Math 311-1 Lecture Notes

Andrew Li

Fall Quarter 2024

Original lecture notes for **Math 311-1: MENU Probability and Stochastic Processes**, from Fall Quarter 2024, taught by Professor Benjamin Weinkove. This course follows Basic Probability Theory by Robert Ash, ISBN 978-0-4886-46628-6.

Table of Contents

§1 September 24, 2024

§1.1 Introduction

Introducing a probability course first requires a rigourous definition of a probability space, and some brief review of set theory.

Proposition 1.1 Under the classical definition of probability, the probability of some event is defined as

$$\mathbb{P}(\text{event}) = \frac{\#\,\text{favourable outcomes}}{\#\,\text{total outcomes}} \tag{1.1}$$

For example, for rolling a (fair) six-sided dice, the probability of each of the six sides landing up is

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \ldots = \mathbb{P}(\{6\}) = 1/6 \tag{1.2}$$

For flipping two coins, the notation more clearly implicates why events are defined as sets as opposed to distinct elements.

$$\mathbb{P}(\{HH, HT, TH\}) = 3/4 \tag{1.3}$$

or phrased in English, the probability of flipping at least one head after flipping two (fair) coins is $3/4$. Note here that

$$\{HH, HT, TH\}^C = \{\text{TT}\} \tag{1.4}$$

which obviously has a $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$ probability.

§1.2 Probability Spaces

Definition 1.2 A **sample space** Ω is defined as the set of possible outcomes of some random experiment.

Definition 1.3 An **event space** $\mathcal{F}$ is defined as some set of events, which are subsets of some sample space Ω. That is, an event is some set of outcomes, such as $\{2, 4, 6\}$ being the even sides of a dice.

Definition 1.4 A **probability space** is a triplet

$$(\Omega, \mathcal{F}, \mathbb{P}) \tag{1.5}$$

consisting of Ω a sample space (a set), $\mathcal{F}$ an event space (a σ-algebra of subsets/events), and $\mathbb{P} : \mathcal{F} \to [0, 1]$ (a probability measure on $\mathcal{F}$).

1.2.1 Sigma Algebra

Definition 1.5 A collection of subsets ($\mathcal{F}$) of some set (Ω) is a **σ-algebra** if

$$\begin{cases} \Omega \in \mathcal{F} \\ A_1, ..., A_\infty \in \mathcal{F} \implies \bigcup_i^\infty A_i \in \mathcal{F} \\ A \in \mathcal{F} \implies A^C \in \mathcal{F} \end{cases} \tag{1.6}$$

It follows trivially that

Lemma 1.6 A σ-algebra always contains $\emptyset$

Proof. Suppose some σ-algebra $\mathcal{F}$ does not contain the empty set. By definition, $\Omega \in \mathcal{F}$, and by definition, $\Omega^C \in \mathcal{F}$. However, $\Omega^C = \emptyset$, which is a contradiction. □

It follows slightly less trivially that

Example 1.7 It is not necessarily true that $\mathcal{F}$ contains **all** subsets of Ω. As a trivial example, let $\Omega = \{1, 2, 3, 4, 5, 6\}$. Then,

$$\mathcal{F} = \{\emptyset, \{1\}, \{2, 3, ..., 6\}, \{1, 2, ..., 6\}\} \tag{1.7}$$

is clearly a σ-algebra, and is easy to see per Definition 1.5.

1.2.2 Probability Measure

Definition 1.8 A **probability measure** is some function

$$\mathbb{P} : \mathcal{F} \to [0, 1] \tag{1.8}$$

such that

$$\mathbb{P}(\Omega) = 1 \tag{1.9}$$

where Ω is a sample space, i.e. all possible outcomes.

The probability of an event (a set) corresponds to the sum of all outcomes within that event. Suppose $\Omega = \{1, 2, 3, ..., N\}$; then

$$\mathbb{P}(\Omega) = \mathbb{P}(1) + \mathbb{P}(2) + ... + \mathbb{P}(N) \tag{1.10}$$

Proposition 1.9 In general, let $A_1, A_2, ...$ be disjoint subsets of Ω; then

$$\mathbb{P}\left(\bigcup_i^\infty A_i\right) = \sum_i^\infty \mathbb{P}(A_i) \tag{1.11}$$

Definition 1.10 Two sets A_i and A_j are disjoint if

$$A_i \cap A_j = \emptyset \impliedby i \neq j \tag{1.12}$$

A simple example for this is rolling a six-sided (fair) dice. In this case, $\mathcal{F}$ is the set of all $2^6 = 64$ subsets of Ω. We can easily see that

$$\mathbb{P}(\{1\}) = 1/6 \tag{1.13}$$
$$\mathbb{P}(\{2\}) = 1/6 \tag{1.14}$$
$$\mathbb{P}(\{1, 2\}) = 1/6 + 1/6 = 2/6 \tag{1.15}$$

Note that the probabilities are not derived based on anything (though we could use physics); we use probability as a model for the world based on how we define the probabilities of certain events.

§1.3 Digression on Set Theory

Suppose A, B, C are sets. The operations $\cup$, $\cap$, and ^{C} are closed and have the following properties:

1. Commutativity

$$A \cap B = B \cap A \tag{1.16}$$
$$A \cup B = B \cup A$$

2. Associativity

$$A \cap (B \cap C) = (A \cap B) \cap C \tag{1.17}$$
$$A \cup (B \cup C) = (A \cup B) \cup C$$

3. Distributivity

$$A \cup (B \cap C) = (A \cup B) \cap (B \cup C) \tag{1.18}$$
$$A \cap (B \cup C) = (A \cap B) \cup (B \cap C)$$

1.3.1 De Morgan's Laws

Suppose $A_1, A_2, \ldots$ are sets. Then,

Lemma 1.11

$$\left(\bigcap_{n}^{\infty} A_n\right)^C = \bigcup_{n}^{\infty} A_n^C \tag{1.19}$$

Proof. For some element x,

$$x \in \left(\bigcap_{n}^{\infty} A_n\right)^C \iff \exists n \mid x \notin A_n \tag{1.20}$$
$$\iff \exists n \mid x \in A_n^C \tag{1.21}$$
$$\iff x \in \bigcup_{n}^{\infty} A_n^C \tag{1.22}$$

(1.20) follows because if x is in the complement of the intersection of all of the sets, that necessarily means it must not be in that intersection, i.e. not be in at least one set. (1.21) follows trivially: given the previous statement, x must be in the complement of one of the sets. So, (1.22) follows because x is in at least one of the complements which is a subset of the union of all of them. □

Lemma 1.12

$$\left(\bigcup_{n}^{\infty} A_n\right)^C = \bigcap_{n}^{\infty} A_n^C \tag{1.23}$$

Proof.

$$x \in \left(\bigcup_{n}^{\infty} A_n\right)^C \iff x \notin A_n \mid \forall n \tag{1.24}$$

$$\iff x \in A_n^C \mid \forall n \tag{1.25}$$

$$\iff x \in \bigcap_{n}^{\infty} A_n^C \tag{1.26}$$

(1.24) follows because for x to not be in the union of all of these sets, then x cannot be an element of any of them, which implies (1.25) because that means x must simultaneously be an element of the complement of all of the sets. For that to be true requires x to be an element of the intersection of all A_n^C. ☐

§2 September 25, 2024

Recall that a probability space is defined as a triplet $(\Omega, \mathcal{F}, \mathbb{P})$. Further recall the three conditions that define a σ-algebra. Finally, recall the definition of a probability measure. Using these, we can define some properties.

