Probability Lecture Notes

December 11

1 Basic Notions

1.1 Distribution of a Random Variable

Let $X : \Omega \to \mathbb{R}$ be a random variable. We consider $\mu_X \in \text{Prob}(\mathbb{R})$, defined as:

$$\mu_X(A) := \mathbb{P}(X \in A)$$

We say that X and Y are equal in distribution (or in Law), denoted $X \overset{d}{=} Y$, if and only if $\mu_X = \mu_Y$.

Remark 1. *$X \overset{d}{=} Y \iff \mu_X((-\infty, t]) = \mu_Y((-\infty, t])$ for all $t \in \mathbb{R}$.*
This is because the set system $\{(-\infty, a] : a \in \mathbb{R}\}$ is a π-system that generates the Borel σ-algebra $\mathcal{B}(\mathbb{R})$.

1.2 Change of Variables Formula

For a continuous function f:

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(t)\, d\mu_X(t)$$

2 Cumulative Distribution Function (CDF)

For a probability measure $\mu \in \text{Prob}(\mathbb{R})$, the CDF is a function $F_\mu : \mathbb{R} \to \mathbb{R}$ defined by:

$$F_\mu(t) = \mu((-\infty, t])$$

For a random variable X:

$$F_X(t) = \mathbb{P}(X \leq t) = F_{\mu_X}(t)$$

2.1 Properties of the CDF

The CDF of a probability measure on $\mathbb{R}$ (or of a random variable) is a function $F : \mathbb{R} \to \mathbb{R}$ satisfying:

1. $0 \leq F(t) \leq 1$.
2. F is non-decreasing.

3. Limits at infinity:
$$\lim_{t \to -\infty} F(t) = 0, \quad \lim_{t \to \infty} F(t) = 1$$

4. F is right-continuous:
$$\lim_{x \searrow a} F(x) = F(a)$$

Remark 2. *F is continuous at $a \in \mathbb{R}$ if and only if $\mu(\{a\}) = 0$. This follows from the Monotone Convergence Theorem applied to sets:*

$$\lim_{x \nearrow a} \mu((-\infty, x]) = \mu((-\infty, a)) \iff \mu(\{a\}) = 0$$

Therefore, F_μ is continuous at $a \iff \mu(\{a\}) = 0$.

Remark 3. *Since F is monotone, it has at most a countable set of discontinuities. Moreover, it is differentiable almost everywhere (a.e.).*

Theorem 1. *Given $F : \mathbb{R} \to \mathbb{R}$ satisfying properties 2–4, there exists a unique $\mu \in$ Prob$(\mathbb{R})$ such that $\mu((-\infty, t]) = F(t)$. Therefore, there is a bijection between CDFs and Borel probability measures on $\mathbb{R}$.*

3 Density Functions

Let μ_X be the distribution of a random variable. If $\mu_X \ll \lambda$ (absolute continuity with respect to the Lebesgue measure λ on $\mathbb{R}$), then by the Radon-Nikodym theorem, there exists a function $f_X \geq 0$ such that:

$$d\mu_X = f_X \, d\lambda \quad (\text{or } d\mu_X = f_X(x)dx)$$

and $\int_{\mathbb{R}} f_X(x)dx = 1$. This implies:

$$\forall E \in \mathcal{B}(\mathbb{R}), \quad \mu_X(E) = \int_E f_X(x)dx$$

In this case, f_X is called the **Probability Density Function (PDF)** of X. If $\mu_X \ll \lambda$, then $F'_X = f_X$ almost surely (CDF' = PDF).

3.1 Examples of Distributions

3.1.1 Continuous Uniform Distribution

Measure: $d\mu = \mathbb{I}_{[0,1]}dx$.
PDF: $f(t) = \mathbb{I}_{[0,1]}(t)$.
CDF:
$$F(t) = \mu((-\infty, t]) = \begin{cases} 0 & t < 0 \\ t & t \in [0, 1] \\ 1 & t > 1 \end{cases}$$

"X is drawn uniformly from $[0, 1]$". Similarly for uniform distribution on $[a, b]$, the density is $\frac{1}{b-a}\mathbb{I}_{[a,b]}$.

