Chapter 2: Some Mathematical Preliminaries

Notes | Exercise Solutions

Notes

- Here, we rigorously define some important concepts in probability theory.
- Notation note: $P(A)$ used for the probability of A, and $\mathcal{P}(A)$ used for the power set of A.
- **Def** (Probability space): The 3-tuple $(\Omega, \mathcal{F}, P)$, where
  1. Ω is a set containing elements which represents the sample space of outcomes
  2. $\mathcal{F}$ (σ-algebra/σ-field) is a set of subsets of Ω (that is, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$) that satisfies 3 conditions:
     1. It contains $\emptyset$, the empty set
     2. Closure under unions: If $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$
     3. Closure under complements wrt Ω: If $A \in \mathcal{F}$, then $\Omega \backslash A \in \mathcal{F}$
  3. $P : \mathcal{F} \to [0, 1]$ is a function that assigns values between 0 and 1 to elements of $\mathcal{F}$ (i.e: subsets of Ω) that satisfies:
     1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
     2. Countable additivity: For a countable collection of disjoint events $A_1, A_2, \cdots \in \mathcal{F}$, $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
  - **Def** (Complete probability space): Say that $(\Omega, \mathcal{F}, P)$ is a complete probability space if $\mathcal{F}$ contains all nullsets. What's a nullset? This builds up to it:
    - **Def** (Outer measure): An outer measure P^* is a function that maps all possible events in the sample space to the extended real numbers, $P^* : \mathcal{P}(\Omega) \to [0, \infty]$, which satisfies:
      1. $P^*(\emptyset) = 0$
      2. Countable additivity: For countable subsets $A, B_1, B_2, \cdots \subseteq \Omega$, if $A \subseteq \cup_{i=1}^{\infty} B_i$ then $P^*(A) \leq \sum_{i=1}^{\infty} P(B_i)$
    - Probability measures satisfy all the conditions of being an outer measure, except the domain is $\mathcal{F}$ instead of $\mathcal{P}(\Omega)$. But we can kind of fix this: Given a probability space $(\Omega, \mathcal{F}, P)$, define the P-outer measure as the mapping
      $P^*(A) := \inf\{P(F) \mid F \in \mathcal{F}, A \subset F\}$
      Our domain gets to be $\mathcal{P}(\Omega)$ now, because P^* is defined for all subsets of Ω, since Ω itself is in every σ-algebra.
    - **Def** (Nullset): Say that $A \subseteq \Omega$ is a nullset if it has P-outer measure 0; that is, $P^*(A) = 0$
  - **Def** (Measurability, of an event): Say that $F \subseteq \Omega$ is F-measurable if $F \in \mathcal{F}$. Then, call F an 'event' and say that F 'occurs with' probability $P(F)$.
  - Some remarks:
    - The smallest σ-algebra: What's the smallest set of subsets of Ω that satisfies the three properties above? If we don't specify anything else, the smallest possible $\mathcal{F} = \{\phi, \Omega\}$. The smallest σ-algebra containing A is called $\sigma(A)$ or $\mathcal{H}_A$, the σ-algebra generated by A

- The largest σ-algebra: Why not just use $P(\Omega)$ as F? Each element of F is essentially an 'event' that can be assigned a probability. We care about carving out subsets of Ω to use for F in our assignment of probabilities because we run into trouble when trying to define P in a nice way for certain Ω (it's not a problem for discrete ones). It might be too big.
- On P: A probability space is a special case of a measure space. It's a special case because of the $P(\Omega) = 1$ condition.

