MATH 589: Advanced Probability Theory 2
Final Exam: 14 December 2021 18:30-21:30

1 Central Limit Theorem, Characteristic Functions, and Convergence of Probability Measures

1.1 Review of Sums of Independent Random Variables

Consider $\{X_n \mid n \in \mathbb{N}\}$ iid random variables with $\mathbb{E}[X_1] = 0$ (WLOG) and $\mathbb{E}\left[X_1^2\right] = 1$. Set $S_n := \sum_{j=1}^n X_J$. From the SSLN,

$$\frac{S_n}{n} \to 0$$

almost surely. In other words, $|S_n|$ has sub-linear growth as $n \to \infty$. In fact, given any sequence $\{b_n \mid n \geq 1\} \subseteq]0, \infty[$ such that $b_n \uparrow \infty$, if

$$\sum_{n=1}^{\infty} \frac{1}{b_n^2} < \infty,$$

i.e., b_n grows sufficiently fast, then $\frac{S_n}{b_n} \to 0$ almost surely (by Kronecker's Lemma, c.f. MATH 587). Why?

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\left[X_n^2\right]}{b_n^2} < \infty \Rightarrow \sum_{n=1}^{\infty} \frac{X_n}{b_n} \text{ converges almost surely } \Rightarrow \frac{S_n}{b_n} \to 0 \text{ almost surely.}$$

Such a sequence $\{b_n\}$ includes:

- $\{n^p\}$ for $p > \frac{1}{2}$.
- $\{\sqrt{n}(\ln(n))^p\}$ for any $p > \frac{1}{2}$.

This means that I can do better than what I know about the LLN. For example, we know that $|S_n|$ grows slower than $\sqrt{n}(\ln(p))^{1/2}$ for any $p > \frac{1}{2}$. Since the inequality is strict, this means you can always do better. There is not a critical level. Now suppose we are interested in the asymptotic behaviour? Can we find a lower bound for the growth rate of S_n?

On the other hand, if $\{X_n \mid n \geq 1\}$ is iid $N(0,1)$ standard Gaussian random variables. Then, set:

$$\breve{S}_n := \frac{S_n}{\sqrt{n}}. \tag{1}$$

$\breve{S}_n$ is again $N(0,1)$ for all $n \geq 1$. At least, in this case, $\breve{S}_n$ doesn't converge to any constant almost surely. In fact, it's easy to see that $\limsup_n \frac{S_n}{\sqrt{n}} = +\infty$ and $\liminf_n \frac{S_n}{\sqrt{n}} = -\infty$ almost surely. Why is this? Let's consider the limsup. For all $R > 0$,

$$\begin{aligned}\mathbb{P}\left(\breve{S}_n > R\right) &= \frac{1}{\sqrt{2\pi}} \int_R^{+\infty} e^{-\frac{x^2}{2}}\, dx \\ &= p_R \\ &> 0.\end{aligned}$$

Since $\limsup_n \breve{S}_n \in m\mathcal{T}$ (tail σ-algebra, we have from the Kolmogorov 0-1 Law that $\limsup_n \breve{S}_n$ is constant almost surely. What is this constant? Write:

$$\breve{S}_n = \frac{S_n}{\sqrt{n}} = \frac{\sum_{j=1}^n X_j + \sum_{j=N+1}^n X_j}{\sqrt{n}}.$$

As $n \to \infty$, $\frac{\sum_{j=1}^n X_j}{\sqrt{n}}$ goes to infinity. Hence, $\limsup_n \breve{S}_n = \infty$ almost surely. One can do a similar analysis for the liminf.

Remark that $\breve{S}_n \sim N(0,1)$ is also seen for a more general sequence of random variables. This phenomenon is called the **Central Limit Phenomenon**.

Q: Can I have a better description of the asymptotics of S_n?

The answer is the **Law of the Iterated Logarithm**.

