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 $\mathbb{E}\left[X_1^2\right] = 1$. 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:

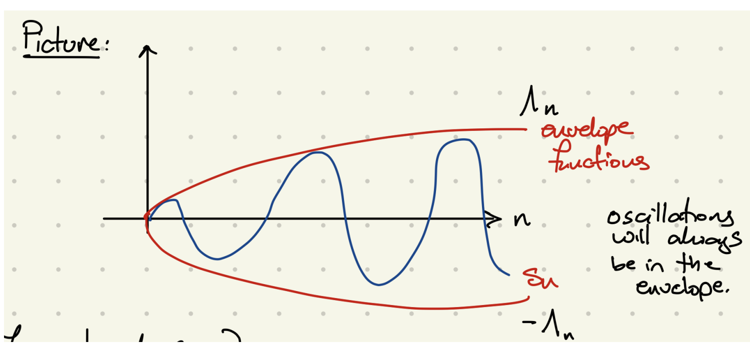


Figure 1: The oscillations of S_n will always be in the envelope given by $\pm\Lambda_n$.

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 $\breve{S}_n$:

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

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 $\breve{S}_n$:

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

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

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

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 $\varepsilon > 0$,

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

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,

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

Moreover, $|R_j(\xi)| \le \left(\frac{1}{6}\xi^3||\varphi'''||_n\right) \wedge \left(\xi^2||\varphi''||_n\right)$, 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:

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

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:

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

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

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

Similarly,

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

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

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

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. □

Corrolary 1. Under the same setting as before, if Lindeberg's condition holds, then for all $\varphi\in C_c^\infty(\mathbb{R})$,

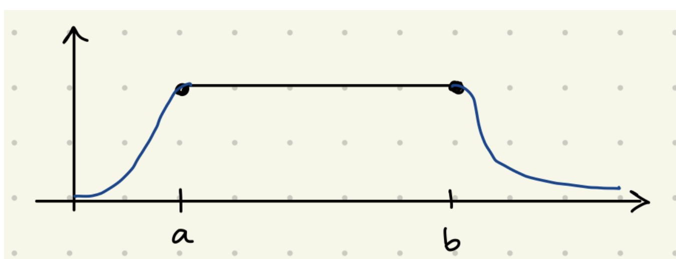
$$\lim_{n\to\infty}\mathbb{E}\left[\varphi(\breve{S}_n)\right]=\gamma_{0,1}(\varphi). \tag{10}$$

In particular, we can show that for all $a,b\in\mathbb{R}$, $a<b$:

$$\mathbb{P}\left(a\le\breve{S}_n\le b\right)=\gamma_{0,1}(]a,b])=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

Proof. The proof only requires a standard fact from analysis, which we will use quite a lot in this course.

Fact. For $[a,b]$ closed, there exists a sequence of functions $\{\varphi_k\mid k\ge1\}\subseteq C_c^\infty(\mathbb{R})$ such that $0\le\varphi_k\le1$ for all $k\ge1$ and $\varphi_k\downarrow\chi_{[a,b]}$. The picture that you want to have in mind is:



Therefore, for all $k\ge1$, we have

$$\limsup_n\mathbb{P}\left(\breve{S}_n\in[a,b]\right)\le\lim_n\mathbb{E}\left[\varphi_k(\breve{S}_n)\right]=\gamma_{0,1}(\varphi_k),$$

where the final equality follows from the Lindeberg's CLT. As $k\to\infty$,

$$\gamma_{0,1}(\varphi_k)\to\gamma_{0,1}([a,b]).$$

Hence, $\limsup_n\mathbb{P}\left(a\le\breve{S}_n\le b\right)\le\gamma_{0,1}([a,b])$. Similarly, for $]a,b[$, there exists a sequence of functions $\{\psi_k\mid k\ge1\}$ such that $0\le\psi_k\le1$ for all $k\ge1$, $\psi_k\uparrow\chi_{]a,b[}$ (so, we approach the indicator function from below). Then,

$$\liminf_n\mathbb{P}\left(a<\breve{S}_n<b\right)\ge\lim_n\mathbb{E}\left[\psi_k(\breve{S}_n)\right]=\gamma_{0,1}(\varphi_k)\to\gamma_{0,1}(]a,b[).$$

Since $\gamma_{0,1}(]a,b[)=\gamma_{0,1}([a,b])$ we have the desired limit statement. □

So, now we want to look at smooth functions that approximates the indicator function χ of a set we are interested in studying. Let's first do some preparation.

Definition 1 (Convolution)**.** Given μ and ν, two probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ given by: for all $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu * \nu := \int_{\mathbb{R}^d} \nu(B - x)\mu(dx), \tag{11}$$

where recall the set $B - x := \{y \in \mathbb{R}^d \mid y + x \in B\}$.

Remarks. It's easy to check with Fubini's Theorem that:

1. $x \mapsto \nu(B - x)$ is a measure with respect to $\mathcal{B}(\mathbb{R}^d)$.
2. $\mu * \nu$ is again a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
3. $\mu * \nu = \nu * \mu$. If ρ is another probability measure on $\mathbb{R}^d$, then,

$$(\mu * \nu) * \rho = \mu * (\nu * \rho). \tag{12}$$

In the next proposition, we will see how convolution corresponds to taking the sum of two independent random variables.