§2.1 Probability Space Properties

2.1.1 Properties of a Sigma Algebra

Let $\mathcal{F}$ be a σ-algebra. Then,

1. $\emptyset \in \mathcal{F}$ which is proven in Lemma 1.6.
2. Closedness under union

$$A_1, \ldots, A_N \in \mathcal{F} \implies \bigcup_{i}^{N} A_i \in \mathcal{F} \tag{2.1}$$

Proof. Using $A_{N+1}, ... \equiv \emptyset$, the union of all of these must be an element of $\mathcal{F}$ □

3. Closedness under intersection

$$A_1, \ldots, A_N \in \mathcal{F} \implies \bigcap_{i}^{N} A_i \in \mathcal{F} \tag{2.2}$$

Proof. Take

$$A_1, \ldots, A_N \in \mathcal{F} \tag{2.3}$$

Then, by definition,

$$A_1^C, \ldots, A_N^C \in \mathcal{F} \tag{2.4}$$

By definition and then De Morgan's Laws,

$$A_1^C, \ldots, A_N^C \in \mathcal{F} \implies \bigcup A_i^C \in \mathcal{F} \implies \left(\bigcap A_i\right)^C \in \mathcal{F} \implies \bigcap A_i \in \mathcal{F} \ \square$$

2.1.2 Generated Sigma Algebras

Let $\Omega \equiv \mathbb{N}$. Then,

$$\mathcal{F} \equiv \{\emptyset, \{1\}, \{1, 2\}, ...\} \tag{2.5}$$

is not a σ-algebra because $\{1, 2\} - \{1\} = \{2\}$ is not an element of $\mathcal{F}$. In spite of this, we can define a σ-algebra $\widetilde{\mathcal{F}}$ to be the intersections of all σ-algebras that contain $\mathcal{F}$, i.e. for some non-sigma-algebra subset of sets $\mathcal{F}$,

$$\widetilde{\mathcal{F}} \equiv \bigcap \mathcal{G} \qquad \forall \mathcal{G}_\sigma \supset \mathcal{F} \tag{2.6}$$

Proposition 2.1 $\widetilde{\mathcal{F}}$ is a σ-algebra.

For the case in (2.5) specifically, we assert that $\widetilde{\mathcal{F}} \equiv 2^{\mathbb{N}}$, i.e. the power set of natural numbers. To show this, see that if $\{1\} \in \mathcal{F}$ then $\{2\} \in \mathcal{F}$ for $\{1, 2\} \in \mathcal{F}$. For similar reasons,

$$\{n\} \in \mathcal{F} \qquad \forall n \in \mathbb{N} \tag{2.7}$$

and by taking the union of these sets, all subsets of $\mathbb{N}$ can be composed of these singleton subsets.

2.1.3 Properties of a Probability Measure

1. The probability of nothing... is nothing!

$$\mathbb{P}(\emptyset) = 0 \tag{2.8}$$

Proof.

$$\mathbb{P}(\emptyset \cup \Omega) = \mathbb{P}(\Omega) = \mathbb{P}(\emptyset) + \mathbb{P}(\Omega) \implies \mathbb{P}(\emptyset) = 0 \tag{2.9}$$

□

2. Probability of unions

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \tag{2.10}$$

To prove this without using pictures, we can express A, B, and $A \cup B$ as disjoint sets

$$A \cup B = (A - B) \cup (A \cap B) \cup (B - A) \tag{2.11}$$
$$A = (A - B) \cup (A \cap B) \tag{2.12}$$
$$B = (B - A) \cup (A \cap B) \tag{2.13}$$

This means that

$$\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A - B) + \mathbb{P}(B - A) + 2\mathbb{P}(A \cap B) \tag{2.14}$$

and clearly, there is one extra $\mathbb{P}(A \cap B)$.

3. Subset probability. Given $B \subseteq A$, $\mathbb{P}(A - B) = \mathbb{P}(A) - \mathbb{P}(B)$ which implies $\mathbb{P}(B) \leq \mathbb{P}(A)$ due to non-negative probabilities being not possible by definition. We can write A and B as disjoint sets,

$$A = B \cup (A - B) \tag{2.15}$$
$$B = B \tag{2.16}$$

and then

$$\mathbb{P}(A) = \mathbb{P}(B) + \mathbb{P}(A - B) \tag{2.17}$$
$$\mathbb{P}(B) = \mathbb{P}(B) \tag{2.18}$$
$$\therefore \mathbb{P}(A) - \mathbb{P}(B) = \mathbb{P}(A - B)(\geq 0) \tag{2.19}$$

4. Union probability less than sum

$$\mathbb{P}\left(\bigcup A_i\right) \leq \sum \mathbb{P}(A_i) \tag{2.20}$$

Proof. We can again write these in a different way

$$\bigcup A_i = (A_1) \cup (A_1^C \cap A_2) \cup (A_1^C \cap A_2^C \cap A_3) \cup \cdots \tag{2.21}$$

Note that by Property 3, $\mathbb{P}(A_1^C \cap A_2) \leq \mathbb{P}(A_2)$ and the same applies to all of the other ones. So,

$$\mathbb{P}\left(\bigcup A_i\right) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots \qquad \square$$

§2.2 Combinatorics and Counting

Take $\Omega = \{a_1, \cdots, a_N\}$ as some sample space with $|\Omega| = N$. Take $\mathcal{F} \equiv 2^\Omega$ as all subsets of Ω and define

$$\mathbb{P}(a_i) = 1/N \qquad \forall i \tag{2.22}$$

which generalizes to

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \tag{2.23}$$

To compute $\mathbb{P}$, we have to be able to count $|A|$, which requires an overview of combinatorics.

