Honors Single Variable Calculus 110.113

December 15, 2023

Contents

1 Project Homework

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

1.1 Introduction

Reading: Grimmet and Welsh's Probability: An Introduction. [1], this is freely available here. 90% of the material will be from this book. I will occasionally steal some content Tao's lecture notes, freely available online on his blog, [4].

> *... 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, 2, and how this captures conditional probability, 3. This is the content of Project 1.
2. Theory of distributions, Sec. 4. This is the beginning of Project 2.
3. Central limit theorem, 6.

1.2 What to hand in

Due December 22nd. This will be 50% of your final grade. The score distribution would be as follows:

1. From Sec 4 to Sec. 12, there are various exercise problems. Select 5. (50%). These are at 5.1.1, 9.0.1, 10.0.1, 11.0.1, .
2. A summary of the central limit theorem, with three real-life examples, this is 1. and 2. of Subsec. 6.0.1. (50%).

2 Defining a probability space following A. Kolmogorov

Reading: Grimmet and Welsh's Probability: an introduction. [1], this is freely available here.

Definition 2.1. 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*.

In most of our set ups, $\mathcal{E}$ is really chosen to be 2^{Ω}.

Definition 2.2. 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 2.3. A *probability space* is the datum of $(\Omega, \mathcal{E}, p)$, where p is a probability measure.

Example

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

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

Thanks to 2.4, p extends to a map $p : \mathcal{E} \to \mathbb{R}_{\geq 0}$. One can show that $(\Omega, \mathcal{E}, p)$ is a probability space, i.e. p satisfies the axiom of def. 2.3.

Proposition 2.4. There is a map

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

uniquely extending the condition

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

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

Proof. Exercise. □

Due to prop. 2.4 we define the following:

Definition 2.5. 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 $p_w := p(w)$

2.1 Problems: modeling of n-coin tosses

Example

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

- Ω_n is the set of all n consecutive ordered sets of letters which are either H or T.[a]
- $\mathcal{E}_n$ 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 $\{\omega\} \in \mathcal{E}_n$ where $\omega \in \Omega_n$. This uniquely extends to a function (why?)

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

[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 2.1, of

   $$\Omega_n$$

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

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

(b) Consider the probability space $(\Omega_3, \mathcal{E}_3, p)$ ($n = 3$ 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 $(\Omega_n, \mathcal{E}_n, p)$. Give a formula for
$$p(E_{\geq i})$$
for $0 \leq i \leq n$.

3. Consider now the probability space $(\Omega_{2n}, \mathcal{E}_{2n}, p)$. How many elements are in the event
$$E := \{\text{exactly } n \text{ heads appear}\}$$
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 $n \to \infty$.

3 Conditional probability

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 3.1. 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].

3.1 Problems

Definition 3.2. 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, 3.2, 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 $(\Omega, \mathcal{E}, q)$.

4 Constructing random variables

Learning Objectives

From random variables we will introduce distributions. There are two types:

- discrete distributions.
- continuous distributions.

A random variable is a quantity which depends on random events. [4] It is a function on the possible outcomes.

Example

Suppose we are in the case of modeling 1 coin toss, 2.1, $(\Omega_1, \mathcal{E}_1, p)$. Then $\Omega_1 = \{H, T\}$ has 2^1 elements. Suppose we want to encode the idea:

- "I go to party if I flip a heads".
- "I don't go to party if I flip tails."

We can describe this using a random variable.

- "go to party" by number 1 and 0 otherwise.
- The idea that "go to party if heads and not if tails" is encoded

$$X : \Omega_1 \to \mathbb{R}$$

$$X(H) = 1, \quad X(T) = 0$$

We may ask "What is the probability that I go to party? " This is:

$$p(X = 1) := p(X^{-1}(1)) = p(H) = \frac{1}{2}$$

Definition 4.1. Let $(\Omega, \mathcal{E})$ and $(\Omega', \mathcal{E}')$ be two measurable spaces. A *measurable map* between the measurable spaces is

1. A map
$$f : \Omega \to \Omega'$$
2. For all $E' \in \mathcal{E}'$, $f^{-1}(E') \in \mathcal{E}$.

[4] It is neither *random* nor a *variable*

We will fix a probability space $(\Omega, \mathcal{E}, p)$ for discourse.

