Honors Single Variable Calculus 110.113

October 24, 2023

Contents

Project homework requires working with new concepts that build upon our lecture material. The problems are not so hard.

0.1 Introduction

Reading: [1], this is freely available here.

> ... *in mathematics you don't understand things. You just get used to them* - von Neumann

Learning Objectives

A recurring theme that you would see throughout your study of more "theoretical" sciences is

- Making *good* definitions.
- *Working* with definitions

This project aims to familiarize you with the foundations of probability theory as set up by A. Komolgorov. In pure mathematics and its applications, it is desirable to have a foundation where one can discuss non deterministic statements, which we will refer as *events*, and non deterministic values, which are *random variables*. The project will proceed in the following order:

1. Probability space, 1.
2. Modeling (statistical)
3. You will then have few choices to explore:
   - Probability: we will explore foundational results as the strong law of large numbers.
   - Statistical: we will explore applications in inference and language models.
   - Social choice theory.

Current status of the available content:

- 1. is available. The problems there are compulsory.
- 2 is not available. Problems are compulsory.
- 3 is not available. You will only have to pick one option depending on your taste.

1 Defining a probability space following A. Kolmogorov

Reading:

Definition 1.1. A *measure space* consists of a pair $(\Omega, \mathcal{E})$ where Ω is a set, and $\mathcal{E}$ is a σ-algebra on Ω.

- elements $E \in \mathcal{E}$ are referred as *events*, *events space* or *measurable sets.*

Definition 1.2. Let $(\Omega, \mathcal{E})$ be a measure space. A *(finite) probability measure* is a map $p : \mathcal{E} \to \mathbb{R}_{\geq 0}$ satisfying

1. $p(\Omega) = 1$

2. Finitely additivity. Let $\{A_i\}_{i\in I}$ be a finite (that is $|I| = n$ for some $n \in \mathbb{N}$) collection of disjoint elements in $\mathcal{E}$ [1]. Then

$$p\left(\bigcup_{i=0}^{N} A_i\right) = \sum_{i=0}^{N} p(A_i)$$

Once we have learnt the definition of series, we will add in another axiom called *countable additivity.*

Definition 1.3. A *probability space* is the datum of $(\Omega, \mathcal{E}, p)$, where p is a probability measure.

The discrete case. Let Ω be a finite set. $\mathcal{E} := 2^{\Omega}$ is the set of all subsets of Ω. This is a σ-algebra. Let p_w be any finite collection of real numbers such that

$$\sum_{w\in\Omega} p_w = 1$$

Proposition 1.4. There is a map

$$p : 2^{\Omega} \to [0, 1]$$

uniquely extending the condition

$$p(\{w\}) = p_w \quad w \in \Omega$$

Proof. Exercise. □

Due to prop. 1.4 we define the following:

Definition 1.5. Let Ω be a finite set. A *probability mass function* on Ω is a map

$$p : \Omega \to [0, 1]$$

satisfying

$$\sum_{w\in\Omega} p(w) = 1$$

we denote $p_w := p(w)$

[1] Remember, these are subsets of 2^{Ω}.

1.1 Problems

Example

Modeling n tosses of a fair coin. We define $(\Omega_n, \mathcal{E}, p)$.

- Ω_n is the set of all n consecutive ordered sets of letters which are either H or T.[a]
- $\mathcal{E}$ is the set of all subsets of Ω_n. One event can be

$$E_{\geq k} := \{\omega \in \Omega_n \, : \text{ at least } k \text{ heads appear in the } n \text{ tosses}\}$$

  This is the set of all sequences with at least k Hs.
- Set $p(\{\omega\}) = \frac{1}{2^n}$ for all singleton subsets $\{\omega\} \in \mathcal{E}$ where $\omega \in \Omega_n$. This uniquely extends to a function (why?)

$$p : \mathcal{E} \to \mathbb{R}_{\geq 0}$$

[a] Of course, from our language of set theory, this is not a valid set. But we can equally use 0 or 1 to model this, in this case, this follows from the axioms.

The following problems are related to the model described above on n-tosses of a fair coin.

1. (a) List out the elements in the events, def 1.1, of

$$\Omega_n$$

   for $n = 1, 2$ and 3. Prove Ω_n has 2^n elements for $n \in \mathbb{N}_{\geq 1}$[2].

   (b) Consider the probability space $(\Omega_3, \mathcal{E}, p)$ ($n = 3$ in example). List out the elements of $E_{\geq i}$ for $i = 1, 2, 3$.

2. For a $n \in \mathbb{N}_{\geq 1}$. Consider the events $E_{\geq i}$ described in example of the probability space $(\Omega_n, \mathcal{E}, p)$. Give a formula for

$$p(E_{\geq i})$$

   for $0 \leq i \leq n$.

3. Consider now the probability space $(\Omega_{2n}, \mathcal{E}, p)$. How many elements are in the event

$$E := \{\text{exactly } n \text{ heads appear}\}$$

[2] This will be a shorthand for a positive integer.

Prove that

$$p(E) = \frac{1}{2^{2n}} \binom{2n}{n}$$

[3]

[3] One can apply *Stirling's* formula to show that this is $\sim \frac{1}{\sqrt{\pi n}}$ as $n \to \infty$.

2 Conditional Expectation

Let us consider the discrete case for warm-up. Once we have learned integration, we will repeat the same story for density functions. The definition below is often referred as Baye's rule. Fix a probability space $(\Omega, \mathcal{E}, p)$.

Definition 2.1. Let $A, B \in \mathcal{E}$. The *conditional probability of A given B*

$$p(A|B) := \frac{p(A \cap B)}{p(B)}$$

provided $p(B) > 0$.

This is what often leads to a formulation of Baye's rule. One of the earliest applications is in the field of *Bayesian inference*, and was used in text classification by Mosteller and Wallace (1964), see [2, 4].

2.1 Problems

Definition 2.2. A *partition of a X* is a collection of subsets X_i, indexed by a set $i \in I$ such that

1. $\bigcup_{i \in I} X_i = X$
2. The sets X_is are pairwise disjoint: for any $i, j \in I$, the intersection (Def. **??**) of X_i and X_j is empty, $X_i \cap X_j = \emptyset$.

We will now work on this definition by proving some important results.

1. (**) Let I be a finite set. Let $\{B_1, B_2, \ldots\}_{i \in I}$ be a finite partition, 2.2, of Ω and $p(B_i) > 0$ for all $i \in I$. Prove that

   $$p(A) = \sum_{i \in I} p(A|B_i) p(B_i)$$

   using the additivity axiom.

2. (**) By conditioning on something, we would expect that we get a *new* probability space. If $B \in \mathcal{E}$ such that $p(B) > 0$ show that $q : \mathcal{E} \to \mathbb{R}$ given by $q(A) := p(A|B)$ defines a probability space $(\Omega, \mathcal{E}, q)$.

References

[1] Geoffrey Grimmet and Dominic Welsh, *Probability: an introduction*, Oxford, 2014.

[2] Dan Jurafsky and James H. Martin, *Speech and language processing (3rd ed. draft)*, 2023.