An Analytical Introduction to Probability Theory

11. Lindeberg's Central Limit Theorem

DASN, NTU https://personal.ntu.edu.sg/wptay/

11.1 Lindeberg's Method

Recall the CLT for an i.i.d. sequence from the last session: Let $X_1, X_2, \ldots$ be an i.i.d. sequence with $\mathbb{E}X_i = 0$ and $\operatorname{var} X_i = 1$. We have

$$\frac{S_n}{\sqrt{n}} \xrightarrow{\text{d}} \mathcal{N}(0,1).$$

In this session, we show another proof of this CLT using Lindeberg's method. Let $G_1, G_2, \ldots$ be i.i.d. $\mathcal{N}(0,1)$, independent of $X_1, X_2, \ldots$. Let

$$T_{m,n} = \frac{1}{\sqrt{n}}(G_1 + \ldots + G_{m-1} + X_m + \ldots + X_n)$$

with

$$\begin{aligned} T_{n+1,n} =& \frac{1}{\sqrt{n}}(G_1 + \ldots + G_n) \sim \mathcal{N}(0,1), \\ T_{1,n} =& \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n). \end{aligned}$$

We aim to show that $T_{1,n} \xrightarrow{\text{d}} T_{n+1,n}$ as $n \to \infty$. Let $f \in C_b(\mathbb{R})$ with $c_2 = \sup\left|f^{(2)}\right| < \infty$ and $c_3 = \sup\left|f^{(3)}\right| < \infty$. We have

$$\begin{aligned} &|\mathbb{E}f(T_{1,n}) - \mathbb{E}f(T_{n+1,n})| \\ &= \left|\sum_{m=1}^{n} (\mathbb{E}f(T_{m,n}) - \mathbb{E}f(T_{m+1,n}))\right| \\ &\leq \sum_{m=1}^{n} |\mathbb{E}f(T_{m,n}) - \mathbb{E}f(T_{m+1,n})|. \end{aligned}$$

Let

$$U_m = \frac{1}{\sqrt{n}}\left(G_1 + \ldots + G_{m-1} + X_{m+1} + \ldots + X_n\right).$$

Then,

$$\begin{aligned} T_{m,n} &= U_m + \frac{X_m}{\sqrt{n}}, \\ T_{m+1,n} &= U_m + \frac{G_m}{\sqrt{n}}. \end{aligned}$$

Using Taylor series expansion at U_m (see Section 10.2), for any $\epsilon > 0$, we obtain

$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}} - f''(U_m)\frac{X_m^2}{2n}\right|\mathbf{1}_{\left\{|X_m|\le\epsilon\sqrt{n}\right\}} \le \frac{c_3|X_m|^3}{6n^{3/2}}\mathbf{1}_{\left\{|X_m|\le\epsilon\sqrt{n}\right\}} \le \frac{c_3\epsilon}{6n}X_m^2,$$
$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}}\right|\mathbf{1}_{\left\{|X_m|>\epsilon\sqrt{n}\right\}} \le \frac{c_2X_m^2}{2n}\mathbf{1}_{\left\{|X_m|>\epsilon\sqrt{n}\right\}}.$$

By combining the inequality $\left|f''(U_m)\frac{X_m^2}{2n}\right| \le \frac{c_2X_m^2}{2n}$ with the second inequality above, we obtain

$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}} - f''(U_m)\frac{X_m^2}{2n}\right|\mathbf{1}_{\left\{|X_m|>\epsilon\sqrt{n}\right\}} \le \frac{c_2X_m^2}{n}\mathbf{1}_{\left\{|X_m|>\epsilon\sqrt{n}\right\}}.$$

Therefore, we have

$$\left|f(T_{m,n}) - f(U_m) - f'(U_m)\frac{X_m}{\sqrt{n}} - f''(U_m)\frac{X_m^2}{2n}\right| \le \frac{c_3\epsilon}{6n}X_m^2 + \frac{c_2\ X_m^2}{n}\mathbf{1}_{\left\{|X_m|>\epsilon\sqrt{n}\right\}}.$$

Similarly, we can obtain

$$\left|f(T_{m+1,n}) - f(U_m) - f'(U_m)\frac{G_m}{\sqrt{n}} - f''(U_m)\frac{G_m^2}{2n}\right| \le \frac{c_3\epsilon}{6n}G_m^2 + \frac{c_2G_m^2}{n}\mathbf{1}_{\left\{|G_m|>\epsilon\sqrt{n}\right\}}.$$