- Random variables:
  - **Def** (Measurability, of a function): Say that a function $Y : \Omega \to \mathbb{R}^n$ is $\mathcal{F}$-measurable if the pre-images of all open sets are $\mathcal{F}$-measurable, $Y^{-1}(U) = \{\omega \in \Omega \mid Y(\omega) \in U\} \in \mathcal{F}$, for all open sets $U \in \mathbb{R}^n$. Basically, the pre-image of Y by every open set in $\mathbb{R}$ is in $\mathcal{F}$.
  - **Def** (Random variable): Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. Then a random variable $X : \Omega \to \mathbb{R}^n$ is an $\mathcal{F}$-measurable function.
  - **Lemma** (Doob-Dynkin, special case): Let $X, Y : \Omega \to \mathbb{R}^n$, and let $\mathcal{H}_X$ be the σ-algebra generated by X (i.e: smallest σ-algebra containing X). Then Y is $\mathcal{H}_X$-measurable if and only if there exists a mapping $g : \mathbb{R}^n \to \mathbb{R}^n$ between X and Y s.t. $Y = g(x)$, and g is Borel-measurable. That is, all of the pre-images of g on open sets are Borel sets
  - **Def** (Probability distributions): We didn't need P to define what a random variable is, since it's just a mapping from Ω to the reals. But now we need P to measure probabilities, because no shit.
    Let X be a random variable. Then the distribution of X is the probability measure induced by X, $\mu_X(B) = P(X^{-1}(B))$ for all $B \in \mathbb{R}^n$
  - **Def** (Expectation): Using the 'weights' assigned to each event by X, we define the weighted average of all possible events as $E[X] := \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}^n} x d\mu_X(x)$

- L^p spaces
  - We'd like to to measure the norm of a random variable, and the distance between random variables. So we make them live in a function space, like an L^p-space
  - **Def** (Of a random variable X, its L^p-norm): Of any $x \in \mathbb{R}^n$, its L^p norm is $||x||_p = (x_1^p + \cdots + x_n^p)^{1/p}$ (for $p = 2$, this is the Euclidean norm). Of a random variable X on sample space Ω with probability measure P, for finite p it is $||X||_p = \int_\Omega |X(\omega)|^p dP(\omega))^{1/p}$ and for $p = \infty$ it is $||X||_\infty = ||X||_{L^\infty(P)} = \sup\{|X(\omega)| \mid \omega \in \Omega\}$
  - **Def** (L^p space, $L^p(P) = L^p(\Omega)$): Contains all random variables $X : \Omega \to \mathbb{R}^n$ with finite L^p norms, with distance metric induced by the L^p norm, $d(X, Y) = ||X - Y||_p$. Note that under this metric, L^p spaces are Banach spaces. In the special case of $p = 2$, it is a Hilbert space.

- **Def** (Independence):
  - Of events: Say that $A, B \in \mathcal{F}$ are independent if $P(A \cap B) = P(A) \cdot P(B)$
  - Of collections of events: Say that $A = \{A_i\}_{i \in I}$ are independent if for all pairs $i, j \in I, i \neq j$, A_i and A_j are independent
  - Of random variables: Say that $X, Y : \Omega \to \mathbb{R}$ are independent if $\mathcal{H}_X$ and $\mathcal{H}_Y$, the σ-algebras generated by X and Y, are independent