Theorem 1 (Law of Iterated Logarithm)**.** *Let $\{X_n\}$ be a sequence of iid RVs with $\mathbb{E}[X_1] = 0$ and $\mathbb{E}[X_1^2] = 1$. For every $n \geq 1$, set $S_n = \sum_{j=1}^n X_j$, and define Λ_n to be the iterated logarithm:*

$$\Lambda_n := \sqrt{2n \ln(\ln(n \vee 3))}.$$

It turns out that Λ_n will give us the accurate oscillation rate of S_n. Recall that the notation $n \vee 3 = \max\{n, 3\}$. Then, we can conclude:

- $\limsup_n \frac{S_n}{\Lambda_n} = 1$ *almost surely.*
- $\liminf_n \frac{S_n}{\Lambda_n} = -1$ *almost surely.*

In fact, for every $c \in [-1, 1]$, for almost every sample point $\omega \in \Omega$, there exists a subsequence $\{n_k\}_\omega \subseteq \mathbb{N}$ such that

$$\lim_{k \to \infty} \frac{S_{n_k}(\omega)}{\Lambda_{n_k}} = c. \tag{2}$$

The picture you want to have in mind is the following:

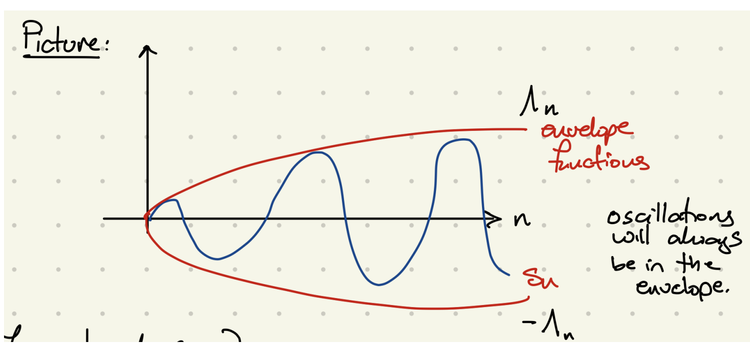

Figure 1: The oscillations of S_n will always be in the envelope given by $\pm\Lambda_n$.

In particular, note that LIL $\Rightarrow$ SLLN. The LIL is a refinement of the SLLN; Λ_n is sub-linear. Another perspective is by looking at it from the Kolmogorov 0-1 Law perspective: the liminf and limsup are constant almost surely.

Task # 1: Prove the Law of Iterated Logarithm.

Q: What can we say about the distribution?

The Central Limit Theorem will answer this question. For now, we will provide a heuristic overview; in the coming sections, we will rigorously do everything.

Idea: in the study of LLN, we consider $\bar{S}_n := \frac{S_n}{n}$, where $\mathbb{E}\left[\bar{S}_n\right] = \mathbb{E}\left[S_1\right] = 0$ for all $n \in \mathbb{N}$. Here, this means that $\bar{S}_n$ preserves the first moment. In **(CLT)** we will consider $\breve{S}_n := \frac{S_n}{\sqrt{n}}$, where $\mathbb{E}\left[\breve{S}_n\right] = 0$ (so, $\breve{S}_n = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}}$, where $\mathbb{E}\left[\breve{S}_n\right] = 0$. Moreover,

$$\mathbb{E}\left[(\breve{S}_n)^2\right] = \frac{n\mathbb{E}\left[X_1^2\right]}{n} = 1.$$

Note that in the CLT, the first and second moments are preserved.

1. The expected value tells us where the mass is centred.
2. The variance measures how the mass is spread out: how random the random variable is.

Heuristically, the CLT studies how the randomness will replace itself under the assumption / condition that the amount of randomness is preserved or fixed. For sure, it will not be going to a constant, and it will resemble a $N(0,1)$ as $n \to \infty$.

We work in the following set-up: $\{X_n\}$ iid random variables with $\mathbb{E}\left[X_1\right] = 0$, $\mathbb{E}\left[X_1^2\right] = 1$, and $S_n = \sum_{j=1}^n X_j$.

Remark: by preserving / stabilizing the second moments, $\breve{S}_n$ stabilizes all the moments. We can see this with the following computation / proof.