Furthermore, by definition, we have

$$\begin{aligned}\mathbb{E}[f'(U_m)X_m] =& \mathbb{E}[f'(U_m)]\mathbb{E}[X_m] = 0,\\ \mathbb{E}\Big[f''(U_m)X_m^2\Big] =& \mathbb{E}[f''(U_m)]\mathbb{E}\Big[X_m^2\Big] = \mathbb{E}[f''(U_m)] = \mathbb{E}\Big[f''(U_m)G_m^2\Big],\end{aligned}$$

and

$$|\mathbb{E}[f(T_{m,n})] - \mathbb{E}[f(T_{m+1,n})]| \le \frac{c_3\epsilon}{3n} + \frac{c_2}{n}\left(\mathbb{E}X_m^2\mathbf{1}_{\left\{|X_m|>\epsilon\sqrt{n}\right\}} + \mathbb{E}G_m^2\mathbf{1}_{\left\{|G_m|>\epsilon\sqrt{n}\right\}}\right).$$

Summing over $1 \le m \le n$, we have

$$|\mathbb{E}[f(T_{1,n})] - \mathbb{E}[f(T_{n+1,n})]| \le \frac{c_3\epsilon}{3} + c_2\left(\mathbb{E}X_1^2\mathbf{1}_{\left\{|X_1|>\epsilon\sqrt{n}\right\}} + \mathbb{E}G_1^2\mathbf{1}_{\left\{|G_1|>\epsilon\sqrt{n}\right\}}\right).$$

By the DCT (Theorem 4.3), we obtain

$$\begin{aligned}&\mathbb{E}X_1^2\mathbf{1}_{\left\{|X_1|>\epsilon\sqrt{n}\right\}} \to \mathbb{E}X_1^2\mathbf{1}_{\{|X_1|=\infty\}} = 0,\\ &\mathbb{E}G_1^2\mathbf{1}_{\left\{|G_1|>\epsilon\sqrt{n}\right\}} \to \mathbb{E}G_1^2\mathbf{1}_{\{|G_1|=\infty\}} = 0,\end{aligned}$$

as $n \to \infty$. Therefore, we have $|\mathbb{E}[f(T_{1,n})] - \mathbb{E}[f(T_{n+1,n})]| \to 0$ by taking $n \to \infty$ and $\epsilon \to 0$.

11.2 Lindeberg's CLT

Theorem 11.1 (Lindeberg's CLT)**.** *For each $n \geq 1$, let $(X_{n,m})_{m=1}^{n}$ be a sequence of independent random variables with $\mathbb{E}X_{n,m} = 0$. Then, $S_n = \sum_{m=1}^{n} X_{n,m} \xrightarrow{\text{d}} \mathcal{N}(0,1)$ as $n \to \infty$ if*

(i) $\sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 \to 1$ *as* $n \to \infty$*; and*

(ii) $\forall\, \epsilon > 0$, $\sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| > \epsilon\}} \to 0$ *as* $n \to \infty$.

Remark 11.1. *Let $Y_1, Y_2, \ldots$ be i.i.d. $\mathbb{E}Y_1 = 0$, $\mathbb{E}Y_1^2 = 1$, and $X_{n,m} = Y_m/\sqrt{n}$. Then $\sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 = 1$ and for all $\epsilon > 0$, we have*

$$\sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| > \epsilon\}} = \mathbb{E}Y_1^2 \mathbf{1}_{\left\{|Y_1| > \epsilon\sqrt{n}\right\}} \xrightarrow{DCT} 0.$$

Therefore, the CLT for i.i.d. sequence follows from Lindeberg's CLT.

Proof. By Chebyshev's inequality, we have

$$\mathbb{P}(|S_n| > M) \leq \frac{1}{M^2} \sum_{m=1}^{n} \mathbb{E}X_{n,m}^2 \leq \frac{2}{M^2}, \ \forall\, n \text{ sufficiently large.}$$

Therefore, $(\mathbb{P}_{S_n})_{n \geq 1}$ is uniformly tight. From Lemma 10.5, to prove the theorem is equivalent to showing that the characteristic function $\varphi_{S_n}(t) \to e^{-\frac{t^2}{2}}$ as $n \to \infty$. We have

$$\log \varphi_{S_n}(t) = \log \left(\prod_{m=1}^{n} \mathbb{E}e^{\mathrm{i}tX_{n,m}} \right) = \sum_{m=1}^{n} \log \left(1 + \mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1 \right).$$