Definition 4.2. A *random variable* is a measurable map[5], def 4.1,

$$X : (\Omega, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

where $\mathbb{R}$ is made a measurable space with the Borel σ-algebra.

1. It is *discrete*, if its *image*, $X(\Omega)$, is *countable* or *finite*.
2. *Continuous*, if it satisfies the property 4.8.

We denote the collection of discrete random variables on $(\Omega, \mathcal{E}, p)$ as $\mathrm{RV}^{\mathrm{disc}}(\Omega, \mathcal{E}, p)$.

A few practical remarks:

- The condition that the map is measurable is:

  $$\text{for all } E \in \mathcal{B}(\mathbb{R}),\ X^{-1}(E) \in \mathcal{E}$$

  This is there to make sense of notions as "what is the probability of the random variable X taking values in the range E?" Mathematically, this is written as

  $$p(X \in \mathcal{E}) := p(X^{-1}(E))$$

4.1 Probability mass function of discrete random variable

However, what is most interesting about a random variable?

- The values it takes.
- The probabilities it associates.
  - For discrete rv, we associate pmf, def. 2.5.
  - For continuous rv, we associate a cdf, or *distribution*, 4.3.

  Note that in the literature, some may conflate the two terms together, as in the case of the pmfs we discuss below, def. 4.4, def. 4.5.

Definition 4.3. The *probability mass function (pmf)* of a discrete random variable X is a function

$$p_X : \mathbb{R} \to [0, 1]$$

such that

$$p_X(x) := p(X^{-1}(x))$$

[5] *Maps* and *functions* are used interchangeably.

4.2 Three examples of pmf

Definition 4.4. Bernoulli distribution. Let X be a random variable. [6]

- It takes the value 0 and 1.
- $p_X(1) = p$ and $p_X(0) = q := 1 - p$.

In this case we write $X \sim \mathrm{Ber}p$.

Definition 4.5. X has *Poisson* distribution with parameter $\lambda > 0$.

$$p(X = k) = \frac{1}{k!}\lambda^k e^{-\lambda} \quad k \in \mathbb{N}_{\geq 0}$$

In this case we write $X \sim \mathrm{Poi}(\lambda)$

As mentioned earlier, we do not mention the model. By definition, this should be a pmf but we can also check

$$\sum_{k=0}^{\infty} \frac{1}{k!}\lambda^k e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty} \frac{1}{k!}\lambda^k = e^{-\lambda}e^{\lambda} = 1$$

Here are some common models that realize this random variables.

Example

World cup soccer match.

- Ω the set of data of each world cup soccer match.
- $X : \Omega \to \mathbb{R}$, takes a match data, and out put the number of goals scored. e.g (fake) $X(\text{TW vs USA }) = 50$.
- The average event rate is 2.5 goals

We can thus use this random variable to model

$$p(\text{k goals in a match}) = \frac{2.5^k e^{-2.5}}{k!}$$

]

Definition 4.6. Say that X has the *binomial distribution* with parameters n, p if

- X takes values in $\{0, 1 \ldots, n\}$.
- $P(X = k) = \binom{n}{k} p^k q^{n-k}$ for $k = 0, 1, \ldots, n$

To model such a random variable, we can consider the case of n-coin tosses where heads is given probability p. Ω_n has 2^n elements. $\mathcal{E}_n = 2^{\Omega_n}$ and that $p(\{\omega\}) = p^{h(\omega)}q^{n-h(\omega)}$. where $h(\omega)$ is the number of heads appearing in the sequence ω.

[6] This implies it is defined on some sample space.

4.3 Continuous random variables

Learning Objectives

- Discuss the normal distribution.

The study of random variables is often through their distribution functions.

Definition 4.7. Let $(\Omega, \mathcal{E}, p)$ be a measure space. If $X \in \mathrm{RV}(\Omega, \mathcal{E}, p)$, the *distribution of X* is

$$F_X : \mathbb{R} \to [0,1]$$
$$F_X(x) := p(X \leq x)$$

Definition 4.8. A random variable X is *continuous* if its distribution may be written

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du \quad x \in \mathbb{R}$$

for some nonnegative function $f_X : \mathbb{R} \to \mathbb{R}$. We say X has *probability density function (pdf)* f_X[7]. Here the integral means the *improper integral* of f_X, **??**.

