{n \to \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} E[|Y_i|^{2+\delta}] = 0, \] (0.7)
then \(T_n/s_n \xrightarrow{d} N(0, 1)\).

Suppose we consider \(\delta = 1\) for our problem. Then
\[ E[|Y_p|^3] = E[|X_p - \mu_p|^3] \]
\[ = \left(1 - \frac{1}{p}\right)^3 P[X_p = 1] + \left(\frac{1}{p}\right)^3 P[X_p = 0] \]
\[ = \left(1 - \frac{1}{p}\right)^3 \left(\frac{1}{p}\right) + \left(\frac{1}{p}\right)^3 \left(1 - \frac{1}{p}\right) \]
\[
\begin{align*}
&= \frac{1}{p} \left( 1 - \frac{1}{p} \right) \left[ \left( 1 - \frac{1}{p} \right)^2 + \frac{1}{p^2} \right] \\
&= \frac{1}{p} \left( 1 - \frac{1}{p} \right) \left[ 1 - \frac{2}{p} + \frac{2}{p^2} \right] \\
&= \frac{1}{p} \left[ 1 - \frac{2}{p} + \frac{2}{p^2} - \frac{1}{p} + \frac{2}{p^2} - \frac{2}{p^3} \right] \\
&= \frac{1}{p} \left[ 1 - \frac{3}{p} + \frac{4}{p^2} - \frac{2}{p^3} \right] \\
&= \frac{1}{p} - \frac{3}{p^2} + \frac{4}{p^3} - \frac{2}{p^4}. \\
\end{align*}
\] (0.8)

First, note that the sum of reciprocals of primes is actually divergent, and has the following asymptotics:
\[
\sum_{p \in \mathbb{P}(n)} \frac{1}{p} \sim \log \log n, \quad \text{as } n \to \infty. \quad (0.9)
\]

But because the sum
\[
\sum_{n \in \mathbb{N}} \frac{1}{n^2} < \infty,
\]
hence the sum
\[
\sum_{p \text{ prime}} \frac{1}{p^2}
\]
also converges.

This implies that
\[
s_n^2 = \sum_{p \in \mathbb{P}(n)} \frac{1}{p} - \sum_{p \in \mathbb{P}(n)} \frac{1}{p^2} \sim \log \log n, \quad \text{as } n \to \infty. \quad (0.10)
\]

Similarly, the sums
\[
\sum_{p \text{ prime}} \frac{1}{p^3} \quad \text{and} \quad \sum_{p \text{ prime}} \frac{1}{p^4}
\]
also converge. Hence, from (0.8) we conclude that
\[
\sum_{p \in \mathbb{P}(n)} E \left[ |Y_p|^3 \right] \sim \sum_{p \in \mathbb{P}(n)} \frac{1}{p} \sim \log \log n. \quad (0.11)
\]

Hence, combining (0.10) and (0.11), we conclude that
\[
\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{p \in \mathbb{P}(n)} E \left[ |Y_p|^3 \right] = \lim_{n \to \infty} \frac{\log \log n}{(\log \log n)^{3/2}} = 0. \quad (0.12)
\]

Hence by the Lyapunov central limit theorem we can conclude that \( Z_n / s_n \xrightarrow{d} N(0,1) \).