By Taylor's series, we have the following elementary facts:

$$|\log(1+\xi) - \xi| \leq \xi^2 \text{ for } |\xi| \leq \frac{1}{2}, \tag{1}$$

$$\left| e^{\mathrm{i}a} - \sum_{k=0}^{n} \frac{(\mathrm{i}a)^k}{k!} \right| \leq \frac{|a|^{n+1}}{(n+1)!} \text{ for } a \in \mathbb{R}. \tag{2}$$

From (2) and $\mathbb{E}X_{n,m} = 0$, we have

$$\begin{aligned} & \left| \mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1 \right| \\ = & \left| \mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1 - \mathrm{i}t\mathbb{E}X_{n,m} \right| \\ \leq & \frac{t^2}{2} \mathbb{E}X_{n,m}^2 \\ \leq & \frac{t^2}{2}\epsilon^2 + \frac{t^2}{2} \mathbb{E}X_{n,m}^2 \mathbf{1}_{\{|X_{n,m}| > \epsilon\}}. \end{aligned} \tag{3}$$

Because $\mathbb{E}X_{n,m}^2\mathbf{1}_{\{|X_{n,m}>\epsilon|\}} \to 0$ as $n \to \infty$, there exists $n_0(m)$ such that for $n \geq n_0(m)$, we have

$$\left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right| \leq t^2\epsilon^2.$$

For $\epsilon \leq \dfrac{1}{t\sqrt{2}}$,

$$\left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right| \leq \frac{1}{2},$$

thus from (1) and (3), we have

$$\begin{aligned}
&\sum_{m=1}^{n}\left|\log\left(1 + (\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1)\right) - \left(\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right)\right| \\
\leq &\sum_{m=1}^{n}\left|\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right|^2 \\
\leq &\sum_{m=1}^{n}\frac{t^4}{4}\left(\mathbb{E}X_{n,m}^2\right)^2 \\
\leq &\frac{t^4}{4}\max_{1\leq m\leq n}\mathbb{E}X_{n,m}^2\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2.
\end{aligned}$$

We have

$$\max_{1\leq m\leq n}\mathbb{E}X_{n,m}^2 \leq \epsilon^2 + \max_{1\leq m\leq n}\mathbb{E}X_{n,m}^2\mathbf{1}_{\{|X_{n,m}|>\epsilon\}} \to \epsilon^2 \text{ as } n \to \infty,$$

where the convergence follows from condition (ii). Combining the above result with condition (i), we obtain

$$\sum_{m=1}^{n}\left|\log\left(1 + (\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1)\right) - \left(\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right)\right| \leq \frac{t^4}{2}\epsilon^2 \text{ as } n \to \infty. \tag{4}$$

Next we show

$$\left|\sum_{m=1}^{n}\left(\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right) + \frac{t^2}{2}\right| \to 0,$$

which, due to condition (i), is equivalent to showing that

$$\left|\sum_{m=1}^{n}\left(\mathbb{E}e^{\mathrm{i}tX_{n,m}} - 1\right) + \frac{t^2}{2}\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2\right| \to 0.$$

From the Taylor series expansion and using a similar argument in Lindeberg's method, we obtain

$$\begin{aligned}
\left|e^{\mathrm{i}tX_{n,m}} - 1 - \mathrm{i}tX_{n,m} + \frac{t^2}{2}X_{n,m}^2\right|\mathbf{1}_{\{|X_{n,m}|>\epsilon\}} \leq& t^2X_{n,m}^2\mathbf{1}_{\{|X_{n,m}|>\epsilon\}}, \\
\left|e^{\mathrm{i}tX_{n,m}} - 1 - \mathrm{i}tX_{n,m} + \frac{t^2}{2}X_{n,m}^2\right|\mathbf{1}_{\{|X_{n,m}|\leq\epsilon\}} \leq& \frac{t^3\epsilon}{6}X_{n,m}^2.
\end{aligned}$$

We then obtain

$$\begin{aligned}
&\left|\sum_{m=1}^{n}\left(\mathbb{E}e^{\mathrm{i}tX_{n,m}}-1\right)+\frac{t^2}{2}\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2\right| \\
\le & t^2\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2\mathbf{1}_{\{|X_{n,m}|>\epsilon\}}+\frac{t^3\epsilon}{6}\sum_{m=1}^{n}\mathbb{E}X_{n,m}^2 \\
\le & \frac{t^3\epsilon}{6} \text{ as } n\to\infty.
\end{aligned}$$