Suppose $X_1 \in L^p$ for all $p \geq 1$. We will show this stablization by induction. For some $m \in \mathbb{N}$, define:

$$L_j := \lim_{n\to\infty} \mathbb{E}\left[(\breve{S}_n)^j\right] \text{ exists for } 1 \leq j \leq m. \tag{3}$$

Consider the $(m+1)$st moment of $\breve{S}_n$:

$$\begin{aligned}
\mathbb{E}\left[S_n^{m+1}\right] &= \mathbb{E}\left[S_n S_n^m\right] \\
&= \sum_{j=1}^n \mathbb{E}\left[X_j (X_j + S_{n\backslash j})^m\right] \\
&= \sum_{j=1}^n \sum_{k=0}^m \binom{m}{k} \mathbb{E}\left[X_j^{k+1}\right] \mathbb{E}\left[S_{n\backslash j}^{m-k}\right] \text{ (by the binomial formula)} \\
&= n\left(\mathbb{E}\left[X_1\right]\mathbb{E}\left[S_{n\backslash 1}^m\right] + m\underbrace{\mathbb{E}\left[X_1^2\right]}_{=1}\mathbb{E}\left[S_{n\backslash 1}^{m-1}\right] + \sum_{k=2}^m \binom{m}{k}\mathbb{E}\left[X_1^{k+1}\right]\mathbb{E}\left[S_{n\backslash 1}^{m-k}\right]\right),
\end{aligned}$$

where $\mathbb{E}\left[X_1\right] = 0$ means the first term vanishes. Since $\mathbb{E}\left[X_1^2\right] = 1$, we get, by applying the definition of $\breve{S}_n$:

$$\begin{aligned}
\mathbb{E}\left[(\breve{S}_n)^{m+1}\right] &= n^{-\frac{m+1}{2}}\mathbb{E}\left[S_n^{m+1}\right] \\
&= n^{-\frac{m+1}{2}}\left(m\mathbb{E}\left[S_{n\backslash 1}^{m-1}\right] + \sum_{k=2}^m \binom{m}{k}\mathbb{E}\left[X_1^{k+1}\right]\mathbb{E}\left[S_{n\backslash 1}^{m-k}\right]\right).
\end{aligned}$$

Substituting in the definition of $\breve{S}_n$, we obtain:

$$= \left(\frac{n-1}{n}\right)^{\frac{m-1}{2}} m\underbrace{\mathbb{E}\left[(\breve{S}_{n\backslash 1})^{m-1}\right]}_{:=L_{m-1}} + \sum_{k=2}^m \underbrace{\frac{(n-1)^{\frac{m-k}{2}}}{n^{\frac{m-1}{2}}}}_{\to 0 \text{ as } n\to\infty}\binom{m}{k}\mathbb{E}\left[X_1^{k+1}\right]\underbrace{\mathbb{E}\left[(\breve{S}_{n-1})^{m-k}\right]}_{:=L_{m-k}}.$$

So as $n \to \infty$, we obtain:

$$1 \cdot m \cdot L_{m-1}. \tag{4}$$

This gives us the following recursive relationship: $L_{m+1} = mL_{m-1}$. Since $L_1 = 0$ and $L_2 = 1$, the *second moment stabilizes all the moments*:

$$L_{2m+1} = 0 \text{ (all odd indices)} \tag{5}$$
$$L_{2m} = 1 \cdot 3 \cdot 4 \cdot ... \cdot (2m-1) \text{ (product of all the odd numbers) } = (2m+1)!! \tag{6}$$

These are the moments of the standard Gaussian. So, the moments of $\breve{S}_n$ converge to the corresponding moments of a $N(0,1)$ random variable as $n \to \infty$. Therefore, intuitively, the distribution of $\breve{S}_n$ "approximates" $N(0,1)$ as $n \to \infty$. As a corollary, if φ is a polynomimal of any degree, then

$$\lim_{n\to\infty} \mathbb{E}\left[\varphi\left(\breve{S}_n\right)\right] = \frac{1}{\sqrt{2\pi}} \int \varphi(x) e^{-\frac{x^2}{2}} dx = \gamma 0,1(\varphi)$$

where $\gamma_{0,1} = N(0,1)$.