2.2.1 Ordered with Replacement

Suppose we have a license plate with five letters. Then, there are 26^5 possible combinations because we can reuse letters, and the order matters. In general, for a set of size N and R repeats, there are

$$N^R \qquad \text{permutations} \tag{2.24}$$

2.2.2 Ordered without Replacement

If we do not replace then we cannot reuse letters. So, for the license plate we have $26 \cdot 25 \cdots 22$ combinations. In general, for N and R, we have

$$\frac{N!}{(N-R)!} = {}^N P_R = (N)_R \qquad \text{permutations} \tag{2.25}$$

2.2.3 Unordered without Replacement

By the same logic, we now have

$$\frac{N!}{R!(N-R)!} = \binom{N}{R} \qquad \text{permutations} \tag{2.26}$$

§3 September 27, 2024

§3.1 Counting Problems (cont)

Recall counting formulas:

1. Ordered samples of r objects out of n with replacement is n^r
2. Ordered samples of r objects out of n with*out* replacement is $\frac{n!}{(n-r)!}$
3. Unordered samples of r objects out of n without replacement is $\frac{n!}{r!(n-r)!} = \binom{n}{r}$

Example 3.1 There are 10 balls in an urn numbered 1 through 10. Randomly draw 5 balls without replacement. What is the probability of the second largest number being 8?

Solution. Ways to choose 5 balls out of 10 is

$$\binom{10}{5} \tag{3.1}$$

if we do not care about the order. How many of these combinations have the second largest number of 8? There are two possibilities: largest number is 9 or largest number is 10. So there are

$$\underbrace{2}_{\text{9 or 10}} \times \underbrace{\binom{7}{3}}_{\text{choose 3 from 7 remaining}} \tag{3.2}$$

This sets one choice as 8, one as one of 9 or 10, and the rest as arbitrary picks that are not 8, 9, or 10. So the probability is

$$\mathbb{P} = \frac{2 \times \binom{7}{3}}{\binom{10}{5}} \tag{3.3}$$

□

3.1.1 Unordered Samples With Replacement

How many Scrabble combinations of 7 letters are there if there are only A, B, and C? It is not as simple as $N^R/R!$ because there can be repeated elements which adds a degree of nuance. We can rewrite some sequence using "stars and bars" into

$$AABCABA \implies * * * * \mid * * \mid * \tag{3.4}$$

To count the number of combinations, there is one slot for each star and one slot for each bar. Each slot can either be a star or a bar. So we pick 7 positions from 9 positions:

$$\binom{9}{7} \equiv \binom{9}{2} = 36 \tag{3.5}$$

ways to arrange these stars and bars.

Proposition 3.2 The number of unordered samples of R objects out of N is

$$\binom{R+N-1}{R} = \binom{R+N-1}{N-1} \tag{3.6}$$

§3.2 Independence

Proposition 3.3 Let A and B be two independent events. We say A and B are independent if $\mathbb{P}(A \text{ and } B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

What if we have N events? How can this definition be generalized?

Definition 3.4 A family of events $\mathcal{A} \equiv \{A_i\}_{i \in I}$ (where I is some set of indices such as $\mathbb{N}$) are **independent** if and only if for every finite subset $A' \subseteq \mathcal{A}$,

$$\mathbb{P}\left(\bigcap A_i'\right) = \prod \mathbb{P}(A_i') \tag{3.7}$$

3.2.1 Properties of Independent Events

Recall the properties of a sample space defined previously.

Lemma 3.5 Let Ω be a sample space and let A be an event with probability $\mathbb{P}(A)$. Then, $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$.

Proof. It must be true that

$$A \cup A^C = \Omega \tag{3.8}$$

so

$$\mathbb{P}(A) + \mathbb{P}(A^C) = \mathbb{P}(A \cup A^C) = \mathbb{P}(\Omega) = 1 \tag{3.9}$$

□

1. If A and B are independent, then $\mathbb{P}(A \cap B^C) = \mathbb{P}(A) \cdot \mathbb{P}(B^C)$. That is, if an event is independent with another, then the first event is independent

with the other not happening as well.

$$A, B \text{ indep.} \implies \left\{A, A^C\right\} \times \left\{B, B^C\right\} \text{ all indep.} \tag{3.10}$$

Proof. Given A and B are independent, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Then,

$$\mathbb{P}(A \cap B^C) = \mathbb{P}(A - B) = \mathbb{P}(A - (A \cap B)) \tag{3.11}$$

But, if $B \subseteq A$, then $\mathbb{P}(A - B) = \mathbb{P}(A) - \mathbb{P}(B)$. So,

$$\begin{aligned} \mathbb{P}(A \cap B^C) \overset{...}{=} \mathbb{P}(A - (A \cap B)) &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) &(3.12)\\ &= \mathbb{P}(A)\left[1 - \mathbb{P}(B)\right] &(3.13)\\ &= \mathbb{P}(A)\mathbb{P}(B^C) &(3.14) \end{aligned}$$

□

This works for the other two non-trivial elements of (3.10) as well.

Proposition 3.6 For some family $\{A_i\}_{i\in I}$ of independent events, define $B_\alpha \equiv (A_\alpha \text{ or } A_\alpha^C)$. Then,

$$\mathbb{P}\left(\bigcap B_i\right) = \prod \mathbb{P}(B_i) \tag{3.15}$$

where here, each event is either some event in A or its complement.

§3.3 Bernoulli Trials

Suppose some factory produces batteries and 5% of all batteries are defective. These are independent defective batteries which makes the QA job difficult. Suppose the factory makes 10 batteries. What is the probability that exactly 3 of them are defective? The first three being defective has probability

$$\mathbb{P}\left[\text{first 3 are defective; rest are ok}\right] = \left(\frac{1}{20}\right)^3 \cdot \left(\frac{19}{20}\right)^7 \tag{3.16}$$

But, there are multiple ways to pick three, so the probability there is

$$\binom{10}{3} \cdot \mathbb{P}\left[\text{first 3 are defective; rest are ok}\right] \tag{3.17}$$

Definition 3.7 For repeating some binary event with probability of success p for N independent trials, the probability of succeeding exactly k times is

$$\mathbb{P}(k \text{ successes}) = \binom{N}{k} \cdot p^k \cdot (1-p)^{N-k} \tag{3.18}$$

§4 September 30, 2024

§4.1 Generalized Bernoulli Trials

Recall for n independent trials and the probability of success for each trial is p. Then,

$$\mathbb{P}(k \text{ successes}) = \binom{n}{k} p^k (1-p)^{n-k} \tag{4.1}$$

What if there are more than two outcomes? Let there be n independent trials and k possible outcomes $b_1, \ldots, b_k$ for each trial with associated probabilities $p_1, \ldots, p_k$ ($\sum p_i = 1$), e.g. rolling a dice n times.