Taking $\epsilon \to 0$, the theorem is proved. □

Theorem 11.2 (Berry-Essen Theorem)**.** *Let $(X_m)_{m\ge 1}$ be a sequence of independent random variables, $\mathbb{E}X_m = 0$. Let $\hat{F}_n$ be the cdf of $\frac{S_n}{\sqrt{\sum_{m=1}^n \mathbb{E}X_m^2}}$ and $G \sim \mathcal{N}(0,1)$. We have*

$$\sup_x\left|\hat{F}_n(x)-G(x)\right| \le \frac{10\sum_{m=1}^n \mathbb{E}|X_m|^3}{\left(\sum_{m=1}^n \mathbb{E}X_m^2\right)^{3/2}}.$$

For the special case where (X_m) is an i.i.d. sequence and $\mathbb{E}X_1^2 = \sigma^2$, we have

$$\sup_x\left|\hat{F}_n(x)-G(x)\right| \le \frac{10\mathbb{E}|X_1|^3}{\sigma^3\sqrt{n}}.$$

This theorem shows how fast the convergence is. The proof is omitted here as it is quite technical and tedious. Please see Durrett if interested.

11.3 Kolmogorov's Three Series Theorem

Theorem 11.3 (Kolmogorov's Three Series Theorem)**.** *Let $(X_i)_{i\ge 1}$ be a sequence of independent random variables and let $Z_i = X_i\mathbf{1}_{\{|X_i|\le 1\}}$. Then, $\sum_{i=1}^n X_i$ converges a.s. if and only if*

(i) $\sum_{i\ge 1}\mathbb{P}(|X_i|>1)<\infty$*.*

(ii) $\sum_{i\ge 1}\mathbb{E}Z_i$ *converges.*

(iii) $\sum_{i\ge 1}\operatorname{var} Z_i<\infty$*.*

Proof. "$\Leftarrow$": Using condition (i), we have

$$\sum_{i\ge 1}\mathbb{P}(X_i\ne Z_i)=\sum_{i\ge 1}\mathbb{P}(|X_i|>1)<\infty.$$

From the Borel-Cantelli Lemma (Lemma 6.1), we have

$$\mathbb{P}(X_i\ne Z_i \text{ i.o.})=0.$$

Therefore,

$$\sum_{i\geq 1} X_i \text{ converges} \iff \sum_{i\geq 1} Z_i \text{ converges.}$$

Using condition (ii), we have

$$\sum_{i\geq 1} Z_i \text{ converges} \iff \sum_{i\geq 1}(Z_i - \mathbb{E}Z_i) \text{ converges.}$$

The proof now follows from condition (iii) and the variance convergence criterion (Theorem 7.4).

"$\Rightarrow$": Suppose that $\sum_{i\geq 1} X_i$ converges a.s.
We first show condition (i). Since $\mathbb{P}(|X_i| > 1 \text{ i.o.}) = 0$, the Borel-Cantelli Lemma (Lemma 6.1) yields

$$\sum_{i\geq 1} \mathbb{P}(|X_i| > 1) < \infty.$$

We next show condition (iii) by contradiction. From condition (i), we have

$$\sum_{i\geq 1} X_i \text{ converges a.s.} \implies \sum_{i\geq 1} Z_i \text{ converges a.s.} \tag{5}$$

Therefore,

$$S(m,n) = \sum_{i=m}^{n} Z_i \to 0 \text{ a.s as } m,n \to \infty.$$

Thus, for any $\delta > 0$, we have

$$\mathbb{P}(|S(m,n)| > \delta) \leq \delta \text{ for sufficiently large } m,n. \tag{6}$$

Now assume $\sum_{i\geq 1} \operatorname{var} Z_i = \infty$. Then we have $\sigma^2(m,n) \triangleq \operatorname{var}(S(m,n)) = \sum_{i=m}^{n} \operatorname{var}(Z_i) \to \infty$ as $n \to \infty$ for fixed m. Let

$$T(m,n) = \frac{S(m,n) - \mathbb{E}S(m,n)}{\sigma(m,n)} = \sum_{i=m}^{n} \frac{Z_i - \mathbb{E}Z_i}{\sigma(m,n)}.$$