1.2 Central Limit Theorems

Theorem 2 (Lindeberg's Central Limit Theorem (CLT))**.** *Assume that $\{X_n\}$ is a sequence of independent square-integrable random variables on a probability space, $\mathbb{E}[X_n] = 0$. For every $n \in \mathbb{N}$, set:*

$$\sigma_n := \sqrt{Var(X_n)}$$
$$\Sigma_n := \sqrt{Var(S_n)} = \sqrt{\sum_{j=1}^{n} \sigma_j^2},$$

where the final equality is true only if the X_n are independent. Set

$$\breve{S}_n = \frac{S_n}{\Sigma_n}$$

(so $\mathbb{E}\left[\breve{S}_n\right] = 0$ and $\mathbb{E}\left[\breve{S}_n\right] = 1$). For all $\varepsilon > 0$, set:

$$g_n(\varepsilon) := \frac{1}{\Sigma_n^2} \sum_{j=1}^{n} \mathbb{E}\left[X_j^2; |X_j| > \varepsilon \Sigma_n\right] \quad \textit{or}$$
$$g_n(\varepsilon) := \sum_{j=1}^{n} \mathbb{E}\left[\left(\frac{X_j}{\Sigma_n}\right)^2; \left|\frac{X_j}{\Sigma_n}\right| > \varepsilon\right].$$

Under this setting, for every $\varphi \in C^3(\mathbb{R})$ with φ'' and φ''' being bounded on $\mathbb{R}$ and for every $\varepsilon > 0$,

$$\left|\mathbb{E}\left[\varphi(\breve{S}_n)\right] - \gamma_{0,1}(\varphi)\right| \le \frac{1}{2}(\varepsilon + \sqrt{g_n(\varepsilon)})||\varphi'''||_n + g_n(\varepsilon)||\varphi''||_n. \tag{7}$$

In particular, if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} g_n(\varepsilon) = 0, \tag{8}$$

(this is called ***Lindeberg's Condition****), then*

$$\lim_{n\to\infty} \mathbb{E}\left[\varphi(\breve{S}_n)\right] = \gamma_{0,1}(\varphi).$$

Before the proof, we first make a quick remark. In the case when $\{X_n \mid n \geq 1\}$ is iid with $\mathbb{E}[X_1] = 0$, $\mathbb{E}\left[X_1^2\right] = 1$ for all $n \geq 1$, $\sigma_n = 1$, $\Sigma_n = \sqrt{n}$. Hence,

$$\breve{S}_n = \frac{S_n}{\sqrt{n}},$$

and so, for all $\varepsilon > 0$,

$$\begin{aligned} g_n(\varepsilon) &= \frac{1}{\Sigma_n^2} \sum_{j=1}^{n} \mathbb{E}\left[X_j^2; |X_j| > \varepsilon \Sigma_n\right] \\ &= \frac{1}{n} \sum_{j=1}^{n} \mathbb{E}\left[X_j^2; |X_j| > \varepsilon \sqrt{n}\right] \\ &= \mathbb{E}\left[X_1^2; |X_1| > \varepsilon \sqrt{n}\right] \to 0 \text{ as } n \to \infty. \end{aligned}$$

So, in this case, Lindeberg's Condition is always satisfied.

Proof. Before the proof, the insight is as follows: as $n \to \infty$, the contribution of the X_j's are getting closer and closer to a centered Gaussian $N(0, \sigma_j^2)$ random variable.

Introduce $\{Z_n \mid n \geq 1\}$ iid random variables independent of $\{X_n \mid n \geq 1\}$. For all $n \geq 1$, set $Y_n := \sigma_n Z_n$. Then, as we know Y_n is a $N(0, \sigma_n^2)$ random variable. Further define $\breve{T}_n := \frac{1}{\Sigma_n} \sum_{j=1}^{n} Y_j$. Note that $\breve{T}_n$ is a $N(0,1)$ random variable. Hence,

$$\gamma_{0,1}(\varphi) = \mathbb{E}\left[\varphi(\breve{T}_n)\right] \Rightarrow \mathbb{E}\left[\varphi(\breve{S}_n)\right] - \gamma_{0,1}(\varphi) = \mathbb{E}\left[\varphi(\breve{S}_n) - \varphi(\breve{T}_n)\right].$$