The sample space Ω = set of all finite sequences of length n where each entry is one of $b_1, \ldots, b_k$ (there are k^n such sequences). We want to compute the probability of exactly n_1 occurrences of b_1, $n_2 \to b_2, \ldots, n_k \to b_k$. Define this as

$$p(n_1, n_2, \ldots, n_k) \qquad \sum n_i = n \tag{4.2}$$

First, we can compute the probability of the first n_1 being b_1; next n_2 being b_2; etc. That equals

$$p_1^{n_1} \cdots p_k^{n_k} \tag{4.3}$$

We also need to scale by the total number of arrangements, i.e. number of ways to get n_1 occurrences of $b_1, \ldots, n_k$ of b_k. That equals

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n_k}{n_k} \tag{4.4}$$

Can we simplify this? After some trivial arithmetic (write it out), this reduces to

$$\frac{n!}{n_1! n_2! n_3! \cdots n_k!} = n! \cdot \prod_i (n_i)!^{-1} \tag{4.5}$$

so the total probability is

Definition 4.1 Generalized Bernoulli Trials

$$p(n_1, n_2, \ldots, n_k) = n! \cdot \left[\prod_i (n_i)!^{-1}\right] \cdot p_1^{n_1} \cdots p_k^{n_k} \tag{4.6}$$

Example 4.2 Take an urn with black, white, red, and green balls. Randomly and independently draw four balls with replacement. What is the probability that I have exactly two distinct colors?

Solution. Let $(b_1, b_2, b_3, b_4) = $ (black, white, red, green). Each is equally likely. What is the probability I have exactly two black and two white?

$$\mathbb{P}(b_1 = 2, b_2 = 2, 0, 0) = \frac{4!}{2!2!} \cdot (1/4)^2 \cdot (1/4)^2 = 3/128 \tag{4.7}$$

That is a simple case. For two of one color and two of another, there are $\binom{4}{2}$ ways to pick two colors, which is six. So, that sub-probability is $18/128 = 9/64$ (two of one color; two of another). We still have to do one of one color and three of another.

$$\mathbb{P}(3, 1, 0, 0) = \frac{4!}{3!1!} \cdot (1/4)^3 \cdot (1/4)^1 = 1/64 \tag{4.8}$$

There are twelve ways to do this (six in one way; six in the other), so this sub-probability is $12/64$. The total is $21/64 \approx 1/3$. □

§4.2 Conditional Probability

Take two events A, B in some probability space. What is

$$\mathbb{P}(A \mid B) \quad \text{and} \quad \mathbb{P}(B \mid A) \tag{4.9}$$

or the probabilities of some event happening given another has happened? Intuitively,

Definition 4.3 The probability B occurs *given* A occurs is

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \tag{4.10}$$

Lemma 4.4 If A and B are independent,

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(A)} = \mathbb{P}(B) \tag{4.11}$$

which is quite intuitive.

Example 4.5 Roll a fair die once. A is rolling an odd number and B is rolling a 5.

$$\mathbb{P}(B \mid A) = \frac{1/6}{1/2} = \frac{1}{3} \tag{4.12}$$

Example 4.6 Throw two dice. Let A be that the highest roll is a six; let B be that the sum is a ten.

$$\mathbb{P}(B \mid A) = \frac{2/36}{11/36} = 2/11 \qquad \mathbb{P}(A \mid B) = \frac{2/36}{3/36} = 2/3 \tag{4.13}$$

Notice that

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A) = \mathbb{P}(B)\mathbb{P}(A \mid B) \tag{4.14}$$

This yields

Theorem 4.7 Bayes' Theorem

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\mathbb{P}(B \mid A)}{\mathbb{P}(B)} \tag{4.15}$$

What happens when there are multiple events, namely A, B, C, D?

$$\begin{aligned} \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A \cap B)\mathbb{P}(C \mid A \cap B) && (4.16) \\ &= \mathbb{P}(A)\mathbb{P}(B \mid A)\mathbb{P}(C \mid A \cap B) && (4.17) \end{aligned}$$

We can keep going.

Proposition 4.8

$$\mathbb{P}\left(\bigcap A_i\right) = \prod_i \mathbb{P}\left(A_i \mid \bigcap_j^{i-1} A_j\right) \tag{4.18}$$

Example 4.9 Draw three cards randomly from a regular deck without replacment. Find the probability that there is no ace in the three cards. Let $A_i = i -$ th card is not an ace

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \tag{4.19}$$

We can enumerate some probabilities

$$\begin{aligned} A_1 &= 48/52 && \text{52 total cards; 4 aces} && (4.20) \\ A_2 &= 47/51 && && (4.21) \\ A_3 &= 46/50 && && (4.22) \end{aligned}$$

§4.3 Law of Total Probability

Definition 4.10 Events $B_1, B_2, \ldots$ are **mutually exclusive** if they are disjoint.

Definition 4.11 Events $B_1, B_2, \ldots$ are exhaustive if $\Omega = \bigcup_i B_i$

Combining these two yields the law of total probability

Theorem 4.12 Let $B_1, B_2, \ldots$ be mutually exclusive and exhaustive. Then,

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) \tag{4.23}$$

and

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i)\mathbb{P}(A \mid B_i) \tag{4.24}$$

for all $i \mid \mathbb{P}(B_i) > 0$.

§5 October 2, 2024

§5.1 Law of Total Probability (properly)

Theorem 5.1 Take some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $B_1, B_2, \ldots \in \mathcal{F}$ are mutually exclusive and exhaustive (see 4.3), for $A \in \mathcal{F}$

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) \tag{5.1}$$

Proof. Trivially,

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) \tag{5.2}$$

But, $B_1 \cap \cdots = \Omega$. So this equals

$$\mathbb{P}(A) = \mathbb{P}(A \cap (B_1 \cap \cdots)) \tag{5.3}$$

$$= \mathbb{P}\left(\bigcup (A \cap B_i)\right) \tag{5.4}$$

Since B_i are disjoint, these are all disjoint, so this becomes

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) \tag{5.5}$$

□

Recall conditional probability if $\mathbb{P}(A) \neq 0$

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \tag{5.6}$$

Theorem 5.2 For $\mathbb{P}(B_i) > 0$, Theorem 5.1 using conditional probabilities equals

$$\mathbb{P}(A) = \sum_i \mathbb{P}(B_i)\mathbb{P}(A \mid B_i) \tag{5.7}$$

§5.2 Bayes' Formula

Proposition 5.3 For $B_1, B_2, \ldots$ mutually exclusive and exhaustive,

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} \tag{5.8}$$

which can be rewritten into

$$\boxed{\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(B_k)\mathbb{P}(A \mid B_k)}{\sum_i \mathbb{P}(B_i)\mathbb{P}(A \mid B_i)}} \tag{5.9}$$

Example 5.4 Throw a die with outcome $i \in \{1, \ldots, 6\}$. Then, flip a coin i times. Find the conditional probability that the dice landed on 3 given at least one head was obtained.