Let $\widetilde{Z}_i = Z_i - \mathbb{E}Z_i$. We have $\left|\widetilde{Z}_i\right| \leq 2$, and $\left|\frac{\widetilde{Z}_i}{\sigma(m,n)}\right| \to 0$ a.s as $n \to \infty$ because $\sigma(m,n) \to \infty$. For any $\epsilon > 0$, $\exists\, n(m)$ such that for all $n \geq n(m)$, we have $\left|\frac{\widetilde{Z}_i}{\sigma(m,n)}\right| \leq \epsilon$ a.s. Therefore, we obtain

$$\sum_{i=m}^{n} \mathbb{E}\left[\frac{\widetilde{Z}_i^2}{\sigma^2(m,n)} \mathbf{1}_{\left\{\left|\frac{\widetilde{Z}_i}{\sigma(m,n)}\right| > \epsilon\right\}}\right] = 0 \text{ a.s. for } n \geq n(m).$$

Furthermore, we have

$$\sum_{m=1}^{n} \mathbb{E}\left[\frac{\widetilde{Z}_i^2}{\sigma^2(n,m)}\right] = 1.$$

By Lindeberg's CLT, we therefore have

$$T(m,n) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \text{ as } m,n \to \infty. \tag{7}$$

But at the same time, we have

$$1 - \delta \leq \mathbb{P}(|S(m,n)| \leq \delta) = \mathbb{P}\left(\left|T(m,n) + \frac{\mathbb{E}S(m,n)}{\sigma(m,n)}\right| \leq \frac{\delta}{\sigma(m,n)}\right).$$

When $m, n \to \infty$, $\sigma(m,n) \to \infty$. Thus, $T(m,n)$ concentrates at a constant, which contradicts (7). Therefore, the assumption that $\sum_{i\geq 1} \operatorname{var} Z_i = \infty$ is false.

Finally, we show condition (ii). From the variance convergence criterion (Theorem 7.4), $\sum_{i\geq 1}(Z_i - \mathbb{E}Z_i)$ converges a.s. Thus, convergence of $\sum_{i\geq 1} \mathbb{E}Z_i$ follows from (5). □

11.4 Levy's Equivalence Theorem

Theorem 11.4 (Levy's Equivalence Theorem)**.** *Let $(X_i)_{i\geq 1}$ be a sequence of independent random variables. Then, $\sum_{i\geq 1} X_i$ converges a.s. $\iff$ converges in probability $\iff$ converges in distribution.*

To prove the Theorem 11.4, we start with a few lemmas.

Lemma 11.1. *Suppose $X_1, X_2, \ldots$ are independent. Let $S_n = \sum_{i\leq n} X_i$. If $\mathbb{P}(|S_n - S_j| \geq a) \leq p < 1$ for all $j \leq n$, then for all $x > a$, we have*

$$\mathbb{P}\left(\max_{1\leq j\leq n} |S_j| \geq x\right) \leq \frac{1}{1-p}\mathbb{P}(|S_n| > x - a).$$

Proof. Let $\tau = \min\{j \leq n : |S_j| \geq x\}$ with $\tau = n+1$ if $|S_j| < x$ for all $j \leq n$. If $\tau = j$, then $|S_j| \geq x$ and

$$\{\omega : |S_n - S_j| < a,\ \tau = j\} \subset \{\omega : |S_n| > x - a,\ \tau = j\}.$$

This gives us

$$
\begin{aligned}
\mathbb{P}\Big(\max_{j\le n}|S_j|\ge x\Big) &= \mathbb{P}(\tau\le n)\\
&= \sum_{j=1}^{n}\mathbb{P}(\tau=j)\\
&\le \frac{1}{1-p}\sum_{j=1}^{n}\mathbb{P}(|S_n-S_j|<a)\mathbb{P}(\tau=j)\\
&= \frac{1}{1-p}\sum_{j=1}^{n}\mathbb{P}(|S_n-S_j|<a,\ \tau=j) \text{ since } \{\tau=j\} \text{ depends only on } X_1,\dots,X_j\\
&\le \frac{1}{1-p}\sum_{j=1}^{n}\mathbb{P}(|S_n|>x-a,\ \tau=j)\\
&= \frac{1}{1-p}\mathbb{P}(|S_n|>x-a).
\end{aligned}
$$

□

Lemma 11.2. $Y_n\to Y$ *a.s. iff* $\max_{i\ge n}|Y_i-Y|\xrightarrow{\text{p}}0$.

Proof. We have $\max_{i\ge n}|Y_i-Y|\to 0$ a. s. $\implies \max_{i\ge n}|Y_i-Y|\xrightarrow{\text{p}}0$.

