**MATH 6480/STAT 9300/AMCS 6481, Fall 2023, Homework Set 7**

1. Let $A_n$ be independent events with probabilities $p_n$ such that $\sum_{n=1}^{\infty}(p_n - p_n^2) = \infty$. Improve on the quantitative Borel-Cantelli theorem (Theorem 1.3 of Unit 3) by stating and proving a Central limit result for $S_n := \sum_{k=1}^{n} \mathbf{1}_{A_k}$.

   <u>*solution*</u>

   Consider the triangular array $X_{n,k} = \frac{\mathbf{1}_{A_k} - p_k}{\sqrt{\sum_{k=1}^{n} p_k(1-p_k)}}$

   where $\mathbb{P}\left(\mathbf{1}_{A_k} = 1\right) = p_k$ and $\mathbb{P}\left(\mathbf{1}_{A_k} = 0\right) = 1 - p_k$, then $\mathbb{E}X_{n,k} = 0$ (the assumption of Lindeberg-Feller CLT satisfies).

   Check the first condition:

   $$\mathbb{E}X_{n,k}^2 = \operatorname{Var} X_{n,k} + \mathbb{E}^2 X_{n,k} = \operatorname{Var} X_{n,k} = \operatorname{Var}\left[\frac{\mathbf{1}_{A_k}}{\sqrt{\sum_{k=1}^{n} p_k\left(1-p_k\right)}}\right] = \frac{p_k\left(1-p_k\right)}{\sqrt{\sum_{k=1}^{n} p_k\left(1-p_k\right)}}$$

   Therefore, the first condition of Lindeberg-Feller CLT satiofies:

   $$\sum_{k=1}^{n} \mathbb{E}X_{n,k}^2 = \sigma^2 = \frac{\sum_{k=1}^{n} p_k\left(1-p_k\right)}{\sum_{k=1}^{n} p_k\left(1-p_k\right)} = 1$$

   Check the second condition:

   Given $|1_{A_k} - p_K| \leqslant |1_{A_K}| + |p_k| \leqslant 2$ bounded, we can focus on the denominator to see if the second condition satisfies. As $\sum_{k=1}^{n} p_k\left(1-p_k\right) \to \infty$ as $n \to \infty$, we can pick $N$ sufficiently large such that $\forall n > N$, there is $\sum_{k=1}^{n} p_k\left(1-p_k\right) > \frac{4}{\varepsilon^2} \quad, \forall \varepsilon > 0$ then

   $$\left|\frac{1_{A_k} - p_k}{\sqrt{\sum_{k=1}^{n} p_k\left(1-p_k\right)}}\right| \leq \frac{|1_{Ak}| + |p_k|}{\sqrt{\sum_{k=1}^{n} p_k\left(1-p_k\right)}} < \frac{2}{2/\varepsilon} = \varepsilon$$

   then $\mathbf{1}_{|X_{n,k}| > \varepsilon \to 0} \to$ as $n \to \infty$ Therefore, the second condition holds as following

   $$\forall \varepsilon > 0, \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}X_{n,k}^2 \mathbf{1}_{|X_{n,k}| > \varepsilon} = 0$$

   By Lindeberg-Feller CLT,

   $$S_n = \sum_{k=1}^{n} X_{n,k} = \frac{\sum_{k=1}^{n} \mathbf{1}_{A_k} - \sum_{k=1}^{n} p_k}{\sqrt{\sum_{k=1}^{n} p_k(1-p_k)}} \to \sigma\chi = \chi$$

   in distribution as $n \to \infty$, which is stronger and improves the quantitative BC.

2. Let $X_1, X_2, \ldots$ be IID with $\mathbf{E}X_1 = 0$ and $\mathbf{E}X_1^2 = \sigma^2 \in (0, \infty)$. Prove or find a counterexample: this implies

$$\frac{\sum_{k=1}^n X_k}{\left(\sum_{k=1}^n X_k^2\right)^{1/2}} \Rightarrow \mathcal{N}(0,1)\,.$$

<u>*solution*</u> Proof given:

i) Apply SLLN to the denominator:

Given IID $\{X_n\}$ with $\mathbb{E}\left|X_1^2\right| = \sigma^2 < \infty$, $\left\{X_n^2\right\}$ are IID as well, there is

$$\frac{\sum_{k=1}^n X_k^2}{n} \xrightarrow{\text{a.s.}} \sigma^2$$

by SLLN. Then

$$Y_n = \frac{\left(\sum_{k=1}^n X_n^2\right)^{\frac{1}{2}}}{\sigma\sqrt{n}} \xrightarrow{\text{a.s.}} 1$$

(Consider convergence a.s. as pointwise convergence, then by continuity, the square root convergence is valid. And convergence a.s. implies convergence in distribution)

ii) Apply Classical CLT to the numerator:

Given $\{X_n\}$ IID with $EX_1 = 0$ and $\operatorname{var} X_1 = \sigma^2 \in (0, \infty)$

$$X_n = \frac{\sum_{k=1}^n X_k}{\sigma\sqrt{n}} \xrightarrow{d} \chi$$

by CLT, where $\chi$ is standard normal.