Proposition 1. Given X and Y two independent random variables, $\mathbb{R}^d$-valued, with $\mathcal{L}_X = \mu$ and $\mathcal{L}_Y = \nu$. If X and Y are independent, then

$$\mathcal{L}_{X+Y} = \mu * \nu. \tag{13}$$

Proof. To see this, we first have that since X and Y are independent,

$$\mathcal{L}_{(X,Y)} = \mathcal{L}_X \cdot \mathcal{L}_Y = \mu \times \nu.$$

So, using Fubini's theorem, we obtain that for all $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
\mathbb{P}\left(X + Y \in B\right) &= \iint_{\mathbb{R}^d \times \mathbb{R}^d} \chi_B(x + y)(\mu \times \nu) \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \chi_B(x + y)\nu(dy) \right) \mu(dx) \\
&= \mu * \nu(B).
\end{aligned}$$

□

Remark. It's also possible to define the convolution of functions. Given f and g two functions on $\mathbb{R}^d$, for all $x \in \mathbb{R}^d$:

$$f * g(x) := \int_{\mathbb{R}^d} f(x - y)g(y)dy,$$

provided that the integral is defined. Similarly, $f * g = g * f$ and $(f * g) * h = f * (g * h)$.

Corrolary 2. If X and Y are independent, and X has a density f and Y has a density g, then $X + Y$ has a density $f * g$.

Notation. for every $x, \xi \in \mathbb{R}^d$, we will denote by $(\cdot, \cdot)$ the dot product:

$$(x, \xi) := \sum_{j=1}^{d} x_j \xi_j.$$

We will denote by $i := \sqrt{-1}$ the imaginary unit. For $z \in \mathbb{C}$, let $\overline{z}$ be the complex conjugate of z. We consider functions $\varphi : \mathbb{R}^d \to \mathbb{C}$. As we'd expect, φ is Borel $\iff$ the real and imaginary parts of φ are Borel functions in the standard sense. If μ is a probability measure on $\mathbb{R}^d$, then we write that $\varphi \in L^p(\mu)$ if

$$\int_{\mathbb{R}^d} |\varphi(x)|^p \mu(dx) < \infty.$$

(note that $|\varphi(x)|^2 = \mathrm{Re}^2(\varphi) + \mathrm{Im}(\varphi)$). Given two functions ψ and φ, $\mathbb{C}$-valued on $\mathbb{R}^d$, their inner product is given by:

$$\langle \varphi, \psi \rangle = (\varphi, \psi)_{L^2} = \int_{\mathbb{R}^d} \varphi(x) \overline{\psi(x)} dx.$$

Definition 2 (Characteristic Function)**.** Given a probability measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, the **characteristic function** of μ, denoted by $\hat{\mu}$, is a function on $\mathbb{R}^d$ such that for all $\xi \in \mathbb{R}^d$,

$$\hat{\mu}(\xi) := \int_{\mathbb{R}^d} e^{i(x,\xi)} \mu(dx). \tag{14}$$

$\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ is well-defined for every ξ, measurable (Fubini) with respect to $\mathcal{B}(\mathbb{R}^d)$, and $|\hat{\mu}| \leq 1$ for all $\xi \in \mathbb{R}^d$.

We can similarly define the characteristic function of a random variable. If X is a random variable on some probability space such that $\mathcal{L}_X = \mu$, then

$$\hat{\mu}(\xi) = \mathbb{E}\left[e^{i(X,\xi)}\right].$$

We now introduce some remarks on characteristic functions.

1. $\hat{\mu} : \mathbb{R}^d \to \mathbb{C}$ is a continuous function. (Can easily verify this by taking a sequence ξ_n, and use **(DOM)** since everything is bounded by 1).
2. If μ is symmetric, i.e., $\forall\ A \in \mathcal{B}(\mathbb{R}^d)$, $\mu(A) = \mu(-A)$. Then, $\hat{\mu}(\xi) \in \mathbb{R}$ for all $\xi \in \mathbb{R}^d$, since by symmetry, the imaginary part will cancel.
3. If μ and ν are two probability measures, then,

$$\widehat{\mu * \nu}(\xi) = \hat{\mu}(\xi) \cdot \hat{\nu}(\xi) = \hat{\mu}(\xi) \cdot \hat{\nu}(\xi),$$

for all $\xi \in \mathbb{R}^d$. To see this, implement with random variables. Take X and Y independent such that $\mathcal{L}_X = \mu$ and $\mathcal{L}_Y = \nu$. Then,

$$\begin{aligned}
\widehat{\mu * \nu}(\xi) &= \mathbb{E}\left[e^{i(X+Y,\xi)}\right] \\
&= \mathbb{E}\left[e^{i(X,\xi)} e^{i(Y,\xi)}\right] \\
&= \mathbb{E}\left[e^{i(X,\xi)}\right] \mathbb{E}\left[e^{i(Y,\xi)}\right] \\
&= \hat{\mu}(\xi) \cdot \hat{\nu}(\xi).
\end{aligned}$$

4. $\hat{\mu}$ contains information about "moments". To see this, assume that X is a random variable such that $\mathcal{L}_X = \mu$ and $\mathbb{E}[|X|^p] < \infty$ for some $p \geq 1$. Then, for every multi-index $\alpha = (\alpha_1, \alpha_2, ..., \alpha_d) \in \mathbb{N}^d$ such that $|\alpha| = \alpha_1 + \alpha_2 + ... + \alpha_d \leq p$. Then,

$$\begin{aligned} \partial^\alpha \hat{\mu}(\xi) &:= \partial_1^{\alpha_1} \partial_2^{\alpha_2} ... \partial_d^{\alpha_d} \hat{\mu}(\xi) \\ &= \int_{\mathbb{R}^d} (ix)^\alpha e^{i(x,\xi)} \mu(dx). \end{aligned}$$

This follows from **(DOM)** . The notation $(ix)^\alpha$ means $i^{|\alpha|} x_1^{\alpha_1} \cdot x_2^{\alpha_2} ... x_d^{\alpha_d}$. In particular, we have

$$[\partial^\alpha \hat{\mu}(\xi)]_{\xi=0} = i^{|\alpha|} \mathbb{E}[X^\alpha].$$

The $\mathbb{E}[X^\alpha]$ term is called the **cross-moment**. The notation X^α means $X_1^{\alpha_1} X_2^{\alpha_2} \cdots X_d^{\alpha_d}$.

