Math 254B Lecture 1 Notes

Daniel Raban

April 1, 2019

1 Introduction to Probability

1.1 Probability spaces and random variables

Definition 1.1. A **probability space**[1] $(\Omega, \mathcal{F}, \mathbb{P})$ is a measure space such that $\mathbb{P}(\Omega) = 1$. We call Ω the **sample space** and its elements $\omega \in \Omega$ **outcomes**. An **event** is a subset $E \in \mathcal{F}$. The **probability** of an event is $0 \leq \mathbb{P}(E) \leq 1$.

Remark 1.1. We can think of restricting events to lie in the σ-algebra in the following sense: this restricts the amount of information we have and therefore the possible events we can consider.

Definition 1.2. When $|\Omega| < \infty$, and $\mathcal{F} = \mathbb{P}(\Omega)$, we can define a **probability vector** $(p_\omega)_{\omega \in \Omega}$ by $p_\omega := \mathbb{P}(\{\omega\})$.

In this case, $\mathbb{P}(A) = \sum_{\omega \in A} p_\omega$ for all $A \subseteq \Omega$. Also, $p_\omega \geq 0$ and $\sum_\omega p_\omega = 1$.

Definition 1.3. Let $(\Omega, \mathcal{F}), (X, \mathcal{B})$ be measure spaces. An X**-valued random variable** on Ω is a $(\mathcal{F}/\mathcal{B})$-measurable function $\varphi : \Omega \to X$.

Often, $X = \mathbb{R}$ or a finite set. The idea is that once you have an outcome in Ω, the random variable φ tells you what value the outcome corresponds to.

Definition 1.4. $\varphi^{-1}[\mathcal{B}] = \{\varphi^{-1}[A] : A \in \mathcal{B}\}$ is called the σ**-algebra generated by** φ. We denote this as $\sigma(\varphi)$.

Lemma 1.1. *This is indeed a σ-algebra.*

[1] In the 1920s, Kolmogorov realized that measure theory was the perfect language to describe probability in a rigorous mathematical setting.

1.2 Distributions and expectation

Definition 1.5. Let $\varphi : (\Omega, \mathcal{F}, \mathbb{P}) \to (X, \mathcal{B})$ be a random variable. We call the map $A \mapsto \mathbb{P}(\varphi^{-1}[A])$ for $A \in \mathcal{B}$ is called the **image** or **pushforward** of $\mathbb{P}$ under φ. We denote this as $\varphi_*\mathbb{P}$.

Lemma 1.2. *$\varphi_*\mathbb{P}$ is a measure on $(X, \mathcal{B})$.*

Definition 1.6. If $\mathbb{P}$ is clear from context, $\varphi_*\mathbb{P}$ is called the **law** or **distribution** of φ.

Proposition 1.1. *Let $g : X \to \mathbb{R}$ be measurable. Then $g \in L^1(\varphi_*\mathbb{P})$ if and only if $g \circ \varphi \in L^1(\mathbb{P})$. In this case, $\int g \, d\varphi_*\mathbb{P} = \int g \circ \varphi \, d\mathbb{P}$.*

Proof. This is true for indicator functions by definition. It is then true for simple functions by linearity, and the monotone convergence theorem gives that this is true for nonnegative functions. By linearity, it is true for measurable functions. $\square$

Definition 1.7. If f is an $\mathbb{R}$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, we say that f has finite **first moment** if $f \in L^1(\mathbb{P})$. The **expectation** of f is $\mathbb{E}_P[f] = \mathbb{E}[f] = \int f \, d\mathbb{P}$.

1.3 Stochastic processes and independence

Definition 1.8. A **stochastic process** is an indexed family $(\varphi_i)_{i \in I}$ of random variables on the same $(\Omega, \mathcal{F}, \mathbb{P})$. From these, if $\varphi_i : \Omega \to X_i$, define $\varphi : \Omega \to \prod_{i \in I} X_i$ sending $\omega \mapsto (\varphi_i(\omega))_{i \in I}$. φ is a random variable when $\prod_i X_i$ is given the **product σ-algebra** $\bigotimes_i \mathcal{B}_i$, generated by $\{\prod_i A_i : A_i \in \mathcal{B}_i \; \forall i, A_i = X_i \text{ for all but finitely many } i\}$. $\varphi_*\mathbb{P}$ is a probability measure on $(\prod_i X_i, \bigotimes_i \mathcal{B}_i)$ called the **joint distribution** of $(\varphi_i)_i$.

Remark 1.2. Knowing the distribution of each φ_i does not tell you all the information that φ has. In general, φ is much more informative than the individual distributions.

Definition 1.9. Let $A \subseteq \Omega$ have $\mathbb{P}(A) > 0$. For any other $B \subseteq \Omega$, its **conditional probability given** A is $\mathbb{P}(B \mid A) = \mathbb{P}(B \cap A)/\mathbb{P}(A)$.

Lemma 1.3. *Let $\mathbb{P}(A) > 0$. Then $\mathbb{P}(\cdot \mid A)$ is a new probability measure on $(\Omega, \mathcal{F})$ such taht $\mathbb{P}(A \mid A) = 1$.*

Definition 1.10. A and B are **independent** if $\mathbb{P}(B \mid A) = \mathbb{P}(B)$. That is, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition 1.11. $A_1, A_2, \ldots, A_n \subseteq \Omega$ are **independent** if $\mathbb{P}(B_1 \cap \cdots \cap B_n) = \prod_{i=1}^n \mathbb{P}(B_i)$ whenever $B_i \in \{A_i, \Omega \setminus A_i\}$ for all i.

Definition 1.12. If $\mathcal{G}_1, \ldots, \mathcal{G}_n$ are σ-subalgebras of $\mathcal{F}$, they are **independent** if $A_1, \ldots, A_n$ are independent for all $A_i \in \mathcal{G}_i$. Random variables $\varphi_1, \ldots, \varphi_n$ are called **independent** if $\sigma(\varphi_1), \ldots, \sigma(\varphi_n)$ are independent.

Definition 1.13. An infinite collection of events/σ-algebras/random variables is **independent** if all finite subcollections are independent.

Definition 1.14. $(\varphi_i)_i$ are **independent and identically distributed** (**iid**) if they are independent, all take values in the same X¡ and ${\varphi_i}_*\mathbb{P} = {\varphi_j}_*\mathbb{P}$ for all i, j.