Alternatively, the distribution can be encoded as *pushforward measure* of p, see 4.10.
What values can F_X tell us? It tell us the probability of observalues in some interval. For instance, for $(a, b]$,

$$p(a < X \leq b) = p(\{X \leq b\} \setminus \{X \leq a\}) = p(X \leq b) - p(X \leq a) = F_X(b) - F_X(a)$$

In this sense, the distribution of a function can recover the probability mass function, 2.5.

Definition 4.9. The *uniform distribution on* $[a, b]$. Let $a, b \in \mathbb{R}$, where $a < b$. Then define

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b \end{cases}$$

4.3.1 Problems

1. Draw a sketch of the uniform distribution, 4.9. Give one choice of density function of F, and sketch it.

[7] A prior this also mean that f_X is a function for which the integral exists.

4.4 The law of probability distribution

Definition 4.10. As set up in 4.7. Distribution can alternatively be regarded as

$$X_* p : B(\mathbb{R}) \to \mathbb{R}$$

where $B(\mathbb{R})$ is the natural *Borel* σ-algebra on $\mathbb{R}$, see **??**. This is referred as the *law of* X.

4.5 Characterization of functions which are distributions

You may omit this section on first read. Can we characterize functions F which are distributions? Yes, see Thm 4.11. Often, we do not care about the model of the random variable X:

- We only care about the probabilities assigned in the distribution function.
- It can be too hard to describe the models.

So it boils down to the study of certain functions :

Theorem 4.11. F_X is the distribution of a random variable $X \in \mathrm{RV}(\Omega, \mathcal{E}, p)$ iff F_X is a function satisfying the three axioms:

- $F_X(x) \le F_X(y)$ if $x \le y$.
- $F_X(x) \to 0$ as $x - \infty$, and $F_X(x) \to 1$ as $x \to \infty$.
- F_X is right continuous: for all $s \in \mathbb{R}$, $\lim_{t \to s^+} F_X(t) = F(s)$.

Note that the converse says that we can construct a probability space with a random variable such that F_X is the distribution associated to X.

Proof. ($\Rightarrow$) Suppose F is a distribution.

1. If $x \le y$, then $\{X \le x\} \subseteq \{Y \le y\}$. Then by axiom of measure, p, $F_X(x) = p(X \le x) \le p(X \le y) = F_X(y)$
2. These two results are saying

$$p(X \le -\infty) = 0 \quad p(X \le +\infty) = 1$$

   These follow from the continuity of p.

($\Leftarrow$) We omit this for now and leave as a challenge. $\square$

Note: distribution is not necessarily continuous.

4.5.1 Problems

Omitted.

5 What can we associate to random variables?

Learning Objectives

- Learn how to compute expectation in the the discrete and continuous case.

All forms of expectation for X a RV on probability space $(\Omega, \mathcal{E}, p)$ can be expressed as

$$\mathbb{E}[X] := \int_{\Omega} X \, dp$$

as we have not discussed integration in this context, we will do special cases.

Definition 5.1. The expectation for a discrete random variable, **??**. Let $X \in \mathrm{RV}^{\mathrm{disc}}\Omega, \mathcal{E}, p)$ [8] many outcomes. Then

$$\mathbb{E}[X] := \sum_{i=1}^{\infty} x_i p(X = x_i)$$

if the sum absolutely [9]converges, def. **??**.

Definition 5.2. Let $X \in \mathrm{RV}^{\mathrm{cts}}(\Omega, \mathcal{E}, p)$. The *expectation* of X is given by

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f_X(x) \, dx$$

whenever the integral absolutely converges, i.e. $\int_{-\infty}^{\infty} |x f_X(x)| \, dx$ exists.

The next quantity, whose definition is the same for both dicrete and continuous.

Definition 5.3. Let X be a discrete or continuous random variable. Then

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}(X))^2)]$$

This measures the degree of dispersion of a random variable *around its mean.*

5.1 Properties of expectation

Proposition 5.4. Law of subconscious statistician. If $X \in \mathrm{RV}^{\mathrm{disc}}(\Omega, \mathcal{E}, p)$, and $g : \mathbb{R} \to \mathbb{R}$ is any function , then

$$\mathbb{E}(g(X)) = \sum_{x \in \mathrm{im} X} g(x) p(X = x)$$

whenever this sum converges absolutely.

Proof. Left as exercise. □

[8] The countable case is nor hard but requires some technical check.