To show the converse, let $M_n=\max_{i\ge n}|Y_i-Y|$, which is a decreasing sequence bounded below by 0. Therefore $M_n\downarrow M$ for some M, which implies that $\forall\,\epsilon>0$, $\mathbb{P}(M>\epsilon)\le\mathbb{P}(M_n>\epsilon)\to 0$ as $n\to\infty$. Therefore, from continuity of $\mathbb{P}$, $\mathbb{P}(M=0)=1$ and $M_n\to 0$ a.s. This is equivalent to saying that $Y_n\to Y$ a.s. □

Lemma 11.3. (Y_n) *converges in probability iff* $\lim_{n,m\to\infty}\mathbb{P}(|Y_m-Y_n|\ge\epsilon)=0$ *for all* $\epsilon>0$.

Proof. The proof in the "⇒" direction is trivial. To prove the converse, we note that the given condition implies that for all $k\ge 1$, $\exists\, n(k)$ such that $\forall\, n,m\ge n(k)$, we have

$$\mathbb{P}\Big(|Y_m-Y_n|\ge\frac{1}{2^k}\Big)\le\frac{1}{2^k}.$$

We can choose $n(k+1)\ge n(k)$ so that

$$\mathbb{P}\Big(|Y_{n(k+1)}-Y_{n(k)}|\ge\frac{1}{2^k}\Big)\le\frac{1}{2^k}.$$

Summing over $k\ge 1$, we have

$$\sum_{k\ge 1}\mathbb{P}\Big(|Y_{n(k+1)}-Y_{n(k)}|\ge\frac{1}{2^k}\Big)\le 1<\infty.$$

The Borel-Cantelli Lemma (Lemma 6.1) implies that $\mathbb{P}\Big(|Y_{n(k+1)}-Y_{n(k)}|\ge\frac{1}{2^k}\text{ i.o.}\Big)=0$ and therefore $\exists\, l$ such that for all $k\ge l$, $|Y_{n(k+1)}-Y_{n(k)}|<\frac{1}{2^k}$ a.s. and for all $j\ge k$, we have

$$|Y_{n(j)}-Y_{n(k)}|<\sum_{i\ge k}\frac{1}{2^i}\to 0 \text{ as } k\to\infty.$$

Therefore $(Y_{n(k)})_{k\geq 1}$ is Cauchy a.s., and it converges since $\mathbb{R}$ is complete. Hence, $\exists\, Y = \lim_{k\to\infty} Y_{n(k)}$. For any $\epsilon > 0$, we can then choose $n(k)$ sufficiently large so that $\forall\, n \geq n(k)$,

$$\begin{aligned}\mathbb{P}(|Y_n - Y| \geq 2\epsilon) &\leq \mathbb{P}\Big(|Y_n - Y_{n(k)}| \geq \epsilon\Big) + \mathbb{P}\Big(|Y_{n(k)} - Y| \geq \epsilon\Big) \\ &\leq 2\epsilon.\end{aligned}$$

The lemma is now proved. □

Lemma 11.4. *Suppose that $(\mathbb{P}_{X_n})$ and $(\mathbb{P}_{Y_n})$ are both uniformly tight. Then $(\mathbb{P}_{X_n+Y_n})$ is uniformly tight.*

Proof. Exercise. □

Lemma 11.5. *If $\mathbb{P}\mathbb{Q} = \mathbb{P}$, then $\mathbb{Q}(\{0\}) = 1$.*

Proof. Let $X \sim \mathbb{P}$ and $Y \sim \mathbb{Q}$ be independent with characteristic functions φ_X and φ_Y respectively. Since $\mathbb{P}\mathbb{Q} = \mathbb{P}$, we have $\varphi_X(t)\varphi_Y(t) = \varphi_X(t)$ for all $t \in \mathbb{R}$. Since $\varphi_X(t)$ is continuous and $\varphi_X(0) = 1$, $\exists\, \epsilon > 0$ such that $|\varphi_X(t)| > 0\ \forall\, |t| \leq \epsilon$. This implies that $\varphi_Y(t) = 1$ for such values of t. Therefore $\mathbb{E}[\cos tY] = 1$ but since $\cos(\cdot) \leq 1$, we must have $tY = 0 \mod 2\pi$ a.s.