- **Def** (Stochastic process): On $(\Omega, \mathcal{F}, P)$, a collection of random variables $\{X_t\}_{t \in T}$ indexed by time.
  - Fix $t \in T$. Then we have a single random variable X_t with its associated probability measure.
    Fix a path $\omega \in \Omega$. Then we have a function that depends only on time, since there is no more randomness.
  - **Thm** (Kolmogorov, extension): Given a family of probability measures $\{\nu_{t_1}, \dots, \nu_{t_k} \mid k \in \mathbb{N}, t_i \in T\}$ on $\mathbb{R}^{nk}$, there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $\{X_t\}$ on Ω s.t.
    $\nu_{t_1,\dots,t_k}(F_1 \times \dots \times F_k) = P(X_{t_1} \in F_1, \dots, X_{t_k} \in F_k)$ for all $t_i \in T, k \in \mathbb{N}$, Borel sets F_i
    1. For all $t_1, \dots, t_k \in T$, $k \in \mathbb{N}$, $\nu_{t_{\sigma(1)},\dots,t_{\sigma(k)}}(F_1 \times \dots \times F_k) = \nu_{t_1,\dots,t_k}(F_{\sigma^{-1}(1)} \times \dots \times F_{\sigma^{-1}(k)})$ for all permutations σ of $\{1, \dots, k\}$
    2. $\nu_{t_1,\dots,t_k}(F_1 \times \dots \times F_k) = \nu_{t_1,\dots,t_k,t_{k+1},\dots,t_{k+m}}(F_1 \times \dots \times F_k \times \mathbb{R}^n \times \dots \times \mathbb{R}^n)$ for all $m \in \mathbb{N}$
  - **Thm** (Kolmogorov, continuity): A stochastic process with discrete time $X = \{X_t\}_{t \geq 0}$ has a continuous version if, for all $T > 0$, there exist α, β, D s.t. $E[|X_t - X_s|^\alpha] \leq D \cdot |t - s|^{1+\beta}$ for $0 \leq s, t \leq T$
- Very important instantiations
  - **Def** (Borel σ-algebra): Let Ω be a topological space, like $\mathbb{R}^n$. Let U be the collection of all open subsets of Ω. Then the Borel σ-algebra $\mathcal{B}$ is $\sigma(U)$, the smallest σ-algebra that contains U.
    The elements $B \in \mathcal{B}$ that are $\mathcal{B}$-measurable are called Borel sets.
  - **Def** (Brownian motion): This is the most important stochastic process.
    - Construction: We 'construct' this process indirectly, by first constructing a measure we like and then using Kolmogorov's extension theorem to say that it exists. For $x, y \in \mathbb{R}^n$ and $t > 0$, define $p(t, x, y) = (2\pi t)^{n/2} \cdot \exp(-\frac{|x-y|^2}{2t})$. We like the following measure:
      $\nu_{t_1,\dots,t_k}(F_1 \times \dots \times F_k) = \int_{F_1 \times \cdot \times F_k} p(t_1, x, x_1) p(t_2 - t_1, x_1, x_2) \dots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \dots dx_k$
    - This measure satisfies K's 2 conditions, so such a stochastic process in such a probability space exists. We call this process Brownian motion. It has the following properties, which follow from how we defined $v_{t_1,\dots,t_k}$ ...
    - Properties: Say that a collection of real-valued random variables $\{B_t\}_{t \geq 0}$ is a standard Brownian motion if it has the following properties:
      1. $B_0 = 0$
      2. Each sample path is continuous
      3. Stationary and normal increments: For any $t > s$, $B_t - B_s \sim N(0, t - s)$
      4. Independent increments: For all disjoint intervals, increments are independent

Exercise Solutions

Exercises completed: 2.1, 2.2, 2.3, 2.19

2.1. Let $X : \Omega \to \mathbb{R}$ assume only countably many values $a_1, a_2, \dots \in \mathbb{R}$.

(a) Show that X is a random variable if and only if $X^{-1}(a_k) \in \mathcal{F}$ for all $k = 1, 2, \dots$ (2.2.16)

We know that X is a random variable iff X is $\mathcal{F}$-measurable. That is, pre-images of all open sets are measurable, so $X^{-1}(U) = \{\omega \in \Omega \mid X(\omega) \in U\} \in \mathcal{F}$ for all open $U \in \mathbb{R}$.
$\implies$: Let X be a random variable. Then by definition, X's pre-images of all open sets are all contained in $\mathcal{F}$. Since $\mathcal{F}$ is closed under complements (and we know that pre-images of complements = complements of pre-images from set theory), pre-images of all closed sets are also in $\mathcal{F}$. The singleton set $\{a_k\} \subset \mathbb{R}$ is closed, so its pre-image by X is $\mathcal{F}$-measurable.
$\impliedby$: Let $X^{-1}(a_k) \in \mathcal{F}$ for all k. Since $\mathcal{F}$ is closed under infinite unions (and unions of pre-images = pre-images of unions), we also have that $X^{-1}(\cup a_i) \in \mathcal{F}$ for any union of a_i's. Let U be open in $\mathbb{R}$. Then $X^{-1}(U) = X^{-1}(U \cap X(\Omega))$, since the pre-image of a value that X does not take on is empty. Since $X(\Omega)$ is countable, so is $U \cap X(\Omega)$. So it is a union of a_i's, which means that it is measurable.

