Project C Project

Charlie Windolf

Contents

Chapter 1

Probability Theory

Fix a sample space S and for all $A \subset S$ define $A^C = S \setminus A$.

1.1 Set Theory

Theorem 1.1.1 (Set arithmetic). *Set intersection $\cup$ and union $\cap$ are commutative, associative, and distributive, i.e.*

- (*Commutativity*)

$$\begin{aligned} A \cup B &= B \cup A, \\ A \cap B &= B \cap A. \end{aligned}$$

- (*Associativity*)

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C, \\ A \cap (B \cap C) &= (A \cap B) \cap C, \end{aligned}$$

- (*Distributivity*)

$$\begin{aligned} A \cup (B \cap C) &= (A \cup B) \cap (A \cup C), \\ A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), \end{aligned}$$

- (*De Morgan*)

$$\begin{aligned} (A \cup B)^C &= A^C \cap B^C, \\ (A \cap B)^C &= A^C \cup B^C. \end{aligned}$$

De Morgan's laws extend to arbitrary unions and intersections. Actually, by definition of set membership, all of the above also extend to arbitrary unions and intersections.

Definition 1.1.1 (Disjoint events, partition). $A_1, A_2, \ldots$ are *pairwise disjoint* if for all $i \neq j$, $A_i \cap A_j = \varnothing$. If $\{A_i\}_{i \in I}$ are pairwise disjoint and $\bigcup_{i \in I} A_i = S$, then $\{A_i\}$ are said to form a *partition* of S.

1.2 Basics of Probability Theory

Definition 1.2.1 (σ-algebra)**.** A collection $\mathcal{B}$ of subsets of S is called a σ-*algebra* if

a. $\varnothing \in \mathcal{B}$,

b. Closed under complement: $A \in \mathcal{B} \implies A^C \in \mathcal{B}$,

c. Closed under countable union: $A_1, A_2, \cdots \in \mathcal{B} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

Definition 1.2.2 (Borel algebra)**.** The Borel algebra is the smallest σ-algebra containing all open sets. We'll use $\mathcal{B}_{\mathbb{R}}$ to denote the Borel algebra on the reals.

Note that (a.,b.) $\implies S = \varnothing^C \in \mathcal{B}$, and by De Morgan and (c.), $\mathcal{B}$ is also closed under countable intersections. A set S together with a σ-algebra $\mathcal{B}$ is called a *measurable space*.

Definition 1.2.3 (Probability measure)**.** Given a measurable space $(S, \mathcal{B})$, a *probability measure* is a set function $P : \mathcal{B} \to \mathbb{R}$ such that

1. $P(A) \geq 0$ for all A,
2. $P(S) = 1$,
3. For $A_1, A_2, ...$ pairwise disjoint, $P(\bigcup_1^{\infty} A_i) = \sum_1^{\infty} P(A_i)$.

Axiom of Continuity. If $A_1 \supset A_2 \supset ...$ and $\bigcap_n A_n = \varnothing$, then $P(A_n) \to 0$.

As shown in an exercise, continuity combined with the finite union property imply the countable union property.

Proposition 1.2.1 (Basic properties of probability measures)**.** Let $(S, \mathcal{B}, P)$ be a probability space and A a measurable set. Then,

a. $P(\varnothing) = 0$,

b. $P(A) \leq 1$,

c. $P(A^C) = 1 - P(A)$,

d. $P(B \cap A^C) = P(B) - P(B \cap A)$,

e. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, which weakens to P(A∩B)≥P(A)+P(B)-1 (Bonferroni),

f. $A \subset B \implies P(A) \leq P(B)$.

Further,

g. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $\{C_i\}$.

h. $P(\bigcup_1^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ for any $\{A_i\} \subset \mathcal{B}$ (Boole).

Note (On counting)*.*

- Recall that the number of independent choices from m collections of sizes $n_1, \dots, n_m$ is $n_1 \times \cdots \times n_m$.