(a) In general, $\hat{\mu} \in C^p(\mathbb{R}^d)$ does NOT imply that $\mathbb{E}[|X|^p] < \infty$. For example, consider that μ is the probability measure on $\mathbb{R}$ with density:

$$f(x) = \begin{cases} 0 & \text{if } |x| < 2 \\ \frac{c}{x^2 \ln(|x|)} & \text{if } |x| \geq 2, \end{cases}$$

where $c > 0$ is a constant such that $\int_{\mathbb{R}} f(x)dx = 1$. Now let X be a random variable such that $\mathcal{L}_X = \mu$. On one hand,

$$\mathbb{E}[|X|] = 2 \int_{2\infty} \frac{dx}{x \ln(x)} \cdot c = \infty \Rightarrow X \notin L^1.$$

On the other hand,

$$\hat{\mu}(\xi) = 2c \int_2^\infty \frac{\cos(x\xi)}{x^2 \ln(x)} dx.$$

One can verify that $\hat{\mu}$ is differentiable at every $\xi \in \mathbb{R}$ and $\hat{\mu}'(0) = 0$.

Example 1. On $\mathbb{R}$,

$$\hat{\partial}_{m,\sigma^2}(\xi) = e^{im\xi} e^{-\frac{1}{2}\sigma^2 \xi^2}.$$

On $\mathbb{R}^d$,

$$\hat{\partial}_{\vec{m},c}(\xi) = e^{i(\vec{m},\xi)} \cdot e^{-\frac{(\xi,c\xi)}{2}},$$

for all $\xi \in \mathbb{R}^d$. Observe that the characteristic functions have super-exponential decay like the densities.

Definition 3. Given a function φ on $\mathbb{R}^d$, the **Fourier Transform** of φ, denoted by $\hat{\varphi}$, is given by: for all $\varphi \in \mathbb{R}^d$:

$$\hat{\varphi}(\xi) := \int_{\mathbb{R}^d} e^{i(x,\xi)} \varphi(x) dx, \tag{15}$$

provided that the integral is defined. In particular, this means that if μ is a probability measure with density φ, then for all $\xi \in \mathbb{R}^d$:

$$\hat{\mu}(\xi) = \hat{\varphi}(\xi). \tag{16}$$

The next theorem tells us that the characteristic function uniquely defines a probability measure.

Theorem 3. *Let μ and ν be two probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. If $\hat{\mu}(\xi) = \hat{\nu}(\xi)$ for all $\xi \in \mathbb{R}^d$, then $\mu = \nu$.*

The proof will follow from these three lemmas.

Lemma 1. Let μ and ν be two probability measures. If $\mu(\varphi) = \nu(\varphi)$ for all $\varphi \in C_b(\mathbb{R}^d; \mathbb{C})$. Then, $\mu = \nu$.

Proof. Since the open sets form a generating π-system, it's sufficient to show that for every open set $B \subseteq \mathbb{R}^d$, $\mu(B) = \nu(B)$. Take B open, and consider $d(x, B^c) := \inf_{y \in B^c} |x - y|$. Then, we know from analysis that $x \mapsto d(x, B^c)$ is continuous. For all $k \geq 1$, set

$$\varphi_k(x) := \left(\frac{d(x, B^c)}{1 + d(x, B^c)} \right)^{1/k}.$$

Ten, $\varphi_k \in [0, 1]$ and for each k, φ_k is continuous. We have that $\varphi_k \uparrow \chi_B$. Why?

$$\lim_{k \to \infty} \varphi_k(x) = \begin{cases} 1 & \text{if } d(x, B^c) > 0 \iff x \in B^c \text{ since closed} \\ 0 & \text{if } d(x, B^c) = 0 \iff x \in B^c. \end{cases}$$

By **(DOM)** or **(MON)** ,

$$\mu(B) = \lim_k \mu(\varphi_k) = \lim_k \nu(\varphi_k) = \nu(B).$$

□

Lemma 2. For all $\varphi \in C_b(\mathbb{R}^d)$ there exists a sequence $\{\varphi_m \mid m \geq 1\} \subseteq C_c^\infty(\mathbb{R}^d)$ such that $||\varphi_m||_n \leq ||\varphi_n||$ for all $m \geq 1$ and $\lim_{m \to \infty} \varphi_m = \varphi$.

(As a result of Lemma 2, $\mu = \nu$ if $\mu(\varphi) = \nu(\varphi)$ for all $C_c^\infty(\mathbb{R}^d)$.

Lemma 3 (A Generalization of Plancherel's Theorem)**.** If $\psi \in C_c^\infty(\mathbb{R}^d)$ and μ is a probability measure on $\mathbb{R}^d$, then

$$\mu(\psi) = \int_{\mathbb{R}^d} \psi(x) \mu(dx) = (2\pi)^{-d} \int_{\mathbb{R}^d} \hat{\psi}(\xi) \overline{\hat{\mu}(\xi)} d\xi, \tag{17}$$

i.e., $\mu(\psi) = \langle \hat{\psi}, \hat{\mu} \rangle$.