(b) Suppose (a) holds. Show that $E(|X|) = \sum_{k=1}^{\infty} |a_k| P(X = a_k)$

$|X|$ only takes on countably-many values $|a_1|, |a_2|, \dots$, so we use the indicator function trick from calculus that allowed us to integrate over planar regions!: $E(|X|) = \int_\Omega X(\omega) dP(\omega) = \int_\Omega |X(\omega)| \sum_{k=1}^{\infty} \chi(\omega) dP(\omega)$, where $\chi(\omega) = 1$ if $\omega \in X^{-1}(a_k)$, and 0 otherwise. Then we have:
$E(|X|) = \sum_{k=1}^{\infty} \int_\Omega |X(\omega)| \chi(\omega) dP(\omega)$, by Fubini's theorem
$= \sum_{k=1}^{\infty} \int_\Omega |a_k| \chi(\omega) dP(\omega)$, since $X(\omega)$ can only take on the values a_k
$= \sum_{k=1}^{\infty} |a_k| \int_\Omega \chi(\omega) dP(\omega) = \sum_{k=1}^{\infty} |a_k| P(X = a_k)$, since $\int_\Omega \chi(\omega) dP(\omega)$ precisely only takes on the value $P(X = a_k)$ when $\omega = a_k$

(c) If (a) holds and $E(|X|) < \infty$, show that $E(X) = \sum_{k=1}^{\infty} a_k P(X = a_k)$

Partition the indices of a_k: Let $\{i\}_{i \in I}$ be the indices for which $a_i \geq 0$, and let $\{j\}_{j \in J}$ be the indices for which $a_i < 0$. So by (b), $E(|X_1|) = \sum_{i \in I} |a_i| P(X = a_i)$ and $E(|X_2|) = \sum_{j \in J} |a_j| P(X = a_j)$ where X_1 is X restricted to its non-negative-valued domain, and X_2 is X restricted to its negative-valued domain. By assumption, the infinite series absolutely converges, so we can rearrange it to yield $E(X) = E(X_1) - E(X_2) = \sum_{k=1}^{\infty} a_k P(X = a_k)$

(d) If (a) holds and $f : \mathbb{R} \to \mathbb{R}$ is measurable and bounded, show that $E(f(X)) = \sum_{k=1}^{\infty} f(a_k) P(X = a_k)$

$f(X(\Omega))$ is the image of countably-many values $X(\Omega)$, so it also contains at most countably-many values b_k (by definition of a function lol). Suppose this is infinite (finite case follows). So f restricted to the image of X is a random variable, and its expected value is given by part (c) as
$E(f(x)) = \sum_{k}^{\infty} b_k P(f(x) = b_k) = \sum_{k=1}^{\infty} f(a_k) P(X = a_k)$

2.2. Let $X : \Omega \to \mathbb{R}$ be a random variable. The distribution function F of X is defined by $F(x) = P(X \leq x)$.

(a) Prove that F has the following properties:

$0 \leq F \leq 1$, $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$: $0 \leq F \leq 1$ since a probability measure is a function mapping to $[0, 1]$ by definition. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$ follow from monotone convergence

F is non-decreasing: We want to show that $x_1 > x_2 \implies F(x_1) \geq F(x_2)$. That is, $P(X \leq x_1) \geq P(X \leq x_2)$. Since $x_1 > x_2$, $x_1 = x_2 + \epsilon$ for some $\epsilon > 0$. Then
$P(X \leq x_1) = P(X \leq x_2 + \epsilon) = P(X \leq x_2) + P(x_2 \leq X \leq x_2 + \epsilon)$. Then by the nonnegativity of probability measures, this is greater than or equal to $P(X \leq x_2)$

F is right-continuous: We want to show that $F(x) = \lim_{h \to 0} F(x + h)$ for $h > 0$. Let $A_1 \supset A_2 \supset \ldots$ be countable nested subsets of $\mathbb{R}$. Then $P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n)$ by σ-additivity in Definition 2.1.1 (b). We can take $A_n = \{X \leq x + 1/n\}$. Then $P(\lim_{n \to \infty} A_n) = P(X \leq x) = F(x)$

(b) Let $g : \mathbb{R} \to \mathbb{R}$ be measurable s.t. $E(|g(X)|) < \infty$. Prove that $E(g(X)) = \int_{-\infty}^{\infty} g(x) dF(x)$, where the integral on the right is interpreted in the Lebesgue-Stieltjes sense.