By Slutsky's Theorem: there is $\frac{X_n}{Y_n} \xrightarrow{d} \frac{x}{1}$ as $n \to \infty$, i.e.,

$$\frac{\sum_{k=1}^n X_k}{\left(\sum_{k=1}^n X_k^2\right)^{\frac{1}{2}}} \xrightarrow{d} \mathcal{N}(0,1) \text{ as } n \to \infty$$

3. For each of the following eight situations, say whether there exist constants $a_n$ and $b_n$ so that $(T_n - a_n)/b_n \xrightarrow{\mathcal{D}} \chi$, where $\chi$ is a standard normal? If so, identify $a_n$ and $b_n$; **you do not need to prove these values or even that there is such a limit**, just state them. In fact sometimes a proof will be beyond your present capability and you must rely on a near-independence heuristic. If there are no such constants, say why not. You do not need a proof in this case either, just give as good a reason as you can in one sentence.

(a) Let $n$ balls be distributed uniformly and independently into $n$ boxes, and let $T_n$ be the number of balls in the first box.

(b) Let $n$ balls be distributed uniformly and independently into $\sqrt{n}$ boxes, and let $T_n$ be the number of balls in the first box.

(c) Let $n$ balls be distributed uniformly and independently into $n$ boxes, and let $T_n$ be the number of empty boxes.

(d) Let $\{X_i\}$ be i.i.d. with $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$. Let $Z_n$ be 1 if $S_n > 0$ and 0 otherwise and let $T_n = \sum_{i=1}^{n} Z_i$.

(e) Let $\{X_i\}$ be i.i.d. with $\mathbf{P}(X_i = 1) = \mathbf{P}(X_i = -1) = 1/2$. Let $T_n = \prod_{i=2}^{n}(1 + X_i/n)$.

(f) Let $X_n$ be uniform on the interval $[0, 1]$ and let $T_n$ be the sum of the first $n$ decimal digits of $X_n$.

(g) Let $T_n$ be the $n^{th}$ prime number.

(h) Let $X_n$ be i.i.d. uniform on $[0, 1]$ and let $Z_n$ be the first digit of $2^{nX_n}$. Let $T_n = \sum_{i=1}^{n} Z_i$.

<u>*solution*</u>

(a) $a_n, b_n$ do not exist. $T_n$ follows Possion distribution, $T_n \sim P(1)$, is a discrete distribution. The corresponding distribution limit of $\frac{T_n - a_n}{b_n}$ is discontinuous, there is a gap of at least $\frac{1}{b_n}$ between left and right limits, which conflicts with the continuity of $\mathcal{N}(0, 1)$. As for the potential of converging to 0 , it requires $b_n = \infty$. There is no such

$a_n$ that makes $\frac{T_n - a_n}{b_n} \to \mathcal{N}(0,1)$.

(b) $a_n = \sqrt{n}, b_n = \sqrt{\sqrt{n}(1 - \frac{1}{\sqrt{n}})}$. compare to (a), there is no costant mean that can be regarded as parameter in possion. So we need to construct a new R.V. The R.V. describing if the $i$-th ball in the first box follows bernoulli distribution. that is, $X_i \sim \text{Ber}\left(\frac{1}{n}\right), P\left(X_i = 1\right) = \frac{1}{\sqrt{n}}$ and $P\left(X_i = 0\right) = 1 - \frac{1}{\sqrt{n}}$.
$T_n = \sum_{i=1}^n X_i$ is Binomial and $a_n = \sum_{i=1}^n \frac{1}{\sqrt{n}} = \sqrt{n}$.
Use the example 1 in notes unit 07. Given $np = \sqrt{n} \to \infty$, then

$$\frac{T_n - \sqrt{n}}{\sqrt{n\left(1 - \frac{1}{\sqrt{n}}\right)}} \xrightarrow{d.} \mathcal{N}(0,1)$$

.

A rigorous proof given as following:
Consider the triangular array $X_{n \cdot k} = \left(X_k - \frac{1}{\sqrt{n}}\right)/b_n$ where $b_n = \sqrt{\sqrt{n}\left(1 - \frac{1}{\sqrt{n}}\right)}$.
Check conditions for Lindeberg - Feller CLT:
i) $\sum_{k=1}^n \mathbb{E}X_{nk}^2 = 1$
ii) $\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}X_{n,k}^2 \mathbf{1}_{|X_{n,k}| > \epsilon} = 0$ since $|X_{n,k}| \to 0$ as $n \to \infty$ (thus $\mathbf{1}_{X_{|n,k|}} > \varepsilon \to 0$ as $n \to \infty$) then

$$\frac{T_n - \sqrt{n}}{\sqrt{\sqrt{n}\left(1 - \frac{1}{\sqrt{n}}\right)}} \xrightarrow{d.} \mathcal{N}(0,1)$$

(c) $a_n = \frac{n}{e}, b_n = \sqrt{\frac{n}{e}\left(1 - \frac{1}{e}\right)}$ similar to $(b)$. construct a R.V. indicating if a box is empty. $X_i \sim \text{Ber}(p)$ where $p = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n = \frac{1}{e}$

Then use binomial converging to Poission that $T_n = \sum_{i=1}^n X_i \sim$ Binomial and then

$$\frac{Tn - \frac{n}{e}}{\sqrt{\frac{n}{e}\left(1 - \frac{1}{e}\right)}} \xrightarrow{d.} \mathcal{N}(0,1)$$

(d) $a_n, b_n$ do not exist.