[9] If it does not then it may not be well-defined.

5.1.1 problems

These problems are related to expectation.

1. $X \sim \text{Ber}p$. Show
$$\mathbb{E}[X] = p$$

2. $X \sim \text{Bin}(n, p)$, show
$$\mathbb{E}[X] = np$$

3. In the n coin toss example, suppose there is a probability p of coming up heads.
$$p(H) = \alpha, p(T) = 1 - \alpha$$
We aim to compute the expected number of heads in terms of n and α.

   (a) Our probability space would be different. 2.1. How should we modify p? In particular, give a formula for
$$p(\{\omega\})$$
   where ω is a sequence of H and T.

   (b) Let X_i be the random variable that takes 1 when a head occurs at the ith toss, so that
$$X_i(\omega) := \begin{cases} 1 & \text{if } \omega_i = \text{head} \\ 0 & \text{otherwise} \end{cases}$$
   where ω_i denotes the value of the ith toss.
$$\mathbb{E}(X_i) \quad 1 \leq i \leq n?$$

   (c) Let X be the random variable that counts the number of heads after n tosses. Express $\mathbb{E}(X)$ in terms of $\mathbb{E}(X_i)$, and hence find the value of $\mathbb{E}(X)$ in terns of n and α.

4. (*) Let $X \in \text{RV}^{\text{disc}}(\Omega, \mathcal{E}, p)$. Show that
$$\text{Var}(X) := \mathbb{E}(X^2) - E(X)^2$$

5. (**) Prove 5.4.

5.2 Continuity of measure

Omitted.

6 Central Limit Theorem

Learning Objectives

- Understand the statement of the central limit theorem *rigorously*.

$$\begin{tikzcd}
\{\text{improper integrals, } 12\} \arrow[d] & \\
\{\text{Continuous Random variables, } 4 \text{ , } 4.8\} \arrow[r] \arrow[d] & \{\text{indepndent random variables, } 10\} \arrow[dd] \\
\{\text{expectation and variance, } 5\} \arrow[dr] & \\
\{\text{normal distribution, } 11\} \arrow[r] & \{\text{Central Limit Theorem, } 6.2\}
\end{tikzcd}$$

The central limit theorem discusses the *stochastic convergence* of essentially random events eventually limiting to a event. We will make this idea precise.
There are various types of convergence we consider. The "weakest mode" of convergence is *convergence in distribution*.

Definition 6.1. Let $X_i \in \text{Seq}(\text{RV}(\Omega, \mathcal{E}, p))$, with F_i their associated distributions. We say X_n *converges to* X *in distribution*, or write

$$X_n \xrightarrow{\mathcal{D}} X$$

where $X \in \text{RV}^{\text{cts}}(\Omega, \mathcal{E}, p)$, is a random variable, with cdf, F if: for all $x \in X$ where F is continuous

$$\lim_{n\to\infty} F_n(x) \to F(x)$$

where F is continuous.

Theorem 6.2. Let $X_1, X_2, \ldots X_n$ be iid copies of a random variable X. Suppose $\mathbb{E}[X_i] = \mu, \text{Var}[X_i] = \sigma^2 < \infty$ for $i = 1, \ldots, n$. Then

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

$$\bar{X}_n := \frac{X_1 + \cdots X_n}{n}$$

6.0.1 Problems

1. Give 3 real life examples of normal distributions.
2. Give a one-page write-up of the the central limit theorem, 6.2. The content should be:
   (a) Explain every term in the theorem. This is combining and digesting the content of the notes.
   (b) Describe 2 applications of central limit theorem.

7 Independence of discrete random variables

Learning Objectives

- To get an understanding of independence of random variable in the discrete context. This is generalized in the continuous case.