By definition, $E(g(X)) = \int_\Omega f(X(\omega)) dP(\omega) = \int_\mathbb{R} g(x) d\mu_X(x)$, where μ_X is the induced probability measure $\mu_X(B) = P(X^{-1}(B))$. Then $\mu_X(\{X \leq x\}) = P(X^{-1}(X \leq x)) = P(X \leq x) = F(x)$, so $E(g(x)) = \int_\mathbb{R} g(x) dF(X)$

(c) Let $p(x) \geq 0$ be a measurable function on $\mathbb{R}$. Say that X has density p if $F(x) = \int_{-\infty}^{x} p(y) dy$ for all x. Then from (2.2.1)–(2.2.2) we know that 1-dimensional Brownian motion B_t with $B_0 = 0$ has density $p(x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t})$ for $x \in \mathbb{R}$. What is the density of B_t^2?

Let F be the distribution function of B_t^2. Then by definition, the desired density $q(y)$ must satisfy $F(x) = P(B_t^2 \leq x) = \int_{-\infty}^{x} q(y) dy$. We can express this in terms of B_t: $P(B_t^2 \leq x) = P(B_t < \sqrt{x}) = \int_{-\infty}^{\sqrt{x}} p(y) dy$
So we can differentiable both sides to get that $q(x) = p(\sqrt{x}) \cdot \frac{d}{dx} \sqrt{x} = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}) \cdot \frac{1}{2\sqrt{x}}$

.

2.3. Let $\{\mathcal{H}_i\}_{i \in I}$ be a family of σ-algebras on Ω. Prove that $\mathcal{H} = \cap \{\mathcal{H}_i \mid i \in I\}$ is again a σ-algebra.

We'll do this by confirming that $\mathcal{H}$ satisfies our 3 properties of σ-algebras:

1. Contains $\emptyset$: Since each $\mathcal{H}_i$ contains $\emptyset$, so does their intersection $\mathcal{H}$.
2. Closure under complements: Let $A \in \mathcal{H}$. Then $A \in \mathcal{H}_i$ for all $i \in I$, and by closure under complements for each $\mathcal{H}_i$ we have that $A^c \in \mathcal{H}_i$. So A^c is in their intersection $\mathcal{H}$.
3. Closure under unions: Let $A_i, A_2, \dots \in \mathcal{H}$. Then $A_1, A_2, \dots \in \mathcal{H}_i$ for each $i \in I$, and their union is in each $\mathcal{H}_i$ by closure under unions of each $\mathcal{H}_i$. So this union is in their intersection $\mathcal{H}$.

2.19: Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and let $p \in [1, \infty]$. A sequence $\{f_n\}_{n=1}^{\infty}$ of functions $f_n \in L^p(\mu)$ is called a Cauchy sequence if $||f_n - f_m||_p \to 0$ as $n, m \to \infty$. The sequence is called convergent if there exists $f \in L^p(\mu)$ such that $f_n \to f$ in $L^p(\mu)$. Prove that every convergent sequence is a Cauchy sequence.

We are confirming that every convergent sequence in a metric space is Cauchy, in the special case where our metric space is $L^p(\mu)$. This is a simple application of the triangle inequality: Let $\{f_n\}_{n=1}^{\infty} = f_1, f_2, \ldots$ be a sequence converging to f, so $f_n \to f$ in $L^p(\mu)$. By the triangle inequality, we have that
$||f_n - f_m||_p \leq ||f_n - f||_p + ||f_m - f||_p$
By convergence, there exist sufficiently large n and m s.t. $||f_n - f||_p \to 0$ and $||f_m - f||_p \to 0$. So $||f_n - f_m||_p \to 0$, and $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence.