Lecture Notes: Weak Convergence and Characteristic Functions

December 14

1 Weak Convergence of Measures

Recall that given a metric space S (e.g., $S = \mathbb{R}, \mathbb{R}^d$), we denote by $\mathrm{Prob}(S)$ the set of all Borel probability measures on S.

Definition 1. *Let $\mu_n \in Prob(S)$ and $\mu \in Prob(S)$. We say that $\mu_n \Rightarrow \mu$ **weakly** if, by definition:*

$$\int_S g\,d\mu_n \to \int_S g\,d\mu \quad \forall g \in C_b(S)$$

where $C_b(S)$ is the space of continuous bounded functions on S.

1.1 The Portmanteau Theorem

We proved the Portmanteau Theorem, which states that the following are equivalent:

1. $\mu_n \Rightarrow \mu$ (weakly).
2. $\int g\,d\mu_n \to \int g\,d\mu$ for all bounded and uniformly continuous g.
3. $\limsup_n \mu_n(F) \le \mu(F)$ for all closed sets F.
4. $\liminf_n \mu_n(U) \ge \mu(U)$ for all open sets U.
5. $\lim_n \mu_n(A) = \mu(A)$ for all Borel sets A with $\mu(\partial A) = 0$ (continuity sets).

2 Convergence in Distribution

Let $\{X_n\}$ be a sequence of real-valued random variables (RVs), and let X be another real-valued RV.

Definition 2. *We say that $\{X_n\}$ converges to X in distribution (denoted $X_n \xrightarrow{d} X$ or $X_n \Rightarrow X$) if $\mu_{X_n} \Rightarrow \mu_X$ weakly. That is, if $\mu_{X_n} \to \mu_X$ in $Prob(\mathbb{R})$.*

Observations: From the Portmanteau theorem, $X_n \xrightarrow{d} X$ is equivalent to:

- $\int g\,d\mu_{X_n} \to \int g\,d\mu_X \iff E[g(X_n)] \to E[g(X)]$ for all $g \in C_b(\mathbb{R})$.
- $E[g(X_n)] \to E[g(X)]$ for all bounded uniformly continuous g.

- $\limsup P(X_n \in F) \leq P(X \in F)$ for all closed sets F.
- $\liminf P(X_n \in U) \geq P(X \in U)$ for all open sets U.
- $\lim P(X_n \in A) = P(X \in A)$ for all Borel sets A such that $P(X \in \partial A) = 0$.

Example 1. *Consider $S = [0,1]$. Let $X_n = \frac{1}{n}$ and $X = 0$ almost surely (a.s.). Let $F = \{0\}$ (closed) and $U = (0,1)$ (open).*

- *For any $g \in C([0,1])$, $E[g(X_n)] = g(1/n) \to g(0) = E[g(X)]$.*
- *However, strictly examining the open set $U = (0,1)$:*

$$\mu_{X_n}(U) = P\left(\frac{1}{n} \in (0,1)\right) = 1 \text{ (for } n > 1\text{)}$$

$$\mu_X(U) = P(0 \in (0,1)) = 0$$

Here, $\liminf \mu_{X_n}(U) = 1 \geq 0 = \mu_X(U)$, which satisfies condition (4) of the theorem.

2.1 Hierarchy of Convergence

Theorem 1.

$$X_n \to X \text{ a.s.} \implies X_n \to X \text{ in prob.}$$

Theorem 2.

$$X_n \to X \text{ in prob.} \implies X_n \xrightarrow{d} X$$

Proof Sketch. Let g be bounded and uniformly continuous. We want to show $E[g(X_n)] \to E[g(X)]$.

First, we show $g(X_n) \to g(X)$ in probability. Given $\epsilon > 0$, since g is uniformly continuous, there exists $\delta > 0$ such that if $a, b \in \mathbb{R}$ and $|a - b| < \delta$, then $|g(a) - g(b)| < \epsilon$. Thus:

$$|X_n(\omega) - X(\omega)| < \delta \implies |g(X_n(\omega)) - g(X(\omega))| < \epsilon$$

This implies:

$$P(|g(X_n) - g(X)| > \epsilon) \leq P(|X_n - X| > \delta)$$

Since $X_n \to X$ in probability, the RHS goes to 0, so $g(X_n) \to g(X)$ in probability.

Since g is bounded, by the Bounded Convergence Theorem (for convergence in probability), we have:

$$E[g(X_n)] \to E[g(X)]$$

By the Portmanteau theorem, this implies $X_n \xrightarrow{d} X$. □

3 Convergence of CDFs

Theorem 3. *$X_n \xrightarrow{d} X$ if and only if $F_{X_n}(t) \to F_X(t)$ for all points t where F_X is continuous.*

Proof Sketch. (⇒) Let $F_X(t) = \mu_X((-\infty, t])$. Note that F_X is continuous at t implies $\mu_X(\{t\}) = 0$. Since $\partial(-\infty, t] = \{t\}$, if F_X is continuous at t, then $\mu_X(\partial(-\infty, t]) = 0$. By the Portmanteau theorem:

$$\mu_{X_n}(-\infty, t] \to \mu_X(-\infty, t] \implies F_{X_n}(t) \to F_X(t)$$

(⇐) Let $C = \{t \in \mathbb{R} : F_X \text{ is continuous at } t\}$. Then $F_{X_n}(t) \to F_X(t)$ for all $t \in C$. Since F_X is non-decreasing, the set of discontinuity points $\mathbb{R} \setminus C$ is countable.

We will prove that for any open set $U \subset \mathbb{R}$:

$$\liminf \mu_{X_n}(U) \geq \mu_X(U)$$

Let $Y = \{(a, b) : a, b \in C\}$. We first prove that $\mu_{X_n}(I) \to \mu_X(I)$ for all $I \in Y$. For $I = (a, b) \in Y$:

$$\mu_{X_n}(I) = F_{X_n}(b-) - F_{X_n}(a)$$

$$\mu_X(I) = F_X(b-) - F_X(a) = F_X(b) - F_X(a) \quad \text{(since } b \in C\text{)}$$

Since $a \in C$, $F_{X_n}(a) \to F_X(a)$. It is enough to show $\lim F_{X_n}(b-) = F_X(b)$.

Step 1: Since F_{X_n} is non-decreasing:

$$F_{X_n}(b-) \leq F_{X_n}(b) \to F_X(b) \implies \limsup F_{X_n}(b-) \leq F_X(b)$$

Step 2: Since F_X is continuous at b, given $\epsilon > 0$, there exists δ such that $b - \delta \in C$ and $F_X(b - \delta) > F_X(b) - \epsilon$. For $n > n_0$:

$$F_{X_n}(b-) \geq F_{X_n}(b - \delta) \to F_X(b - \delta) > F_X(b) - \epsilon$$

$$\implies \liminf F_{X_n}(b-) \geq F_X(b) - \epsilon$$

Combining Step 1 and 2, we get convergence for intervals in Y.

Step 3: Note that if $I_1, I_2 \in Y$, then $I_1 \cap I_2 \in Y$.

$$\mu_{X_n}(I_1 \cup I_2) = \mu_{X_n}(I_1) + \mu_{X_n}(I_2) - \mu_{X_n}(I_1 \cap I_2) \to \mu_X(I_1 \cup I_2)$$

By induction, this holds for any finite union $I_1 \cup \cdots \cup I_k$.

Step 4: Any open set U can be written as $U = \bigcup_{k \geq 1} I_k$ where $I_k \in Y$.

$$\mu_X(U) = \lim_{k \to \infty} \mu_X\left(\bigcup_{i=1}^{k} I_i\right)$$

Given $\epsilon > 0$, there exists k such that $\mu_X(U) \leq \mu_X(\bigcup_{i=1}^{k} I_i) + \epsilon$.

$$\leq \liminf \mu_{X_n}\left(\bigcup_{i=1}^{k} I_i\right) + \epsilon \leq \liminf \mu_{X_n}(U) + \epsilon$$

Letting $\epsilon \to 0$, we get $\liminf \mu_{X_n}(U) \geq \mu_X(U)$, which satisfies the Portmanteau condition.

□

4 Central Limit Theorem (CLT)

Theorem 4 (Normalized / Standard Version)**.** *Given a sequence* (X_n) *of i.i.d. random variables with* $E[X_1] = 0$ *and* $Var(X_1) = 1$*. Let* $S_n = X_1 + \cdots + X_n$*. Then:*

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

where $\mathcal{N}(0,1)$ *has the density* $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$*.*

5 Characteristic Functions

The characteristic function of a random variable X is a function $\varphi_X : \mathbb{R} \to \mathbb{C}$ defined by:

$$\varphi_X(t) = E[e^{itX}] = E[\cos(tX)] + iE[\sin(tX)]$$

5.1 Properties

1. $\varphi_X(0) = E[e^0] = 1$.
2. $|\varphi_X(t)| \leq E[|e^{itX}|] = 1$.
3. If X has a density f_X, then $\varphi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x)\, dx$.
4. **Fourier Connection:** The characteristic function is essentially the Fourier transform of the probability density.
5. If X_1, X_2 are independent, $\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)$.
6. Scaling: $\varphi_{cX}(t) = \varphi_X(ct)$.