3.1.2 Bernoulli Distribution

Bernoulli(p) is the distribution of a RV with two values $\{a, b\}$ where $\mathbb{P}(X = a) = p$ and $\mathbb{P}(X = b) = 1 - p$.

$$\mu = p\delta_a + (1-p)\delta_b \quad \text{(Dirac measures)}$$

This distribution has no density with respect to Lebesgue measure. Its CDF is a step function.

3.1.3 Poisson Distribution

Poisson(λ) is the distribution of a RV taking values in $\{0, 1, 2, \dots\}$.

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

This is a discrete measure: $\mu = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \delta_k$. It has no PDF with respect to Lebesgue measure.

3.1.4 Standard Normal Distribution $N(0, 1)$

PDF:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

CDF:

$$F(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Exercise 1. *If $X \sim N(0, 1)$, show that $\mathbb{E}[X] = 0$ and Var$(X) = 1$.*

Generally, for $N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

4 Convergence in Distribution

Let $(S, \mathcal{B}(S))$ be a metric space with the Borel σ-algebra.

Definition 1. *Let $\{\mu_n\} \subset Prob(S)$ and $\mu \in Prob(S)$. We say $\mu_n \to \mu$* ***weakly*** *if:*

$$\int_S g \, d\mu_n \to \int_S g \, d\mu$$

for all $g \in C_b(S)$, where $C_b(S) = \{h : S \to \mathbb{R} \mid h \text{ is continuous and bounded}\}$. Note: $\|g\|_\infty = \sup_{x \in S} |g(x)|$.

Exercise 2. *If weakly $\mu_n \to \mu$ and weakly $\mu_n \to$ d then $\mu =$ d*

Definition 2. *We say that a sequence of random variables $\{X_n\}$ converges* ***in distribution*** *(or in Law) to a random variable X, written $X_n \xrightarrow{d} X$ (or $X_n \Rightarrow X$), if $\mu_{X_n} \to \mu_X$ weakly.*

Remark 4. *$X_n \xrightarrow{d} X \iff \mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for all $g \in C_b(\mathbb{R})$. (By a change of variables)*

5 Portmanteau Theorem

Theorem 2 (Portmanteau Theorem)**.** *Let $(S, \mathcal{B}(S))$ be a metric space. Let $\{\mu_n\} \subset$ Prob(S) and $\mu \in$ Prob(S). The following are equivalent:*

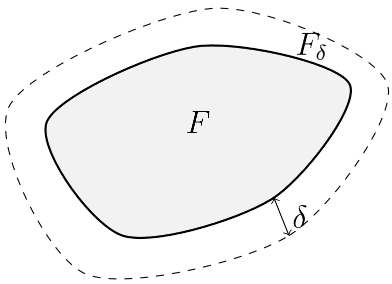
1. *$\mu_n \to \mu$ weakly ($\int g d\mu_n \to \int g d\mu, \forall g \in C_b(S)$).*
2. *$\int g d\mu_n \to \int g d\mu$ for all g bounded and uniformly continuous (or Lipschitz).*
3. *$\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ for all closed sets $F \subset S$.*
4. *$\liminf_{n\to\infty} \mu_n(G) \ge \mu(G)$ for all open sets $G \subset S$.*
5. *$\lim_{n\to\infty} \mu_n(A) = \mu(A)$ for all Borel sets A with $\mu(\partial A) = 0$ (such sets are called μ-continuity sets).*

5.1 Proof of the Portmanteau Theorem

Proof (1) $\implies$ (2): Trivial, since Lipschitz functions are continuous and bounded.

Proof (2) $\implies$ (3): Let F be a closed set. We approximate F from above. Let $\epsilon > 0$. Define $F_\delta := \{x \in S : d(x, F) < \delta\}$ where $d(x, F) = \inf_{y\in F} d(x, y)$. Note that $x \mapsto d(x, F)$ is Lipschitz. Since F is closed, $d(x, F) = 0 \iff x \in F$.