The given $T_n = \sum_{i=1}^{n} Z_i \in [0, n]$. If there exists such $a_n, b_n$, then $b_n$ need to be of order $n$ for the construction limit to be tight, and $a_n$ also need to be of the same order $n$. Given $T_n, a_n, b_n$ same order. $\left|\frac{T_n - a_n}{b_n}\right| \leq c$ for some constant c. In other words. $\lim_{n\to\infty} \frac{T_n - a_n}{b_n}$ is bounded then cannot be normal

(e) $a_n = 1, b_n = \frac{1}{\sqrt{n}}$. Consider $\log T_n = \sum_{i=1}^{n} \frac{X_i}{n} + O\left(\frac{1}{n}\right)$, where $O\left(\frac{1}{n}\right)$ refers to $\exists$ a constant $C$ s.t. $C \leq \frac{1}{n}$. $\mathbb{E}\left(X_i\right) = 0$ and $\operatorname{var}\left(X_i\right) = 1 \in (0, \infty)$, then

$$Y_n = \frac{\sum_{i=1}^{\infty} X_i}{\sqrt{n}} \to \mathcal{N}(0,1) \quad \text{in dist. by CLT}$$

Then

$$\begin{aligned}
\sqrt{n}\left(T_n - 1\right) &= \sqrt{n}\left(e^{\left(\frac{\sum_{i=1}^{\infty} X_i}{\sqrt{n}}\right.} + O\left(\frac{1}{n}\right)) - 1\right) \\
&= \sqrt{n}(e^{\left(\frac{Y_n}{\sqrt{n}} + O\left(\frac{1}{n}\right)\right)} - 1) \\
&\geqslant \sqrt{n}\left(\frac{Y_n}{n} + O\left(\frac{1}{n}\right)\right) \\
&= Y_n + O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Then

$$\lim_{n\to\infty} \sqrt{n}\left(T_n - 1\right) = \lim_{n\to\infty} Y_n = \mathcal{N}(0.1)$$

Therefore, we can take $a_n = 1$ and $b_n = \frac{1}{\sqrt{n}}$

(f) $a_n = 4.5n, b_n = \sqrt{8.25n}$. let $Y_i =$ uniform $\{0, 1, \ldots, 9\}$ be the R.V. of the $i$-th decimal, then $T_n = \sum_{i=1}^{n} Y_i$. $\mathbb{E}\left(T_n\right) = n\mathbb{E}\left(Y_i\right) = 45n$. Thus $\operatorname{Var}\left(Y_i\right) = \frac{\left(10^2 - 1\right)}{12} = \frac{99}{12} = 8.25$. and then $\operatorname{Var}\left(Y_i - E\left(Y_i\right)\right) = 8.25$.
Given $\{Y_i - \mathbb{E}Y_i\}$ IID with mean zero and variance $\sigma^2 = 8.25$. Since $T_n - ET_n = \sum_{i=1}^{n}\left(Y_i - \mathbb{E}Y_i\right)$,

$$\frac{T_n - 4.5n}{\sqrt{8.25n}} \xrightarrow{d.} \mathcal{N}(0,1) \text{ by CLT}$$

(g) $a_n, b_n$ do not exist, since $T_n$ is deterministic, there is no randomness subject to.

(h) $a_n = \mu n, \quad b_n = \sigma\sqrt{n}$

$$\mathbb{P}\left(Z_n = k\right) = \mathbb{P}\left(\log_{10} k \leqslant \{nX_n \log_{10} 2\} \leqslant \log_{10}(k+1)\right).$$

where $k = \{1.2, \ldots 9\}$, and $\{nX_n \log_{10} 2\} = nX_n \log_{10} 2 - \lfloor nX_n \log_{10} 2 \rfloor$. And $\{nX_n \log_{10} 2\}$ is uniform on $[0, 1]$. Then $\mathbb{P}\left(Z_n = k\right) = \log_{10}\left(\frac{k+1}{k}\right)$.
Given $\{Z_n\}$ IID with $\mathbb{E}Z_n = \mu \neq 0$ and $\operatorname{var}\left(Z_n\right) = \sigma^2 < \infty$

$$\frac{T_n - n\mu}{\sqrt{n}\sigma} \rightarrow \mathcal{N}(0, 1) \text{ by CLT}$$

Therefore, $a_n = n\mu$ and $b_n = \sqrt{n}\sigma$