As a result of Lemma 3, $\mu = \nu$ if $\hat{\mu} = \hat{\nu}$. We will neatly collect this into a theorem.

Theorem 4. *Let μ and ν be two probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then,*

$$\begin{aligned} \mu = \nu &\iff \mu(\varphi) = \nu(\varphi) \ \forall \ \varphi \in C_b(\mathbb{R}^d) \\ &\iff \mu(\psi) = \nu(\psi) \ \forall \ \psi \in C_c^\infty(\mathbb{R}^d) \\ &\iff \hat{\mu}(\xi) = \hat{\nu}(\xi) \ \forall \xi \in \mathbb{R}^d. \end{aligned}$$

We can think of $C_b(\mathbb{R}^d)$ and C_c^∞ as classes of test functions that test how measure behaves.

1.3 Weak Convergence of Probability Measures

There are only two types of convergence which will be covered in this course.

Definition 4 (Weak Convergence of Measure)**.** Assume that $\{\mu_n \mid n \geq 1\}$ and μ are probability measures on $\mathcal{B}(\mathbb{R}^d)$. We say that μ_n **converges weakly** to μ, and we write "$\mu_n \Rightarrow \mu$" if for all $\varphi \in C_b(\mathbb{R}^d; \mathbb{C})$:

$$\lim_{n\to\infty} \mu_n(\varphi) = \mu(\varphi). \tag{18}$$

We also have cnvergence in distribution of random variables.

Definition 5. Assume that $\{X_n\}$ and X are $\mathbb{R}^d$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. X_n converges to X **in distribution**, denoted by "$X_n \to X$ in distribution", if $\mathcal{L}X_n \to \mathcal{L}_X$, i.e., for all $\varphi \in C_b(\mathbb{R}^d)$,

$$\lim_{n\to\infty} \mathbb{E}\left[\varphi(X_n)\right] = \mathbb{E}\left[\varphi(X)\right]. \tag{19}$$

Remark.

1. For two probability measures μ and ν on $\mathbb{R}^d$, the most natural way of putting a metric on the space of probability measures is the **total variation distance** between μ and ν:

$$||\mu - \nu||_{\text{var}} := \sup\{|\mu(A) - \nu(A)| \mid A \in \mathcal{B}(\mathbb{R}^d)\}. \tag{20}$$

2. Given $\{\mu_n\}$ and μ probability measures on $\mathbb{R}^d$, if $\lim_{n\to\infty} ||\mu_n - \mu||_{\text{var}} = 0$, then μ_n converges to μ **in the strong sense**.

Exercise. verify that if $||\mu_n - \mu||_{\text{var}} \to 0$ then $\mu_n \Rightarrow \mu$.

It is often inconvenient to work with strong convergence. For example, we know that $\lim_{n\to\infty} \frac{1}{n} = 0$. If $\mu_n := \delta_{1/n}$ for all $n \geq 0$, then $\mu = \delta_0$. So, naturally, μ_n should be getting closer and closer to μ. However, if you look at the total variation distance, for all $n \geq 1$, $||\mu_n - \mu||_{\text{var}} = 1$. Hence, μ_n does not convergence to μ in the strong sense. However, if we relax our standards, $\mu_n \Rightarrow \mu$ because for all $\varphi \in C_b(\mathbb{R})$, $\mu_n(\varphi) = \varphi\left(\frac{1}{n}\right)\varphi(0) = \mu(\varphi)$ where the convergence follows from continuity.

1. Let $\{X_n\}$ and X be $\mathbb{R}^d$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.
   (a) If $X_n \to X$ in probability, then $X_n \to X$ in distribution.
   (b) If $X_n \to X$ in distribution and $X \equiv c$ for some constant c, then $X_n \to X$ in probability.
2. Let $\{X_n\}$ and X be $\mathbb{R}$-valued random variables such that X_n has the distribution function F_n for all $n \geq 1$ and X has distribution function F. Then,
   (a) If $X_n \to X$ in distribution, then $\lim_{n\to\infty} F_n(x) = F(x)$at every continuous point x of F.
   (b) If X_n has density f_n for all $n \geq 1$ and X has density f, and $f_n \to f$ a.e. with respect to the Lebesgue measure on $\mathbb{R}$, then $X_n \to X$ in distribution.

Proposition 2. Let $\{\mu_n\}$ and μ be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. If for every subsequence $\{n_k\} \subseteq \mathbb{N}$, there exists a further subsequence, $\{n_{k_l} \subseteq \{n_k\}$ such that $\mu_{n_{k_l}}\} \Rightarrow \mu$, then $\mu_n \Rightarrow \mu$.

Proposition 3. If $\{\mu_n \mid n \geq 1\}$ is a sequence of probability measures on $\mathbb{R}^d$ and $\mu_n \Rightarrow \mu$ and $\mu_n \Rightarrow \nu$, then $\mu = \nu$ (the limit of weak convergence is unique).

Proof. We can very briefly sketch the argument: for all $\varphi \in C_b(\mathbb{R}^d)$,

$$\begin{aligned} \mu(\varphi) &= \lim_{n\to\infty} \mu_n(\varphi) = \nu(\varphi) \\ &\Rightarrow \text{ integrals match on all continuous and bounded functions} \\ &\Rightarrow \mu = \nu. \end{aligned}$$

∎

The following proposition will be useful for the homework.