We construct a function g_δ to approximate the indicator $\mathbb{I}_F$.



Define $g_\delta : S \to \mathbb{R}$ by:

$$g_\delta(x) = \left(1 - \frac{d(x, F)}{\delta}\right)^+$$

where $a^+ = \max(a, 0)$. Properties of g_δ:

- g_δ is Lipschitz (composition of Lipschitz functions).
- $0 \le g_\delta \le 1$.
- On F, $d(x, F) = 0 \implies g_\delta(x) = 1$. Thus $\mathbb{I}_F \le g_\delta$.
- On F_δ^c, $d(x, F) \ge \delta \implies g_\delta(x) = 0$. Thus $g_\delta \le \mathbb{I}_{F_\delta}$.

So, $\mathbb{I}_F \le g_\delta \le \mathbb{I}_{F_\delta}$.

Since F is closed, as $\delta \to 0$, $F_\delta \downarrow F$. By continuity of measure from above (using $\bigcap F_{\delta_k} = F$), we have $\mu(F_\delta) \to \mu(F)$. Choose δ such that $\mu(F_\delta) < \mu(F) + \epsilon$.

Then:

$$\mu_n(F) = \int \mathbb{I}_F \, d\mu_n \le \int g_\delta \, d\mu_n$$

Taking limit sup:

$$\limsup_{n\to\infty} \mu_n(F) \le \lim_{n\to\infty} \int g_\delta \, d\mu_n = \int g_\delta \, d\mu \le \int \mathbb{I}_{F_\delta} \, d\mu = \mu(F_\delta) < \mu(F) + \epsilon$$

Since ϵ is arbitrary, $\limsup \mu_n(F) \le \mu(F)$.

Proof (3) $\iff$ (4): By taking complements. Open $G = F^c$. $\mu_n(G) = 1 - \mu_n(F)$.

$$\liminf \mu_n(G) = 1 - \limsup \mu_n(F) \ge 1 - \mu(F) = \mu(G).$$

Proof (3) + (4) $\implies$ (5): Let A be a Borel set with $\mu(\partial A) = 0$. Recall $\partial A = \overline{A} \setminus A^\circ$ and $A^\circ \subset A \subset \overline{A}$. Since $\mu(\partial A) = 0$, $\mu(\overline{A}) = \mu(A^\circ) = \mu(A)$.

$$\limsup \mu_n(A) \le \limsup \mu_n(\overline{A}) \le \mu(\overline{A}) = \mu(A)$$
$$\liminf \mu_n(A) \ge \liminf \mu_n(A^\circ) \ge \mu(A^\circ) = \mu(A)$$

Thus, $\mu_n(A) \to \mu(A)$.

Proof (5) $\implies$ (1): We use the layer cake representation.

Lemma 1. *If $f \in L^1(S, \mu)$ and $f \ge 0$, then $\int_S f d\mu = \int_0^\infty \mu(\{f > t\}) dt$.*

Let $g \in C_b(S)$. Since g is bounded ($m \le g \le M$), we can look at $g - m \ge 0$. So assume without loss of generality that $g \ge 0$.

$$\int g \, d\mu_n = \int_0^\infty \mu_n(\{g > t\}) dt$$

The set $A_t = \{g > t\}$ has boundary $\partial A_t \subset \{g = t\}$. We need $\mu(\partial A_t) = 0$ for condition (5) to apply. It is sufficient to show $\mu(\{g = t\}) = 0$.

Exercise 3. *The set of $t \in \mathbb{R}$ such that $\mu(\{g = t\}) > 0$ is at most countable.*

Since the set of atoms is countable, for almost every t (with respect to Lebesgue measure), $\mu(\{g = t\}) = 0$. Thus, for a.e. t, $\mu_n(\{g > t\}) \to \mu(\{g > t\})$. By the Dominated Convergence Theorem (since measures are bounded by 1),

$$\int_0^\infty \mu_n(\{g > t\}) dt \to \int_0^\infty \mu(\{g > t\}) dt \implies \int g d\mu_n \to \int g d\mu.$$