5.2 Fourier Analysis Review

If $g \in L^1(\mathbb{R})$, we define:

$$\hat{g}(t) = \int_{\mathbb{R}} e^{-itx} g(x) dx$$

This is well defined. However, we often work with functions that are not immediately in L^1.

We consider the **Schwartz Space** $\mathcal{S}(\mathbb{R})$, which is the space of smooth functions that vanish at infinity (and their derivatives vanish as well). $\mathcal{S}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$. The fourier transform of functions in the Schwartz space is again in the Schwartz space and so we can apply it multiple times.

Exercise 1. *The Fourier transform of the Gaussian function* $e^{-x^2/2}$ *is* $e^{-t^2/2}$*.*

Hint: if $G(t) = \hat{f}(t)$ *then what is the relation between* $G'(t)$ *and* $\hat{f}(t)$ *?*

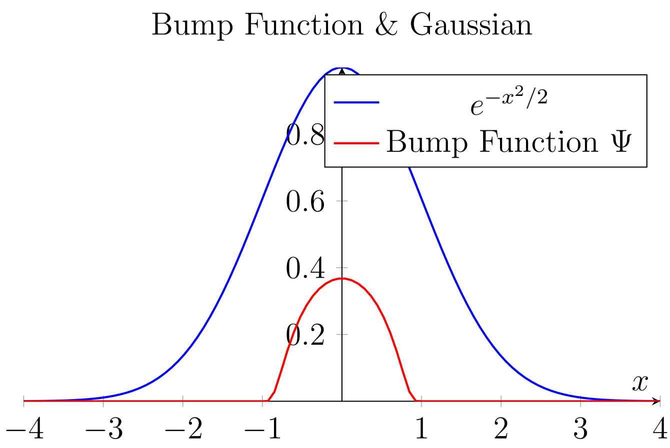

Figure 1: **Schwartz space functions**

5.3 Proof Sketch of CLT using Characteristic Functions

Assume $E[X] = 0$ and $E[X^2] = 1$. The Taylor expansion of φ_X around 0 is:

$$\varphi_X(s) = 1 + iE[X]s - \frac{E[X^2]s^2}{2} + o(s^2) = 1 - \frac{s^2}{2} + o(s^2)$$

Now consider the normalized sum $Z_n = \frac{S_n}{\sqrt{n}}$. Since S_n is a sum of i.i.d. variables:

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \left[\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Plugging in the expansion for large n:

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) \approx \left(1 - \frac{t^2}{2n}\right)^n$$

As $n \to \infty$:

$$\lim_{n\to\infty} \left(1 - \frac{t^2}{2n}\right)^n = e^{-t^2/2}$$

This limit, $e^{-t^2/2}$, is exactly the characteristic function of the standard normal distribution $\mathcal{N}(0, 1)$.

5.4 Levy Continuity Theorem

Theorem 5. *$X_n \xrightarrow{d} X$ if and only if $\varphi_{X_n}(t) \to \varphi_X(t), \forall t \in \mathbb{R}$*

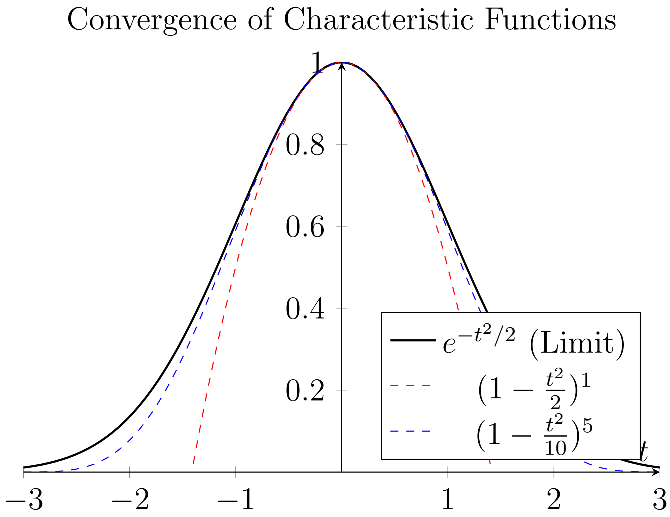

Figure 2: **The Approximation**