Proposition 4. Suppose $\mu_n \Rightarrow \mu$. Then:

1. For all open sets $G \subseteq \mathbb{R}^d$,
$$\mu(G) \leq \liminf_n \mu_n(G). \tag{21}$$
2. For all closed sets $F \subseteq \mathbb{R}^d$,
$$\mu(F) \geq \limsup_n \mu_n(F). \tag{22}$$

Proof. We will only prove (i). Given an open set $G \subseteq \mathbb{R}^d$, there exists a sequence $\{\varphi_k \mid k \geq 1\} \subseteq C_b(\mathbb{R}^d)$ such that $\varphi_k \uparrow \chi_G$. By **(MON)** or **(DOM)** ,
$$\mu(G) = \lim_{k\to\infty} \mu(\varphi_k) = \lim_{k\to\infty} \lim_{n\to\infty} \mu_n(\varphi_k) \leq \liminf_{n\to\infty} \mu_n(G),$$
where the second equality follows from weak convergence, and the final inequality follows from the fact that for all $k \geq 1$, $\varphi_k \leq \chi_G$. □

In fact, if $\mu(G) \leq \liminf_n \mu_n(G)$ for every open set $G \subseteq \mathbb{R}^d$, then $\mu_n \Rightarrow \mu$.

1.4 Tightness of a Family, Class, or Collection of Probability Measures

Definition 6. Let $\{\mu_n \mid n \geq 1\}$ be a sequence of probability measures on $\mathbb{R}^d$. We say that $\{\mu_n\}$ is **tight** if for all $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq \mathbb{R}^d$ such that
$$\sup_n \mu_n(K_\varepsilon^c) < \varepsilon. \tag{23}$$
This is telling us that we can make the whole family uniformly small.

Remark. "tightness" means that the mass is concentrated in a way that is uniform for the μ_n's. For example:

- $\{\mu_n = \gamma_{0,1/n} \mid n \geq 1\}$ is a tight family (the variance goes down as $n \to \infty$.
- $\{\nu_n = \gamma_{0,n} \mid n \geq 1\}$ is *not* tight: as n grows, the variance gets more spread out.

Theorem 5 (Prokhorov's Theorem)**.** *Let $\{\mu_n \mid n \geq 1\}$ be a sequence of probability measures on $\mathbb{R}^d$. Then:*

1. *If there exists a probability measure on $\mathbb{R}^d$ such that $\mu_n \Rightarrow \mu$, then $\{\mu_n \mid n \geq 1\}$ is tight.*
2. *If $\{\mu_n \mid n \geq 1\}$ is tight, then there exists a subsequence $\{n_k \mid n \geq 1\} \subseteq \mathbb{N}$ and a probability measure μ on $\mathbb{R}^d$ such that along the subsequence, $\mu_{n_k} \Rightarrow \mu$ as $k \to \infty$.*

Proof. **(i).** Assume that $\mu_n \Rightarrow \mu$. For a contradiction, assume that $\{\mu_n \mid n \geq 1\}$ is not tight: there exists an $\eta > 0$ such that for all compact sets $K \subseteq \mathbb{R}^d$,
$$\sup_n \mu_n(K_n^c) > \eta.$$
We will use this statement to extract a subseqeuence: for all $k \geq 1$, there exists an n_k such that $\mu_{n_k}(\overline{B(0,k)}^c) > \eta$. Then, for every $R > 0$ when k is sufficiently large, i.e., $k \geq R$, we get from $\mu_{n_k} \Rightarrow \mu$:
$$\begin{aligned}\mu(B(0,R)) &\leq \liminf_{k\to\infty} \mu_{n_k}(B(0,R)) \text{ (weak convergence)}\\ &\leq \liminf_{k\to\infty} \mu_{n_k}(B(0,k))\\ &\leq 1-\eta.\end{aligned}$$

Therefore, for all $R > 0$,

$$\mu(B(0,R)) \leq 1-\eta \Rightarrow \mu(\mathbb{R}^d) < 1-\eta,$$

where the implication follows from sending $R \to \infty$ and **(MON)** . However, this is not possible, since $\mu(\mathbb{R}^d) = 1$ since μ is a probability measure. $\square$

Task: Give a rigorous proof of **(i)** of Prokhorov's Theorem. You may use:

1. Riesz-Representation Theorem
2. Stone-Weierstrass Theorem (separability of space of continuous functions on compact sets).

Theorem 6. *Let $\{\mu_n \mid n \geq 1\}$ and μ be probability measures. If $\mu_n \Rightarrow \mu$, then*

$$\lim_{n\to\infty} \hat{\mu_n}(\xi) = \hat{\mu}(\xi) \ \forall \xi \in \mathbb{R}^d,$$

(so weak convergence of measure gives us convergence of characteristic functions) and this convergence is uniform on compact sets, i.e., for all compact $K \subseteq \mathbb{R}^d$,

$$\lim_{n\to\infty} \sup_{\xi\in K} |\hat{\mu}_n(\xi) - \hat{\mu}(\xi)| = 0. \tag{24}$$

Furthermore, for all $\varphi \in C_b(\mathbb{R}^d)$ if $\{\varphi_n \mid n \geq 1\} \subseteq C_b(\mathbb{R}^d)$ such that $\sup_n ||\varphi_n||_n < \infty$ and $\varphi_n \Rightarrow \varphi$ uniformly on compact sets, then $\lim_{n\to\infty} \mu_n(\varphi_n) = \mu(\varphi)$.