Solution. The professor drew a beautiful, branching tree. Each outcome of a dice corresponds to some probabilities in terms of number of heads. This can be used to easily compute the conditional probability, using Equation 5.9. □

§5.3 Borel Sets

Take $\Omega = \mathbb{R}$.

Definition 5.5 The Borel σ-algebra on $\mathbb{R}$, $\mathcal{B}$, is the σ-algebra of subsets of $\mathbb{R}$ generated by[a] the closed intervals for all $[a, b]$ where $a \leq b$.

[a] smallest σ-algebra containing

The elements of $\mathcal{B}$ are Borel sets.

Proposition 5.6 Some facts

$$(a, b) = \bigcup_{n=1}^{\infty} [a + 1/n, b - 1/n] \implies (a, b) \in \mathcal{B} \tag{5.10}$$

$$[a, \infty) = \bigcup_{n=1}^{\infty} [a, a + n] \implies [a, \infty) \in \mathcal{B} \tag{5.11}$$

$$[z, z] \in \mathcal{B} \, \forall z \in \mathbb{Z} \implies \mathbb{Z} \in \mathcal{B} \tag{5.12}$$

5.3.1 Example of a probability space involving $\mathcal{B}$

Let $\Omega = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}$. Let $f(x)$ be a normalized and non-negative (Riemann) integrable function on $\mathbb{R}$. Then, define the probability measure

$$\mathbb{P}(B) = \int_B f(x)\,\mathrm{d}x \tag{5.13}$$

But how do we integrate over a Borel set? We use the unique probability measure $\mathbb{P}$ on $\mathcal{B}$ and redefine

$$\mathbb{P}([a,b]) = \int_a^b f(x)\,\mathrm{d}x \tag{5.14}$$

Proof. This is measure theory so the professor refused. □

§5.4 Random Variables

A random variable is a function on the sample space Ω which we want to measure.

Definition 5.7 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is a real-valued function

$$R : \Omega \to \mathbb{R} \tag{5.15}$$

such that for $a, b \in \mathbb{R}$ with $a \leq b$, the set $\{\omega \in \Omega \mid a \leq R(\omega) \leq b\}$ is an element of $\mathcal{F}$. This is equivalent to the inverse image of $R^{-1}([a,b])$.

Proposition 5.8 If $\mathcal{F} = 2^\Omega$, then every function $R : \Omega \to \mathbb{R}$ is a random variable.

Example 5.9 Flip a coin 6 times. Ω is all of the possible outcomes. The function

$$R = \text{number of heads} \tag{5.16}$$

is a random variable. Then, the probability

$$\mathbb{P}(0 \leq R \leq 1) = \frac{1}{2^6} + \frac{6}{2^6} \tag{5.17}$$

Example 5.10 Roll two dice. Ω is all of the possible outcomes. So

$$\Omega = \{(i, j) \mid i, j \in \{1, \dots, 6\}\} \tag{5.18}$$

Define a random variable $R = i + j$. We can compute

$$\mathbb{P}(0 \leq R \leq 3) = 0 + 0 + 1/36 + 2/36 = 1/12 \tag{5.19}$$

Then, a less trivial example:

Example 5.11 Take $\Omega = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}$ and probability measure on function $f(x)$. Define

$$R : \mathbb{R} \to \mathbb{R} \qquad R(x) = x + 1 \tag{5.20}$$

Then,

$$R^{-1}([a, b]) = [a - 1, b - 1] \tag{5.21}$$

which is obviously an element of $\mathcal{B}$.

Most functions (and most continuous ones) are random variables.

§6 October 4, 2024

§6.1 Random Variables

Take $(\Omega, \mathcal{F}, \mathbb{P})$ as a probability space. Recall

Definition 6.1 A random variable is some function

$$R : \Omega \to \mathbb{R} \tag{6.1}$$

such that $R^{-1}([a,b]) \in \mathcal{F}$. Note that

$$R^{-1}([a,b]) \tag{6.2}$$

describes all points $\chi \in \Omega$ such that $R(\chi)$ maps into $[a,b] \in \mathbb{R}$.

Definition 6.2 A random variables is some function

$$R : \Omega \to \mathbb{R} \tag{6.3}$$

such that $R^{-1}(B) \in \mathcal{F}$ for all Borel sets $B \in \mathcal{B}$ on $\mathbb{R}$.

Proposition 6.3 The above two definitions are equivalent.

Proof. Suppose R is a random variable as defined in Definition 6.1. Define

$$\mathcal{G} = \{G \subset \mathbb{R} \mid R^{-1}(G) \in \mathcal{F}\} \tag{6.4}$$

By Definition 6.1, all closed intervals are in $\mathcal{G}$. We now claim that $\mathcal{G}$ is a σ-algebra on $\mathbb{R}$. If this is true, then $\mathcal{G}$ being a σ-algebra containing all closed intervals means it contains all Borel sets $B \in \mathcal{B}$, i.e. $\mathcal{B} \subseteq \mathcal{G}$. This means Definition 6.2 $\subseteq$ Definition 6.1.

Lemma 6.4 **The Claim** $\mathcal{G}$ is a σ-algebra on $\mathbb{R}$

Proof of Lemma 6.4.

$$R^{-1}(\mathbb{R}) = \Omega \in \mathcal{F} \implies \mathbb{R} \in \mathcal{G} \tag{6.5}$$

$$G_i \in \mathcal{G} \implies R^{-1}(G_i) \in \mathcal{F} \implies \bigcup_i R^{-1}(G_i) \in \mathcal{F} \tag{6.6}$$

$$\implies R^{-1}\left(\bigcup_i G_i\right) \in \mathcal{F} \implies \left(\bigcup_i G_i\right) \in \mathcal{G} \tag{6.7}$$

$$G \in \mathcal{G} \implies R^{-1}(G) \in \mathcal{F} \implies \left(R^{-1}(G)\right)^C \in \mathcal{F} \implies G^C \in G \tag{6.8}$$

□

On the converse, if we have a random variable as defined in Definition 6.2, then that implies that

$$R^{-1}([a,b]) \in \mathcal{F} \tag{6.9}$$

for all $a \leq b$ because all closed intervals are Borel sets. Then that trivially proves the definition, i.e. Definition 6.1 $\subseteq$ Definition 6.2. Thus,

$$\text{Definition 6.1} \equiv \text{Definition 6.2}$$ □

§6.2 Discrete Random Variables

Take some probability space. Take some random variable

$$R : \Omega \to \mathbb{R} \tag{6.10}$$

Definition 6.5 A random variable $R : \Omega \to \mathbb{R}$ is discrete if $\mathrm{Im}(R)$ is a finite or countably infinite set of points, i.e. R hits a countable number of points.