We can formulate the notion of events being independent as follows:

Definition 7.1. Let $(\Omega, \mathcal{E}, p)$ be probability space. $A, B \in \mathcal{E}$ are *independent* if

$$p(A \cap B) = p(A)p(B)$$

This allows us to generalize into the context of random variables. If $X, Y \in \mathrm{RV}^{\mathrm{disc}}(\Omega, \mathcal{E}, p)$, we can ask what their *joint mass function* is. The two functions induces a map

$$(X, Y) : \Omega \to \mathbb{R}^2$$

$$(X, Y) : \omega \mapsto (X(\omega), Y(\omega))$$

For a given fixed point $(x, y) \in \mathbb{R}^2$, we have

$$(X, Y)^{-1}(x, y) := \{\omega \in \Omega \,:\, X(\omega) = x \text{ and } Y(\omega) = y\}$$

Definition 7.2. The *joint mass function* $p_{X,Y}$ of X, Y is the function

$$p_{X,Y} : \mathbb{R}^2 \to [0, 1]$$

$$p_{X,Y}(x, y) := p((X, Y)^{-1}(x, y)) \quad (x, y) \in \mathbb{R}^2$$

This is also written as

$$p_{X,Y} := p(X = x, Y = y)$$

What does it mean for the two random variables to be independent?

Definition 7.3. $X, Y \in \mathrm{RV}^{\mathrm{disc}}(\Omega, \mathcal{E}, p)$ are independent if the pair $\{X = x\}$ and $\{Y = y\}$ are independent, 7.1 for all $x, y \in \mathbb{R}$: mathematically

$$p(X = x, Y = y) = p(X = x)p(Y = y) \quad x, y \in \mathbb{R}$$

8 Generating sigma algebras

Learning Objectives

- To understand a common way to generate σ-algebras.

Definition 8.1. Let $\mathcal{E}_0 \subseteq 2^\Omega$ be any collection of subsets of Ω. The *smallest σ-algebra* containing $\mathcal{E}_0$, denoted

$$\sup{}^\sigma(\mathcal{E}_0)$$

is the unique σ-algebra on Ω satisfying

1. $\sup^\sigma(\mathcal{E}_0)$ is a σ-algebra on Ω which contains $\mathcal{E}_0$.
2. If $\mathcal{E}'$ is any other σ-algebra that contains $\mathcal{E}_0$, then it contains $\sup^\sigma(\mathcal{E}_0)$, i.e. $\sup^\sigma(\mathcal{E}_0) \subseteq \mathcal{E}$.

This is a common operation in many areas of maths. This is taking the least upper bound, *least upper bound*, def. **??** [10] in the context of the power set, $(2^\Omega, \subseteq)$. This operation only makes sense if it exists.

Proposition 8.2. Let $\mathcal{E}_0 \subseteq 2^\Omega$ be any collection of subsets of Ω. The operation of *smallest σ-algebra* 8.1 exists and is unique.

Proof. We show that $\sup^\sigma(\mathcal{E}_0)$ exists. To do this:

1. Note that there exists at least *one* σ algebra which contains $\mathcal{E}_0$. $\mathcal{E}_0 \subseteq 2^\Omega$.
2. Let $\mathcal{A}$ be the collection of all σ-algebras $\mathcal{E}'$ which contains $\mathcal{E}_0$. Then

$$\bigcap_{\mathcal{E}' \in \mathcal{A}} \mathcal{E}'$$

is the smallest σ-algebra which contains $\mathcal{E}_0$. (Why?)

□

9 The Borel σ-algebra on $\mathbb{R}$

This section may be technical on the first read. See comment after

Definition 9.1. $E \subseteq \mathbb{R}$ is open if for all $x \in E$, there is some $\varepsilon > 0$, such that $(x - \varepsilon, x + \varepsilon) \subseteq E$.

[10] but where as the relation there is the order relation, $(\mathbb{R}, \leq)$

Then

Definition 9.2. The *Borel σ-algebra on* $\mathbb{R}$ is

$$\mathcal{B}(\mathbb{R}) := \sup{}^{\sigma} \text{open}(\mathbb{R})$$

where open($\mathbb{R}$) is the set of opens in $\mathbb{R}$.

But how do we have a handle of such a complicated set?

Proposition 9.3. 1. If $E \subseteq_{\text{open}} \mathbb{R}$ then there exists at most countably many disjoint open interval I_j such that

$$E = \bigcup_{j=1}^{\infty} I_j$$

2. We have the following equality

$$\mathcal{B}(\mathbb{R}) = \sup{}^{\sigma}(\{(-\infty, a] \,:\, a \in \mathbb{Q}\})$$

9.0.1 Problems

1. Complete the proof of Prop. 8.2
   - Show why the construction explained in proof yields a sigma algebra.
   - Show that the newly constructed is unique.
2. (****) Prove 1. of 9.3.
3. (****) Prove 2. of 9.3.