Proof. For all $\xi \in \mathbb{R}^d$, the map $x \in \mathbb{R}^d \mapsto e^{i(x,\xi)} \in \mathbb{C}$ is continuous and bounded. Hence,

$$\mu_n \Rightarrow \mu \Rightarrow \hat{\mu}_n(\xi) \to \hat{\mu}(\xi).$$

We will now prove the last statement. Sicne $\mu_n \Rightarrow \mu$, the sequence $\{\mu_n \mid n \geq 1\}$ is tight for all $\varepsilon > 0$. Hence, there exists a compact set $K_\varepsilon \subseteq \mathbb{R}^d$ such that $\sup_n \mu_n(K_\varepsilon^c) < \varepsilon$ and $\mu(K_\varepsilon^c) < \varepsilon$. Hence,

$$\begin{aligned}
|\mu_n(\varphi_n) - \mu(\varphi)| &= |\mu_n(\varphi_n) - \mu_n(\varphi)| + \underbrace{|\mu_n(\varphi) - \mu(\varphi)|}_{\to 0 \text{ as } n\to\infty} \\
&\leq |\mu_n(\chi_{K_\varepsilon} \cdot (\varphi_n - \varphi))| + \left|\mu_n(\chi_{K_\varepsilon^c} \cdot (\varphi_n - \varphi))\right| \\
&\leq \underbrace{\sup_{x\in K_\varepsilon} |\varphi_n(x) - \varphi(x)|}_{\to 0 \text{ as } n\to\infty} \cdot 1 + \underbrace{\sup_n \mu_n(K_\varepsilon^c)}_{<\varepsilon} \cdot \underbrace{(\sup_n ||\varphi_n||_k + ||\varphi||_n)}_{<\infty},
\end{aligned}$$

where the first convergence to zero occurs since $\varphi_n \to \varphi$ uniformly on compact sets K_ε, the second term is assumed to be less than ε and the final term was assumed to be finite. Therefore, $\mu_n(\varphi_n) \to \mu(\varphi)$.

In particular, if $\{\xi_n \mid n \geq 1\} \subseteq \mathbb{R}^d$ such that $\xi_n \to \xi$ as $n \to \infty$, then $\hat{\mu}_n(\xi_n) \to \hat{\mu}(\xi)$ as $n \to \infty$, because we could simply take $\varphi_n = e^{i(\cdot,\xi/n)}$ and $\varphi = e^{i(\cdot,\xi)}$ $(*)$.

We need to now prove that for all compact sets $K \subseteq \mathbb{R}^d$,

$$\sup_{\xi\in K} |\hat{\mu}_n(\xi) - \hat{\mu}(\xi)| \to 0 \text{ as } n \to \infty.$$

For a contradiction, assume otherwise: there exists K compact, $\eta > 0$, a subsequence $\{n_k\} \subseteq \mathbb{N}$ such that

$$\sup_{\xi\in K} |\hat{\mu}_{n_k}(\xi) - \hat{\mu}(\xi)| > \eta.$$

Thus, there exists a subsequence $\xi_{n_k} \in K$ such that

$$|\hat{\mu}_{n_k}(\xi_{n_k}) - \hat{\mu}(\xi_{n_k})| > \eta.$$

Since K is compact, $\{\xi_{n_k} \mid k\} \subseteq K \Rightarrow$ that there exists $\{n_{k_l} \subseteq \{n_k\}$ and there exists a $\xi_0 \in K$ such that $\xi_{n_{k_l}} \to \xi_0$. Hence,

$$\left|\hat{\mu}_{n_{k_l}}(\xi_{n_{k_l}}) - \mu(\xi_{n_{k_l}})\right| \leq \underbrace{\left|\hat{\mu}_{n_{k_l}}(\xi_{n_{k_l}}) - \hat{\mu}(\xi_0)\right|}_{\to 0 \text{ by } (*)} + \underbrace{\left|\hat{\mu}(\xi_0) - \hat{\mu}(\xi_{n_{k_l}})\right|}_{\to 0 \text{ since } \hat{\mu} \text{ is cts}}$$

Contradiction! Therefore, $\hat{\mu}_n \to \hat{\mu}$ uniformly on compact sets. □

2 Lecture 6

The following theorem is another application of Levy's Continuity Theorem.

Theorem 7 (Levy's Equivalence Theorem)**.** *Let $\{X_n \mid n \geq 1\}$ be independent random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Set $S_n := \sum_{j=1}^n X_j$. Then, the following are equivalent:*

1. *S_n converges a.s. to some random variable S, i.e., $\sum_{j=1}^n X_j$ converges almost surely.*
2. *S_n converges in probability to some random variable S.*
3. *S_n converges in distribution to S.*

The idea of this theorem is that it's very hard for independent random variables to converge, so if they converge in one sense, they converge in all senses.