Hence,

$$\begin{aligned} \varphi(\breve{S}_n) - \varphi(\breve{T}_n) =& \varphi\left(\frac{1}{\Sigma_n}(X_1 + ... + X_n)\right) - \varphi\left(\frac{1}{\Sigma_n}(X_1 + ... + X_{n-1} + Y_n)\right) + \varphi\left(\frac{1}{\Sigma_n}(X_1 + ... + X_{n-1} + Y_n)\right) \\ &- \varphi\left(\frac{1}{\Sigma_n}(X_1 + ... + Y_{n-1} + Y_n)\right) + \varphi\left(\frac{1}{\Sigma_n}(X_1 + ... + Y_{n-1} + Y_n)\right) - ... \\ &- \varphi\left(\frac{1}{\Sigma_n}(X_1 + Y_2 + ... + Y_n)\right) + \varphi\left(\frac{1}{\Sigma_n}(X_1 + Y_2 + ... + Y_n)\right) + \varphi\left(\frac{1}{\Sigma_n}(Y_1 + ... + Y_n)\right). \end{aligned}$$

In light of this representation, for all $1 \leq j \leq n$, set:

$$U_j := \frac{1}{\Sigma_n}(X_1 + ... + X_{j-1} + X_{j+1} + Y_{j+2} + ... + Y_n). \tag{9}$$

Then, we can express the above more compactly as:

$$\varphi(\breve{S}_n) - \varphi(\breve{T}_n) = \sum_{j=1}^{n} \left(\varphi\left(U_j + \frac{X_j}{\Sigma_n}\right) - \varphi\left(U_j + \frac{Y_j}{\Sigma_n}\right)\right)$$

The idea is to now use Taylor expansions: recall that the Taylor Expansion of φ is:

$$\varphi(U_j + \xi) = \varphi(U_j) + \xi \varphi'(U_j) + \frac{\xi^2}{2} \varphi''(U_j). + ...$$

Set $R_j(\xi) = \varphi(U_j + \xi) - \varphi(U_j) - \xi \varphi'(U_j) - \frac{1}{2}\xi^2 \varphi'(U_j)$. Then,

$$\mathbb{E}\left[\varphi\left(U_j + \frac{X_j}{\Sigma_n}\right)\right] = \mathbb{E}\left[R_j\left(\frac{X_j}{\Sigma_n}\right)\right] + \mathbb{E}\left[\varphi(U_j)\right] + \mathbb{E}\left[\frac{X_j}{\Sigma_n}\varphi'(U_j)\right] + \frac{1}{2}\mathbb{E}\left[\frac{X_j^2}{\Sigma_n}\varphi''(U_j)\right].$$

Let's simplify all these terms:

- Since X_j is independent of U_j, we can write:

$$\mathbb{E}\left[\frac{X_j}{\Sigma_n}\varphi'(U_j)\right] = \frac{1}{\Sigma_n}\mathbb{E}\left[X_j\right]\mathbb{E}\left[\varphi'(U_j)\right] = 0.$$
$$\frac{1}{2}\mathbb{E}\left[\frac{X_j^2}{\Sigma_n}\varphi''(U_j)\right] = \frac{1}{2}\mathbb{E}\left[\frac{X_j^2}{\Sigma_n^2}\right]\cdot\mathbb{E}\left[\varphi''(U_j)\right] = \frac{\sigma_j^2}{\Sigma_n^2}\mathbb{E}\left[\varphi''(U_j)\right]$$

Similarly,

$$\mathbb{E}\left[\varphi\left(U_j + \frac{Y_j}{\Sigma_n}\right)\right] = \mathbb{E}\left[R_j\left(\frac{Y_j}{\Sigma_n}\right)\right] + \mathbb{E}\left[\varphi(U_j)\right] + 0 + \frac{1}{2}\frac{\sigma_j^2}{\Sigma_n^2}\cdot\mathbb{E}\left[\varphi''(U_j)\right].$$

Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\varphi(\breve{S}_n) - \varphi(\breve{T}_n)\right]\right| &\leq \sum_{j=1}^n \left|\mathbb{E}\left[R_j\left(\frac{X_j}{\Sigma_n}\right)\right] - \mathbb{E}\left[R_j\left(\frac{Y_j}{\Sigma_n}\right)\right]\right| \\ &\leq \sum_{j=1}^n \left|\mathbb{E}\left[R_j\left(\frac{X_j}{\Sigma_n}\right)\right]\right| + \left|\mathbb{E}\left[R_j\left(\frac{Y_j}{\Sigma_n}\right)\right]\right|\end{aligned}$$