Example 6.6 Flip a coin six times. Ω is the set of all possible outcomes. Take $R : \Omega \to \mathbb{N}$ as defined as the number of heads. The image of R is

$$\kappa := \{0, \dots, 6\} \tag{6.11}$$

which is discrete.

Define a probability function $\mathbb{P}_R(x)$ as

$$\mathbb{P}(R = x) \tag{6.12}$$

Then, $\mathbb{P}(R \in B)$ where $B \in \mathcal{B}$ is the sum

$$\sum_{x \in B} \mathbb{P}_R(x) \tag{6.13}$$

Define a distribution function $F_R(x)$ as

$$F_R(x) = \mathbb{P}(\{R \leq x\}) = \sum_{t \leq x} \mathbb{P}_R(t) \tag{6.14}$$

In Example 6.6,

$$\mathbb{P}_R(k \in \kappa) = \binom{6}{k} \left(\frac{1}{2}\right)^6 \tag{6.15}$$

For $k \notin \kappa$, $\mathbb{P}(k) = 0$. The distribution function is then

$$F_R(x) = \sum_{t=0}^{x} \mathbb{P}_R(t)\, \mathrm{d}t \tag{6.16}$$

§6.3 Absolutely Continuous Random Variables

Take a probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P})$. Define

$$\mathbb{P}(B \in \mathcal{B}) = \int_B f(x)\,\mathrm{d}x \tag{6.17}$$

such that f is a given probability density function which is **non-negative**, **integrable**, and

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1 \tag{6.18}$$

Define a random variable $R(\omega) \equiv \omega$. As in the same way we defined a distribution function,

$$F_R(x) \equiv \mathbb{P}(R \leq x) \tag{6.19}$$

Based on the integral above,

$$F_R(x) = \int_{-\infty}^{x} f(\omega)\,\mathrm{d}\omega \tag{6.20}$$

Definition 6.7 Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $R : \Omega \to \mathbb{R}$. R is absolutely continuous if there exists an integrable density function $f_R \geq 0$ such that $F_R(x) = \int_{-\infty}^{x} f_R(t)\,\mathrm{d}t$

Proposition 6.8 Let R be absolutely continuous with a density function f_R. Then,

$$\mathbb{P}(R \in (a, b]) = \int_a^b f_R(t)\,\mathrm{d}t \tag{6.21}$$

Proof. Trivial □

This is equivalent to

$$\mathbb{P}(R \in B) = \int_B f(x)\,\mathrm{d}x \tag{6.22}$$

for Borel set $B \in \mathcal{B}$.

Lemma 6.9 For an **absolutely continuous** random variable R with density function f_R, $\mathbb{P}(R = \chi)$ for some fixed value χ equals zero. This does not mean χ *never* occurs, but the probability is zero.

Proof. Trivial □

Example 6.10 Uniform Distribution Take probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ where $\mathbb{P}$ is some density function.

$$\mathbb{P}(B) = \int_B f(x)\,\mathrm{d}x \tag{6.23}$$

such that the density function is defined as

$$f_R(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{else} \end{cases} \tag{6.24}$$

for some constants a, b. This has uniform probability for a subset of $\mathbb{R}$. Then, the cumulative distribution function equals

$$F_R(x) = \mathbb{P}(R \leq x) = \begin{cases} 0 & x < a \\ \int_a^x \frac{1}{b-a}\,\mathrm{d}t = \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases} \tag{6.25}$$

§7 October 7, 2024

§7.1 Recap: Absolutely Continuous Random Variables

Given some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $R : \Omega \to \mathbb{R}$, recall the distribution function

$$F_R(x) = \mathbb{P}(R \leq x) \tag{7.1}$$

R is absolutely continuous if there exists a non-negative, integrable density function $f_R : \mathbb{R} \to \mathbb{R}$

$$F_R(x) = \int_{-\infty}^{x} f_R(x)\,\mathrm{d}x \tag{7.2}$$

In this case, then

$$\mathbb{P}(R \in B) = \int_B f_R(x)\,\mathrm{d}x \tag{7.3}$$

Some remarks:

1. If R is a absolutely continuous random variable, then F_R is continuous.

   Hand Waving. We need to show that

   $$\lim_{x \to a} F_R(x) = F_R(a) \tag{7.4}$$

   But, $F_R(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, so

   $$\lim_{x \to a} F_R(x) = \underbrace{\cdots\cdots}_{\text{measure theory}} = F_R(a) \tag{7.5}$$

   $\varepsilon - \delta$ was not done. $\square$

2. If the density function f_R is continuous, then the fundamental theorem of calculus yields

   $$\frac{\mathrm{d}F_R(x)}{\mathrm{d}x} = f_R(x) \tag{7.6}$$

   That is, the (probability) density function is the derivative of the (cumulative) distribution function.

3. "Let R be an absolutely continuous random variable with density function $f(x)$" means that we can construct R using the probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ where

   $$\mathbb{P}(B) = \int_B f(x)\,\mathrm{d}x \tag{7.7}$$

   using the given density function $f(x)$. Then, R is a random variable with density f. The density and distribution functions well-define R. We don't care about $R(\Omega)$ as much as f_R and F_R.

§7.2 Functions of Random Variables

Let R_1 be a random variable uniformly distributed on $[0,1]$, i.e.

$$f_1(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{else} \end{cases} \tag{7.8}$$

Define R_2 as

$$R_2 = (R_1)^2 \tag{7.9}$$

Then, we want to find $F_2(x)$. First, note

$$F_1(x) = \mathbb{P}(R_1 \le x) = \begin{cases} 0 & x < 0 \\ x & x \in [0,1] \\ 1 & x > 1 \end{cases} \tag{7.10}$$

To find the distribution function $(F_2)_{R_2}$,

$$F_2(u) = \mathbb{P}(R_2 \le u) = \mathbb{P}(R_1^2 \le u) \tag{7.11}$$

If $u \ge 0$, then

$$F_2(u) = \mathbb{P}\big(R_1 \in [-\sqrt{u}, \sqrt{u}]\big) \tag{7.12}$$

$$= \int_{-\sqrt{u}}^{\sqrt{u}} f_1(t)\,\mathrm{d}t \tag{7.13}$$

$$= \begin{cases} 0 & u < 0 \\ \sqrt{u} & 0 \le u \le 1 \\ 1 & \text{else} \end{cases} \tag{7.14}$$

The derivative, i.e. the density function, equals

$$f(u) = \begin{cases} 0 & u < 0 \\ \frac{1}{2\sqrt{u}} & u \in [0,1] \\ 0 & u > 1 \end{cases} \tag{7.15}$$

Example 7.1 A slightly trickier example. Take R_1 with density function

$$f_1(x) = \begin{cases} 0 & x < 0 \\ e^{-x} & x \ge 0 \end{cases} \tag{7.16}$$

and R_2 as

$$R_2 = \begin{cases} R_1 & R_1 \le 1 \\ 1/R_1 & R_1 > 1 \end{cases} \tag{7.17}$$

What is cdf(R_2) (or F_2)?