Proof. This time, we only need to show that $(iii) \Rightarrow (ii)$. Assume that $\mu_n := \mathcal{L}_{S_n}$ for every $n \in \mathbb{N}$ and $\mu_n \Rightarrow \mu$ for some probability measure μ. We will show that $\{S_n\}$ forms a Cauchy sequence in probability. Recall the definition of that:

$$\forall\, \varepsilon > 0,\ \exists\, N > 1 \ \text{ s.t. } \sup_{m \geq N} \mathbb{P}\left(|S_m - S_n| > \varepsilon\right) \leq \varepsilon.$$

For a contradiction, assume otherwise. Then, there exists an $\eta > 0$ and a subsequence $\{n_k\}$ such that along the subsequence, the Cauchy condition above is violated:

$$\mathbb{P}\left(|S_{n_{k+1}} - S_{n_k}| > \eta\right) \geq \eta. \tag{25}$$

For every k, set $v_k := \mathcal{L}_{S_{n_{k+1}} - S_{n_k}} \Rightarrow \mu_{n_{k+1}} = \mu_{n_k} * v_k$ (since this is a sum of independent random variables) $(*)$. Since $\mu_n \Rightarrow \mu$, the sequence $\{\mu_n \mid n \geq 1\}$ is tight $\Rightarrow$ for all $\varepsilon > 0$, there exists an $M > 0$ such that

$$\sup_n \mu_n(\overline{B(0,M)}^c) \leq \varepsilon.$$

Hence, for all $k \geq 1$,

$$\begin{aligned}
v_k(\overline{B(0,2M)}^c) &= \mathbb{P}\left(|S_{n_{k-1}} - S_{n_k}| > 2M\right) \\
&\leq 2 \sup_n \mathbb{P}\left(|S_n| > M\right) \\
&= 2 \sup_n \mu_n(\overline{B(0,M)^c}) \\
&\leq 2\varepsilon.
\end{aligned}$$

This shows that the sequence $\{v_k\}$ is tight. Hence, by the second part of the previous theorem, there exists a subsequence $\{k_l \mid l > 1\} \subseteq \mathbb{N}$ and a probability measure v such that $v_{k_l} \Rightarrow v$ as $l \to \infty$ $(**)$.

Combining $(*)$ and $(**)$, we obtain that $\mu = \mu * v$.

Proposition / Remark: Let v be a probability measure on $\mathbb{R}^d$. If there exists a probability measure μ on $\mathbb{R}^d$ such that $\mu = \mu * v$, then $v = \delta_0$.

Proof: For every $\xi \in \mathbb{R}^d$,

$$\hat{\mu}(\xi) = \hat{\mu}(\xi) \cdot \hat{v}(\xi) \Rightarrow \hat{v}(\xi) = 1 \text{ if } \hat{\mu}(\xi) \neq 0.$$

Hence, there exists some positive number r such that $\hat{\mu}(\xi) \neq 0$ for $\xi \in B(0,r)$. Hence, for all $\xi \in B(0,r)$,

$$\hat{v}(\xi) = 1.$$

Hence, for all $\xi \in B(0,r)$,

$$\int_{\mathbb{R}^d} \cos(x,\xi) v(dx) = 1 \Rightarrow (x,\xi) = 0 \text{ mod } 2\pi \text{ for } v\text{- a.e } x.$$

Given any unit vector $e \in \mathbb{R}^d$, choose $\xi_1, \xi_2 \in B(0,r)$, ξ_1 and ξ_2 are both along the direction e and $\xi_2 = \rho\xi_1$ where $\rho \notin \mathbb{Q}$. Hence,

$$\begin{aligned}
&\Rightarrow (x,\xi_1) = 0 \text{ mod } 2\pi \text{ for } v \text{ almost every } x. \\
&\Rightarrow (x,\xi_2) = \rho(x,\xi_1) = 0 \text{ mod } 2\pi \text{ for } v \text{ almost every } x. \\
&\Rightarrow (x,\xi_1) = 0 \text{ for } v \text{ almost every } x. \\
&\Rightarrow (x,e) = 0 \ v \text{ almost everywhere for all unit vectors} \\
&\Rightarrow v = \delta_0.
\end{aligned}$$

Going back to the main proof, we have therefore proven that $v_{k_l} \Rightarrow \delta_0$, i.e., $S_{n_{k_l}+1} - S_{n_{k_l}} \to 0$ in distribution. Since this is a constant, we have therefore that

$$S_{n_{k_l+1}} - S_{n_{k_l}} \to 0 \text{ in probability.}$$

But, this contradicts (25). Hence, we have that $\{S_n\}$ forms a Cauchy sequence in probability $\Rightarrow S_n \to S$ in probability for some random variable S. $\square$

3 Infinitely Divisible Laws

Definition 7 (Infinitely Divisible)**.** Let μ be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then, we say that μ is **infinitely divisible** if for every $n \geq 1$, there exists a $v_{(n)}$ probability measure on $(R^d, \mathcal{B}(\mathbb{R}^d))$ such that

$$\underbrace{v_{(n)} * v_{(n)} * \ldots * v_{(n)}}_{n \text{ times}} = \mu, \tag{26}$$

Equivalently, $(\hat{v}(\xi))^n = \hat{\mu}(\xi)$ for all $\xi \in \mathbb{R}^d$.

Notation-wise, we write $I(\mathbb{R}^d)$ as the collection of all the infinitely divisible laws on $\mathbb{R}^d$. If $\mu \in I(\mathbb{R}^d)$, then we will write $v_{(n)}$ in the definition as $\mu_{1/n}$, i.e., this means that

$$\underbrace{\mu_{1/n} * \mu_{1/n} * \ldots * \mu_{1/n}}_{n \text{ copies}} = \mu.$$

We introduce a few remarks:

1. If $\mu \in I(\mathbb{R}^d)$, then for every $n \geq 1$,

$$(\hat{\mu}_{1/n}(\xi))^n = \hat{\mu}(\xi) \ \forall \xi \in \mathbb{R}^d.$$

Heuristically, we want to study the "nth" root of $\hat{\mu}$.