10 Independence of random variables

Reading: [1, 6]. The definition of independent discrete random variables comes from studying the joint mass function, 7.2. In the general case of arbitrary random variables, the independence[11] is phrased more appropriately using their joint distribution.

Definition 10.1. Let $X, Y \in \text{RV}(\Omega, \mathcal{E}, p)$ Their *joint distribution function* is the map given by

$$F_{X,Y}(x, y) := p(X \leq x, Y \leq y) := p\left(\{\omega \in \Omega \,:\, X(\omega) \leq x, Y(\omega) \leq y\}\right)$$

[11] This new definition subsumes the old one

Definition 10.2. $X, Y \in \mathrm{RV}(\Omega, \mathcal{E}, p)$ are *independent* if for all $x, y \in \mathbb{R}$, the events $\{X \leq x\}$ and $\{Y \leq y\}$, or equivalently

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad x, y \in \mathbb{R}$$

The independence of n random variables is given as

Proposition 10.3. [4, 2, Exercise 21] Let $X_1, \ldots, X_n \in \mathrm{RV}(\Omega, \mathcal{E}, p)$. Then X_i are jointly independent iff

$$p\left(\bigcap_{i=1}^{n} \{X_i \leq t_i\}\right) = p\left(\bigcap_{i=1}^{n} X_i^{-1}((-\infty, t_i])\right) = \prod_{i=1}^{n} p(X_i \leq t_i)$$

where $\{X_i \leq t_i\} := \{\omega \in \Omega \,:\, X_i(\omega) \leq t_i\}$.

In greater generality: note that the random variables X_i may not just land values in $\mathbb{R}$. [12] So, a more appropriate way to phrase joint independence of n random variables is the following definition.

Definition 10.4. Let $X_1, X_2, \ldots X_n \in \mathrm{RV}(\Omega, \mathcal{E}, p)$ be n random variables. Then $X_1, X_2, \ldots$ are *jointly independent* if for all tuples of events $\{E_i\}_{i=1}^n$ in $B(\mathbb{R})$ (so each $E_i \in B(\mathbb{R})$, the borel set)

$$p(\bigcap_{i=1}^{n} X_i^{-1}(E_i)) = \prod_{i=1}^{n} p(X_i^{-1}(E_i))$$

10.0.1 Problems

[12] for instance, they can be valued in $\mathbb{R}^m$ for some integer $m \in \mathbb{N}_{\geq 0}$. Some people call this a random vector.

11 The normal distribution

Learning Objectives

- Do computations with the standard normal density function, 11.1.

The normal distribution is one of the most important *continuous* distributions. Many populations have distribution that fit closely to an appropriate normal curve.

1. height, weight physical characteristic.

One reason normal distribution is important is because of the central limit theorem, 6.2.

Definition 11.1. The *standard normal density function of mean μ and variance σ^2*

$$f_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2}} \qquad f_X : \mathbb{R} \to \mathbb{R} \tag{1}$$

where $\mu, \sigma \in \mathbb{R}$.

Let us list properties of the graph of f_X.

1. Normal distributions satisfy that they are closed under scaling and shifting.

Definition 11.2. Let $X_1, \ldots, X_n$ be random variables. We denote

$$\bar{X}_n := \frac{\sum_{i=1}^n X_i}{n}$$

Theorem 11.3. Weak Law of Large Numbers. Let $X_1, \ldots, X_n$ be iid random variables, such that $\mathbb{E}(X_i) = \mu < \infty$. Then

$$\lim_{n\to\infty} p(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

11.0.1 Problems

The following problems consists of content that covers various part of previous sections.

1. Graph the density function, $f_{\sigma,\mu}$, for the standard normal distribution, def 11.1.

2. (****) The expoential distribution function X is given by

$$F_X(x) := \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

(a) Describe a desntiy function for F_X. Hence, show that that X is a continuous random variable.

(b) Show that $f_X(a) = F'_X(a)$ at $a \in \mathbb{R}$ where the derivatives of F_X exists.

(c) Give one real-life example which the exponential distribution function models.