Solution. We can split into ranges. When $y \leq 0$,

$$F_2(y) = \mathbb{P}(R_2 \leq y) = \int_{(-\infty,y]} f_1(x)\,\mathrm{d}x = 0 \tag{7.18}$$

For $y > 0$,

$$F_2(y) = \mathbb{P}(R_2 \leq y) \tag{7.19}$$

Now, there are two cases: one where $y \leq 1$ and one where $y > 1$.

$$\mathbb{P}(R_2 \leq y \quad \text{and} \quad R_1 \leq 1) \quad \text{and} \quad \mathbb{P}(R_2 \leq y \quad \text{and} \quad R_1 > 1) \tag{7.20}$$

So,

$$\begin{aligned} F_2(y \mid y > 0) &= \mathbb{P}(R_1 \leq y \text{ and } R_1 \leq 1) + \mathbb{P}\left(\frac{1}{R_1} \leq y \text{ and } R_1 > 1\right) && (7.21) \\ &= F_1(y) + \mathbb{P}(R_1 > 1/y) && (7.22) \\ &= \left[1 - e^{-y}\right] + [1 - \mathbb{P}(R_1 \leq 1/y)] && (7.23) \\ &= \cdots = 1 - e^{-y} + e^{-1/y} && (7.24) \end{aligned}$$

As $y \to 1$, $F_2(y) \to 1$. So, other than this case,

$$F_2(y) = \begin{cases} 1 & y > 1 \\ 0 & y < 0 \end{cases}$$

□

§7.3 Properties of Distributions Functions

Take $(\Omega, \mathcal{F}, \mathbb{P})$ as some probability space and R as a random variable. Take $F(x)$ as

$$\mathbb{P}(R \leq x) \tag{7.25}$$

Then,

Proposition 7.2

1. Let $A_1, A_2, \ldots \in \mathcal{F}$ be an expanding sequence (i.e. $A_m \supseteq A_n$ for all $m > n$). Then,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{7.26}$$

2. Let $A_1, A_2, \ldots \in \mathcal{F}$ be a contracting sequence (i.e. $A_m \subseteq A_n$ for all $m > n$). Then,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n) \tag{7.27}$$

Proof of the first. Take $A = \bigcup_n A_n$. Then,

$$A = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \cdots \tag{7.28}$$

These are clearly disjoint. It follows that

$$\mathbb{P}(A) = \sum_{i=1} \mathbb{P}(A_i - A_{i-1}) \tag{7.29}$$

where $A_0 \equiv \emptyset$. This equals

$$\mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=1}^{\infty} [\mathbb{P}(A_{i+1}) - \mathbb{P}(A_i)] \tag{7.30}$$

But there are cascading cancellations, so

$$\mathbb{P}(A) = \mathbb{P}(A_1) + \lim_{n\to\infty} (\mathbb{P}(A_{n+1}) - \mathbb{P}(A_1)) \tag{7.31}$$

Thus,

$$\mathbb{P}(A) = \lim_{n\to\infty} A_{n+1\equiv n} \tag{7.32}$$

□

Some properties of distribution functions

1. F is non decreasing, i.e. if $a < b$ then $F(a) \leq F(b)$.

   Proof.

   $$a < b \implies \{R \leq a\} \subset \{R \leq b\} \tag{7.33}$$

   □

2. Infinite limit

   $$\lim_{x\to\infty} F(x) = 1 \tag{7.34}$$

   For the absolutely continuous case this is trivial. Let x_n be a sequence of real numbers with $x_n \to \infty$. We want to show that

   $$\lim_{n\to\infty} F(x_n) = 1 \tag{7.35}$$

   This is equivalent to showing that $F(x \to \infty) \to 1$.

   Proof. Define $A_n = \{R \leq x_n\}$. Then,

   $$F(x_n) = \mathbb{P}(R \leq x_n) = \mathbb{P}(A_n) \tag{7.36}$$

   But, this is an expanding sequence because $A_1 \subset A_2 \subset \cdots$. Thus, applying Proposition 7.2,

   $$F\left(x_n \xrightarrow{n\to\infty} \infty\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}(\Omega) = 1 \tag{7.37}$$

   □

§8 October 9, 2024

§8.1 Properties of Distribution Functions

Take some $(\Omega, \mathcal{F}, \mathbb{P})$ probability space, R random variable, and $F(x) = \mathbb{P}(R \leq x)$ distribution function. Recall Lemma 7.2,

$$A_1 \subset A_2 \subset \cdots \implies \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \tag{8.1}$$

and a similar Lemma for contracting sets. Now, we can continue proving some properties

1. F non decreasing. Proven last time.
2. $\lim_{x\to\infty} F(x) = 1$. Proven last time.
3. $\lim_{x\to-\infty} F(x) = 0$

   Proof. Let $x_n \to -\infty$. For every sequence of real numbers tending to $-\infty$,

   $$\lim_{n\to\infty} F(x_n) = 0 \tag{8.2}$$

   Then, define $A_n \equiv \{R \leq x_n\}$. Then, the sets A_n are contracting, and we can apply the Lemma.

   $$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) \tag{8.3}$$

   Because A_n are contracting as $x \to -\infty$, this means $A_n \to \emptyset$, and therefore

   $$\lim_{n\to\infty} F(x_n) \to F(\emptyset) = 0 \tag{8.4}$$

   □

4. Limit probability from the right

   $$\lim_{x\to x_0^+} F(x) = F(x_0) \tag{8.5}$$

   Proof. Let x_n be a monotonically decreasing sequence of real numbers such that $x_n > x_0$ but $x_n \to x_0$. Let

   $$A_n = \{R \leq x_n\} \tag{8.6}$$

The sets A_n are contracting because x_n are decreasing on $x \to x_0^+$. Then,

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) \tag{8.7}$$

The intersection of all of these is $(-\infty, x_0)$. If $\omega \in A_n$ for all n, that implies $R(\omega) \leq x_n$ for all n. As $x_n \to x_0$, this means

$$R(\omega) \leq x_0 \tag{8.8}$$

for all ω. Conversely, □

5. Limit probability from the left

$$\lim_{x\to x_0^-} F(x) = \mathbb{P}(R < x_0) \tag{8.9}$$

(this is not the same as $\leq x_0$).