Take $|s|, |t| \leq \epsilon$ with s/t being irrational. Then for each ω, we have

$$\begin{aligned}tY(\omega) &= 2\pi k,\ \ k \in \mathbb{Z}, \\ sY(\omega) &= 2\pi m,\ \ m \in \mathbb{Z}.\end{aligned}$$

If $Y(\omega) \neq 0$, $s/t = m/k$, a contradiction. Therefore $Y = 0$ a.s. □

Proof of Theorem 11.4. We first show that $S_n = \sum_{i=1}^n X_i$ converges in probability implies convergence a.s. Suppose $S_n \xrightarrow{\text{p}} S$, i.e., for all $\epsilon > 0$, $\exists\ n(\epsilon)$ such that $\forall\, n \geq n(\epsilon)$,

$$\mathbb{P}(|S_n - S| > \epsilon) \leq \epsilon. \tag{8}$$

For $k, j \geq n(\epsilon)$, we have

$$\mathbb{P}(|S_k - S_j| \geq 2\epsilon) \leq \mathbb{P}(|S_k - S| \geq \epsilon) + \mathbb{P}(|S_j - S| \geq \epsilon) \leq 2\epsilon$$

From Lemma 11.1, we obtain

$$\begin{aligned}\mathbb{P}\Big(\max_{n\leq j\leq k} |S_j - S_n| \geq 4\epsilon\Big) &\leq \frac{1}{1-2\epsilon}\mathbb{P}(|S_k - S_n| \geq 2\epsilon) \\ &\leq \frac{2\epsilon}{1-2\epsilon} \\ &\leq 3\epsilon,\end{aligned}$$

for ϵ sufficiently small. The MCT (Theorem 4.2) then yields

$$\mathbb{P}\Big(\max_{j\geq n} |S_j - S_n| \geq 4\epsilon\Big) \leq 3\epsilon.$$

Together with (8), we finally have

$$\mathbb{P}\Big(\max_{j\geq n}|S_j - S| \geq 5\epsilon\Big) \leq 4\epsilon.$$

Lemma 11.2 then implies that $S_j \to S$ a.s. as $j \to \infty$.

We next show that S_n converges in distribution implies convergence in probability. Suppose that $\mathbb{P}_{S_n} \xrightarrow{\text{d}} \mathbb{P}$ for some $\mathbb{P}$. From Lemma 9.7, $(\mathbb{P}_{S_n})_{n\geq 1}$ is uniformly tight. Therefore from Lemma 11.4, $(\mathbb{P}_{S_n - S_k})_{1\leq k\leq n}$ is uniformly tight. We proceed by contradiction. Suppose $\exists\ \epsilon > 0$ and $(n(l))$, $(m(l))$ with $n(l) \leq m(l)\ \forall\, l$ such that

$$\mathbb{P}\Big(|S_{m(l)} - S_{n(l)}| > \epsilon\Big) \geq \epsilon. \tag{9}$$

Let $Y_l = S_{m(l)} - S_{n(l)}$, then since $(\mathbb{P}_{Y_l})$ is uniformly tight, from Helly's Selection Theorem (Theorem 9.2), $\exists\ (l(r))$ such that $\mathbb{P}_{Y_{l(r)}} \xrightarrow{\text{d}}$ some distribution $\mathbb{Q}$. Since $S_{m(l(r))} = S_{n(l(r))} + Y_{l(r)}$ and $S_{n(l)}$, $Y_{l(r)}$ are independent, we have

$$\mathbb{P}_{S_{m(l(r))}} = \mathbb{P}_{S_{n(l(r))}} * \mathbb{P}_{Y_{l(r)}}.$$

Taking $r \to \infty$, we then have $\mathbb{P} = \mathbb{P}\mathbb{Q}$ from Exercise 10.1. From Lemma 11.5, $\mathbb{Q}(\{0\}) = 1$, which implies that $\mathbb{P}\Big(|Y_{l(r)}| > \epsilon\Big) < \epsilon$ for r sufficiently large. This contradicts (9) and the proof is complete. □

11.5 Poisson Convergence

Let $X_{n,m}$, $n, m \geq 1$ be independent Bernoulli r.v.s with $\mathbb{P}(X_{n,m} = 1) = p_{n,m}$. If $p_{n,m} \to 0$ sufficiently fast as $n \to \infty$, then

$$\sum_{m=1}^{n} \operatorname{var} X_{n,m} = \sum_{m=1}^{n} p_{n,m}(1 - p_{n,m}) \to 0.$$

This violates Lindeberg's CLT condition (i), therefore we cannot apply the CLT here. However, we can still obtain a convergence in distribution result.