- How many ways are there to sample k objects from a collection of size n? Well, it depends on the sampling method:
  - *Ordered, with replacement*: n^k.
  - *Ordered, without replacement*:

$$\frac{n!}{(n-k)!} = n \times (n-1) \times \cdots \times (n-k+1).$$

  - *Unordered, without replacement*:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n}{k} \times \frac{n-1}{k-1} \times \cdots \times \frac{n-k+1}{1}.$$

  - *Unordered, with replacement*, aka *stars and bars*: since we're putting k balls into n bins, we are equivalently arranging $n-1$ bars and k stars. So, there are $k+n-1$ total slots, and we want to choose k of them for the stars, putting us in the previous case:

$$\frac{(k+n-1)!}{k!(n-1)!} = \binom{k+n-1}{k}.$$

1.3 Conditional Probability and Independence

Definition 1.3.1 (Conditional probability)**.** Let A, B be measurable with $P(B) > 0$. Then the *conditional probability* $P(A \mid B)$ is given by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Theorem 1.3.1 (Bayes' rule)**.** *For B measurable and $A_1, A_2, \ldots$ a partition of the sample space, we have*

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j)P(A_j)},$$

and in particular for any measurable A,

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

To obtain the latter, take $A = \bigcup_{i=1}^{\infty} A_i$.

Proposition 1.3.1 (Chain rule)**.** For any events $A_1, \ldots, A_n$,

$$\begin{aligned} P(A_n \cap \cdots \cap A_1) &= P(A_n \mid A_{n-1} \cap \cdots \cap A_1) \times \cdots \times P(A_2 \mid A_1)P(A_1) \\ &= \prod_{k=1}^{n} P\Big(A_k \mid \bigcap_{j=1}^{k-1} A_j\Big). \end{aligned}$$

This follows by straight induction from $P(A \cap B) = P(B \mid A)P(A)$.

Definition 1.3.2 (Independence)**.** Two events A, B are said to be *independent*, written $A \perp\!\!\!\perp B$, if

$$P(A \cap B) = P(A)P(B).$$

Note: If $A \perp\!\!\!\perp B$, then also $A \perp\!\!\!\perp B^C$, $A^C \perp\!\!\!\perp B$, $A^C \perp\!\!\!\perp B^C$.

Definition 1.3.3 (Mutual independence)**.** A collection of events $\{A_i\}_{i \in I}$ is said to be mutually independent if for all $J \subset I$,

$$P\Big(\bigcap_{j \in J} A_j\Big) = \prod_{j \in J} P(A_j).$$

1.4 Random Variables

Definition 1.4.1 (Measurable function)**.** Let $(S, \mathcal{A})$, $(T, \mathcal{B})$ be measurable spaces and $f : A \to B$. Then we say that f is measurable if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$, i.e. preimages of measurable sets are measurable.

Note: The book ignores the measurability of random variables, but I wanted to write it down.

Definition 1.4.2 (Random variable)**.** A random variable is a measurable function $X : (S, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Definition 1.4.3 (Pushforward, induced probability measure)**.** Let X be a random variable with range $\mathcal{X} = X(S)$. Then X induces a *pushforward* probability measure on $\mathcal{X}$, $P_X = P \circ X^{-1}$, i.e.

$$P_X(E) = P(X^{-1}(E)).$$

This will most often be written as $P(X \in E)$.

1.5 Distribution Functions

Definition 1.5.1 (Cumulative distribution function)**.** The *cumulative distribution function* or *cdf* of a random variable X is

$$F_X(x) = P(X \leq x).$$

Note: In the book, they have (to me) some problems with notation where $F_X(x) = P_X(X \leq x)$. This does not make sense to me, since $X \leq x$ lives in the sample space and is not a subset of $\mathbb{R}$ under any reasonable interpretation.

Definition 1.5.2 (cádlág)**.** Continuous from the right, with limits existing from the left.

Theorem 1.5.1 (Properties of the .)**.** *$F(x)$ is a cdf if and only if the following hold* *a.* $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$,

b. $F(x)$ *is nondecreasing,*

c. $F(x)$ *is right continuous.*

Note that (b.) and (c.) imply cádlág. Actually that's not really important at this point?

Note: We say that a random variable with a continuous cdf is *continuous* and that one with a step function cdf is *discrete*.

Definition 1.5.3 (Identically distributed)**.** Random variables X, Y are *identically distributed* if $P_X = P_Y$, i.e. if for all $A \in \mathcal{B}_{\mathbb{R}}$, $P(X \in A) = P(Y \in A)$.

Definition 1.5.4 (pdf, pmf)**.** The definition they give is vague. But the point is these are positive functions that sum to 1 over the range of the random variable.

1.6 Important examples

- Binomial pmf: If $X \sim \mathrm{Bin}(n, p)$, then

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Chapter 2

Transformations and Expectations

2.1 Distributions of Functions of a Random Variable