Moreover, $|R_j(\xi)| \leq \left(\frac{1}{6}\xi^3||\varphi'''||_n\right) \wedge \left(\xi^2||\varphi''||_n\right)$, where the first case happens if ξ is small and the second case happens if ξ is not small. Hence, for all $\varepsilon > 0$, we have:

$$\sum_{j=1}^n \left|\mathbb{E}\left[R_j\left(\frac{X_j}{\Sigma_n}\right)\right]\right| \leq \frac{1}{6}||\varphi''||_n \sum_{j=1}^n \mathbb{E}\left[\frac{|X_j|^3}{\Sigma_n^3}; |X_j| \leq \varepsilon\Sigma_n\right] + ||\varphi''||_n \sum_{j=1}^n \mathbb{E}\left[\frac{|X_j|^2}{\Sigma_n^2}; \frac{|X_j|}{\Sigma_n} > \varepsilon\right],$$

where the first term in the sum comes from the bound for ξ being small and the second term in the sum comes fro the bound for ξ being not so small. Pulling one of the $|X_j|$ out of the fraction in the first term of the sum, and using the bound given, we obtain:

$$\leq \frac{\varepsilon}{6}||\varphi''||_n \sum_{j=1}^n \frac{\mathbb{E}\left[X_j^2\right]}{\Sigma_n^2} + ||\varphi''||_n \cdot g_n(\varepsilon),$$

which is good, since we have $\sum_{j=1}^n \frac{\sigma_j^2}{\Sigma_n^2} = 1$. Hence,

$$\sum_{j=1}^n \left|\mathbb{E}\left[R_j\left(\frac{X_j}{\Sigma_n}\right)\right]\right| \leq \frac{\varepsilon}{6}||\varphi''|||_n + ||\varphi''||_n \cdot g_n(\varepsilon).$$

Similarly,

$$\begin{aligned}\sum_{j=1}^n \mathbb{E}\left[\left|R_j\left(\frac{Y_j}{\Sigma_n}\right)\right|\right] &\leq \frac{1}{6}||\varphi'''||_n \mathbb{E}\left[|Z_n|^3\right] \sum_{j=1}^n \frac{\sigma_j^3}{\Sigma_n^3} \\ &\leq \frac{1}{3}||\varphi'''||_n \max_{1\leq j\leq n}\frac{\sigma_j}{\Sigma_n}\cdot \underbrace{\sum_{j=1}^n \frac{\sigma_j^2}{\Sigma_n^2}}_{=1}.\end{aligned}$$

We have that for all $1 \leq j \leq n$,

$$\begin{aligned}\sigma_j^2 = \mathbb{E}\left[X_j^2\right] &= \mathbb{E}\left[X_j^2; |X_j| \leq \varepsilon\Sigma_n\right] + \mathbb{E}\left[X_j^2; |X_j| > \varepsilon\Sigma_n\right] \\ &= \varepsilon^2\Sigma_n^2 + \sum_{l=1}^n \mathbb{E}\left[X_l^2; |X_l| > \varepsilon\Sigma_n\right].\end{aligned}$$

Hence,

$$\max_{1\le j\le n} \frac{\sigma_j^2}{\Sigma_n^2} \le \varepsilon^2 + g_n(\varepsilon) \Rightarrow \max_{1\le j\le n} \frac{\sigma_j}{\Sigma_n} \le \sqrt{\varepsilon^2 + g_n(\varepsilon)} \le \varepsilon + \sqrt{g_n(\varepsilon)}.$$

Collecting all the bounds,

$$\begin{aligned}\left|\mathbb{E}\left[\varphi(\breve{S}_n)\right] - \mathbb{E}\left[\varphi(\breve{T}_n)\right]\right| &\le \frac{\varepsilon}{6}||\varphi'''||_n + g_n(\varepsilon)||\varphi''||_n + \frac{1}{3}||\varphi'''||_n(\varepsilon + \sqrt{g_n(\varepsilon)} \\ &\le \frac{1}{2}\left(\varepsilon + \sqrt{g_n(\varepsilon)}\right)||\varphi'''||_n + g_n(\varepsilon)||\varphi''||_n\end{aligned}$$

which proves the theorem. □