Theorem 11.5. *Suppose*

(i) $\sum_{m=1}^{n} p_{n,m} \to \lambda \in (0, \infty)$ *as* $n \to \infty$, *and*

(ii) $\max_{1\leq m\leq n} p_{n,m} \to 0$ *as* $n \to \infty$,

then $S_n = \sum_{m=1}^{n} X_{n,m} \xrightarrow{\text{d}} \operatorname{Po}(\lambda)$, *where* $\operatorname{Po}(\lambda)(k) = \dfrac{\lambda^k}{k!}e^{-\lambda}$ *for* $k \in \mathbb{Z}_{\geq 0}$ *is the Poisson distribution.*

As in the proof of Lindeberg's CLT, the proof of this result proceeds via Levy's continuity theorem (Theorem 10.1). We need a preliminary lemma.

Lemma 11.6. *If the complex numbers z_i, w_i, $i = 1, \dots, n$, are such that $|z_i|, |w_i| \le \theta$, then*

$$\left|\prod_{i=1}^{n} z_i - \prod_{i=1}^{n} w_i\right| \le \theta^{n-1} \sum_{i=1}^{n} |z_i - w_i|.$$

Proof. We prove by induction on n. The result obviously holds for $n = 1$. Suppose it is true for $n - 1$. Then,

$$\begin{aligned}
&\left|\prod_{i=1}^{n} z_i - \prod_{i=1}^{n} w_i\right| \\
&\le \left|\prod_{i=1}^{n-1} z_i \cdot z_n - \prod_{i=1}^{n-1} w_i \cdot z_n\right| + \left|\prod_{i=1}^{n-1} w_i \cdot z_n - \prod_{i=1}^{n-1} w_i \cdot w_n\right| \\
&\le \theta \left|\prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} w_i\right| + \theta^{n-1}|z_n - w_n| \\
&\le \theta^{n-1} \sum_{i=1}^{n-1} |z_i - w_i| + \theta^{n-1}|z_n - w_n| \\
&= \theta^{n-1} \sum_{i=1}^{n} |z_i - w_i|.
\end{aligned}$$

□

Proof of Lemma 11.6. The characteristic function of $\mathrm{Po}\,(\lambda)$ is $\exp(\lambda(e^{\mathrm{i}t} - 1))$ while that for S_n is

$$\mathbb{E}e^{\mathrm{i}tS_n} = \prod_{m=1}^{n} \Big(1 + p_{n,m}\Big(e^{\mathrm{i}t} - 1\Big)\Big).$$

We have

$$\begin{aligned}
\Big|\exp\Big(p_{n,m}\Big(e^{\mathrm{i}t} - 1\Big)\Big)\Big| &= \exp(p_{n,m}\Re(e^{\mathrm{i}t} - 1)) \\
&= \exp(p_{n,m}(\cos t - 1)) \le 1,
\end{aligned}$$

and also

$$\Big|1 + p_{n,m}(e^{\mathrm{i}t} - 1)\Big| = |1 + p_{n,m}(\cos t - 1) + \mathrm{i}\sin t| \le 1,$$

which together with Lemma 11.6 yield

$$\begin{aligned}
&\left|\prod_{m=1}^{n}\Big(1+p_{n,m}\Big(e^{\mathrm{i}t}-1\Big)\Big)-\exp\left(\sum_{m=1}^{n}p_{n,m}(e^{\mathrm{i}t}-1)\right)\right| \\
&\leq \sum_{m=1}^{n}\Big|\exp\Big(p_{n,m}\Big(e^{\mathrm{i}t}-1\Big)\Big)-\Big(1+p_{n,m}\Big(e^{\mathrm{i}t}-1\Big)\Big)\Big| \\
&\leq \frac{1}{2}\sum_{m=1}^{n}p_{n,m}^{2}\Big|e^{\mathrm{i}t}-1\Big|^{2} \\
&\leq 2\max_{1\leq m\leq n}p_{n,m}\sum_{m=1}^{n}p_{n,m}\to 0 \text{ as } n\to\infty.
\end{aligned}$$

Since $\exp\Big(\sum_{m=1}^{n}p_{n,m}(e^{\mathrm{i}t}-1)\Big)\to\exp(\lambda(e^{\mathrm{i}t}-1))$ as $n\to\infty$, the result follows from Levy's continuity theorem (Theorem 10.1). □