2. If $\mu, \nu \in I(\mathbb{R}^d)$, then for every $n \in \mathbb{N}$,

$$(\mu_{1/n} \hat{*} \nu_{1/n}) = \hat{\mu}(\xi) * \hat{\nu}(\xi) = \mu \hat{*} \nu(\xi).$$

This implies that I is closed under convolution:

$$\mu_{1/n} * \nu_{1/n} = (\mu * \nu)_{1/n} \Rightarrow \nu * \nu \in I(\mathbb{R}^d). \tag{27}$$

3. If $\{\mu_k\}$ is a sequence of infinitely divisible laws, $\mu_k \Rightarrow \mu$ as $k \to \infty$ and for all $n \geq 1$, $\mu_{k,1/n} \Rightarrow \nu_{(n)}$ as $k \to \infty$ for some probability measure $v_{(n)}$, then $\mu \in I(\mathbb{R}^d)$ and $\mu_{1/n} = \nu_{(n)}$.

3.1 Examples of Infinitely Divisible Laws

1. Trivial Examples: for $\vec{m} \in \mathbb{R}^d$, $\delta_m \in I(\mathbb{R}^d)$.

$$(\delta_m)_{1/n} = \delta_{m/n}.$$

2. Gaussian measures: for all $m \in \mathbb{R}^d$, for all $C = (C_{ij})_{d\times d} \geq 0$ (non-negative definite) such that

$$\hat{\gamma_{m,C}}(\xi) = e^{i(m,\xi)} e^{-1/2(\xi, C\xi)}$$

Hence, for all $n \in N$,

$$\left(\gamma_{m/n,C/n}\right)^n = \hat{\gamma}_{m,C}(\xi)$$

for all $\xi \in \mathbb{R}^d$. Hence, all Gaussian measures are infinitely divisible with $(\gamma_{m,C})_{1/n} = \gamma_{m/n,C/n}$.

3. Poisson Measures.
   (a) Standard Poisson distribution / measure is supported on $\{0, 1, 2, ...\}$. Given $\alpha > 0$, let π_α be the Poisson distribution with parameter α, i.e., for all $k \geq 0$,

   $$\pi_\alpha(\{k\}) = e^{-\alpha} \frac{\alpha^k}{k!}$$

   Equivalently, write,

   $$\pi_\alpha = \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} (\delta_k) = \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} (\delta_1)^{*k}$$

   (b) General Poisson Measure on $\mathbb{R}^d$:
      i. Given $\alpha > 0$ and a probability measure ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$,

      $$\pi_{\alpha,\nu} = \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} \nu^{(*k)}$$

      Let's try to understand this from the random variable point of view. Let $\{X_n\}$ be iid random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathcal{L}_{X_1} = \nu$. Let N be a random variable on the same probability space independent of $\{X_n\}$, $\mathcal{L}_N = \pi_\alpha$. Define $S = \sum_{i=1}^{N} X_j$. Hence, for all $\omega \in \Omega$, S is defined point-wise as:

$$S(\omega) = \sum_{j=1}^{N(\omega)} X_j(\omega). \tag{28}$$

Then, for all $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}
\mathbb{P}(S \in B) &= \sum_{k=0}^{\infty} \mathbb{P}(S \in B, N = k) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\left(\sum_{j=1}^{k} X_j \in B, N = k\right) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\left(\sum_{j=1}^{k} X_j \in B\right) \mathbb{P}(N = k) \\
&= \sum_{k=0}^{\infty} \mathbb{P}\left(\sum_{j=1}^{k} X_j \in B\right) e^{-\alpha}\frac{\alpha^k}{k!} \\
&= \sum_{k=0}^{\infty} e^{-\alpha}\frac{\alpha^k}{k!} v^{*k}(B) \\
&= \pi_{\alpha,\nu}(B).
\end{aligned}$$

Note that we have a taylor expansion of the exponential, and so for every $\xi \in \mathbb{R}^d$,

$$\hat{\pi_{\alpha,v}}(\xi) = \sum_{k=0}^{\infty} e^{-\alpha}\frac{\alpha^k}{k!}(\hat{v}(\xi))^k = e^{-\alpha}e^{\alpha\hat{v}(\xi)} = e^{\alpha(\hat{v}(\xi)-1)}.$$

This shows that $\pi_{\alpha,\nu} \in I(\mathbb{R}^d)$ and for all $n \geq 1$,

$$(\pi_{a,\nu})_{1/n} = e^{\alpha/n(\hat{v}(\xi)-1)} = \pi_{\alpha/n,v}.$$

Notation. Given $\alpha > 0$, probability measure v on $\mathbb{R}^d$,

$$\hat{\pi}_{\alpha,v}(\xi) = \exp(\alpha(\hat{v}(\xi) - 1)) = \exp\left(\alpha \int_{\mathbb{R}^d} (e^{i(x,\xi)} - 1)\nu(dx)\right)$$

Set $M := \alpha\nu$, $M(\mathbb{R}^d) = \alpha$. Set $\pi_M := \pi_{\alpha,\nu}$, i.e.,

$$\hat{\pi}_M(\xi) = \exp\left(\int_{\mathbb{R}^d} (e^{i(x,\xi)} - 1)M(dx)\right).$$

Further, $M(\{0\})$ does not affect π_M, so WLOG we assume that $M(\{0\}) = 0$. Set

$$\mathcal{M}_0(\mathbb{R}^d) := \{M \mid M \text{ is a finite Borel measure on } (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \text{ such that } M(\{0\}) = 0\}. \tag{29}$$