3. Let $Z \sim N(0,1)$, be a normal random variable with mean 0 and variance 1.

   (a) find the value of (e.g. on lookup tables available online)

   $$p(Z \leq -1.25), p(Z \geq 1, 25)$$

   (b) Show that

   $$p(Z \leq -x) = p(Z \geq x)$$

4. Let X be a discrete random variable such that

   $$p(X = k) = p_k \quad k = 0, 1, 2, \dots$$

   $$\sum_{k=0}^{\infty} p_k = 1$$

   Explain why

   $$F_X(x) = \begin{cases} 0 & x < 0 \\ p_0 + \cdots + p_{\lfloor x \rfloor} & x \geq 0 \end{cases}$$

5. (**) Continuity of random variable. Let $X \in \mathrm{RV}(\Omega, \mathcal{E}, p)$.

   - Show that if $p(X = c) > 0$, then F_X has a discontinuity at c. Is the converse true?
   - Let X be a continuous random variable. Show that $p(X = x) = 0$ for all $x \in \mathbb{R}$.

12 Improper Integral

Reading: [3, p283].

Definition 12.1. Let $a \in \bar{\mathbb{R}}$, see def. **??**, $G : (a, +\infty) \to \mathbb{R}$ be a function.

$$\lim_{t \to +\infty} G$$

exists if there exists $L \in \bar{\mathbb{R}}$, such that

1. If $L \in \mathbb{R}$ finite: for all $\varepsilon > 0$, there exists $N_\varepsilon > 0$, such that for all $t > N_\varepsilon$,

$$|L - G(t)| < \varepsilon \quad \forall t > N_\varepsilon$$

2. If $L = +\infty$: for all $M \in \mathbb{R}$, there exists N_M such that for all $t > N_M$,

$$G(t) > M$$

3. If $L = -\infty$: for all $M \in \mathbb{R}$, there exists N_M such that for all $t > N_M$,

$$G(t) < M$$

This allows us to define the limit of integral.

Definition 12.2. For all $t > a$, if

$$G(t) := \int_a^t f(x)\, dx$$

exists and is finite, we can define a function

$$G : (a, +\infty) \to \mathbb{R}$$

We define

$$\int_a^\infty f(x)\, dx := \lim_{t \to +\infty} \int_a^t f(x),\, dx$$

if it exists and is *finite*. We say the integral $\int_a^\infty f(x),\, dx$ is *convergent* and *divergent* otherwise.

similarly we can define:

Definition 12.3. For all $t < b$ if

$$G(t) := \int_t^b f(x)\, dx$$

exists and is finite, define

$$G : (-\infty, b) \to \mathbb{R}$$

$$\int_{-\infty}^{b} f(x)\,dx := \lim_{t\to-\infty} \int_{t}^{b} f(x)\,dx$$

provided it exists and is finite. [13] We say the integral $\int_{-\infty}^{t} f(x)$, dx is *convergent* and *divergent* otherwise.

We can then extend this definition to two sided integral.

Definition 12.4. Let $c \in \mathbb{R}$. If

$$\int_{-\infty}^{c} f(x)\,dx \text{ and } \int_{c}^{\infty} f(x)\,dx$$

are convergent, then

$$\int_{-\infty}^{\infty} f(x)\,dx := \int_{-\infty}^{c} f(x)\,dx + \int_{c}^{\infty} f(x)\,dx$$

Note: $\int_{-\infty}^{\infty} f(x)\,dx$ exists, then for any other choice of c',

$$\int_{-\infty}^{c'} f(x)\,dx + \int_{c'}^{+\infty} f(x)\,dx = \int_{-\infty}^{c} f(x)\,dx + \int_{c}^{\infty} f(x)\,dx$$

Example

1. $\int_{1}^{\infty} \frac{1}{x}\,dx$. Recall

12.0.1 Problems:

??

1. Complete the definition of the negative improper integral, 12.3, where I have not defined $\lim_{t\to-\infty} G(t)$. i.e. let $G : (-\infty, b) \to \mathbb{R}$, be a function. Give a definition of $\lim_{t\to-\infty} G$ converging to a vale $L \in \mathbb{R}$, this should be a modification of def. 12.1.

2. Show that

$$\int_{-\infty}^{0} \frac{1}{2-x}\,dx$$

is divergent.

3. Compute

$$\int_{-\infty}^{\infty} xe^{-x^2}\,dx$$

[13] We leave the reader to fill in the definition of $\lim_{t\to-\infty}$ for a function $H : (-\infty, b) \to \mathbb{R}$.

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.

[3] Michael Spivak, *Calculus, 4th edition*.

[4] Terence Tao, *275a* (2015).