Proof. The proof is pragmatically the same as above. Let x_n be a monotonically increasing sequence. Then, set A_n as an increasing sequence. As $n \to \infty$, $F(x_n) \to \mathbb{P}(R < x_0)$.

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \tag{8.10}$$

We want to show that

$$\bigcup_n A_n \equiv \{R < x_0\} \tag{8.11}$$

For this to be true, pick some $\omega \in \bigcup_n A_n$; that means ω must be in any of them. All A_n are some $\{R < x_n\}$. But, $x_n \to x_0$ implies that $\omega < x_0 \implies \omega \in A_n$ for some n. Since the limit x_n does not equal x_0, this effectively means

$$\bigcup_n A_n \equiv \{R < x_0\} \tag{8.12}$$

This can be done in reverse too to complete the $\iff$ proof, but that proof is quite trivial. □

6. Probability of equality

$$\mathbb{P}(R = x_0) = F(x_0^+) - F(x_0^-) \tag{8.13}$$

Proof. Trivially,

$$\mathbb{P}(R = x_0) = \mathbb{P}(R \leq x_0) - \mathbb{P}(R < x_0) \tag{8.14}$$

By previous properties, this equation is pretty obviously equal to Equation 8.13. □

Example 8.1 Pick some discrete probability space where R is the value of a fair die. Then, the probability

$$\mathbb{P}(R = 2) = \mathbb{P}(R \leq 2) - \mathbb{P}(R < 2) = \frac{2}{6} - \frac{1}{6} = \frac{1}{6} \tag{8.15}$$

and the rest of the distribution looks like a staircase for obvious reasons. This also satisfies the other properties, which is trivial.

§8.2 Joint Density Functions

Definition 8.2 Take some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and random variables R_1, R_2. We say the pair (R_1, R_2) is absolutely continuous if there exists some integrable function $f_{12}(x, y) > 0$ such that the joint distribution function

$$F(x, y) = \mathbb{P}(R_1 \leq x \cap R_2 \leq y) = \iint_{(-\infty,-\infty)}^{(x,y)} f_{12}(x, y) \, \mathrm{d}A \tag{8.16}$$

Then, f_{12} is called the density of the pair (R_1, R_2) or the joint density of R_1 **and** R_2.

8.2.1 Borel Sets Tangent

Borel sets can be done on $\mathbb{R}^2$.

Definition 8.3 On $\mathbb{R}^2$, the Borel σ-algebra $\mathcal{B}$ is the σ-algebra of subsets of $\mathbb{R}^2$ generated by rectangles

$$[x_1, x_2] \times [y_1, y_2] \tag{8.17}$$

where rectangle bounds can also have open-interval bounds.

This definition can be generalized to $\mathbb{R}^n$.

Proposition 8.4 Take $\Omega \equiv \mathbb{R}^n$. If f is non-negative and integrable on $\mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1 \tag{8.18}$$

then if $\mathcal{F} \equiv \mathcal{B}^n$, then there exists a unique probability measure $\mathbb{P}$ on $\mathcal{B}^n$ such that

$$\mathbb{P}([x_1^i, x_1^f] \times \cdots \times [x_n^i, x_n^f]) = \int_{\mathbb{R}^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{8.19}$$

Lemma 8.5 A ball is an element of $\mathcal{B}^3$; a disk is an element of $\mathcal{B}^2$.

There were a few more examples, but they did not say anything new.[1]

Example 8.6 Let $f(x, y)$, $R_1 \equiv x$, $R_2 \equiv y$. Set

$$f(x, y) = \begin{cases} 1 & 0 \leq x \leq 1 \cup 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases} \tag{8.20}$$

Let R_1, R_2 have joint density f. Find

$$\mathbb{P}(2R_1 \leq R_2) \tag{8.21}$$

That just equals

$$\mathbb{P}(2x \leq y) \tag{8.22}$$

and finding this can be done with a very simple double integral (or just finding the area of a triangle).

[1] "Once you see some examples it becomes more tractable" but it already is???

§9 October 11, 2024

§9.1 Recap: Joint Distribution Functions

Recall that for R_1, R_2 random variables, we say that a pair (R_1, R_2) is absolutely continuous if there exists a joint density function $f_{12}(x, y)$ such that

$$F(x, y) \equiv \mathbb{P}(R_1 \le x \cap R_2 \le y) = \iint_{\mathbb{R}^2}^{[x,y]} f_{12}(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{9.1}$$

This can be rather trivially extrapolated to $\mathbb{R}^N$.

Example 9.1 Suppose (R_1, R_2) has density

$$f_{12}(x, y) = \begin{cases} 2e^{-x-2y} & 0 < x, y < \infty \\ 0 & \text{else} \end{cases} \tag{9.2}$$

Find $\mathbb{P}(R_1 > 1 \cap R_2 < 1)$.

Solution. We can simply integrate the density over this space

$$\hat{x} = [1, \infty) \cap \hat{y} = (0, 1] \tag{9.3}$$

noting that $y < 0$ has density 0. Thus, we integrate over $\hat{x} \times \hat{y}$

$$\mathbb{P} = \iint_{\hat{x} \times \hat{y}} 2e^{-x-2y}\, \mathrm{d}(\hat{x} \times \hat{y}) \tag{9.4}$$

Pretty trivial from here. □

Again, this can be trivially extrapolated to N random variables. Treat

$$\mathbf{R} = (R_1, \ldots, R_N) \tag{9.5}$$
$$\mathbf{x} = (x_1, \ldots, x_N) \tag{9.6}$$

as a vector of random variables. Then,

$$F_{12\cdots N}(\mathbf{x}) \iint_{\mathbb{R}^N} f_{12\cdots N}\, \mathrm{d}\mathbf{x} \equiv \mathbb{P}\left[R_1 \le x_1, \ldots, R_N \le x_N\right] \tag{9.7}$$

Proposition 9.2 Given F_{12}, we can find f_{12} (assuming continuous) by differentiating

$$\frac{\partial}{\partial x} \frac{\partial}{\partial y} F_{12} \tag{9.8}$$

Recall that taking partial derivatives is a commutative operation.

§9.2 Joint vs Individual Density Functions

Proposition 9.3 If (R_1, R_2) is a absolutely continuous pair, then R_1 and R_2 are absolutely continuous.

Proof. Let (R_1, R_2) have density $f_{12}(x, y)$. Without loss of generality,

$$F_1(x) \equiv \mathbb{P}(R_1 \leq x) \tag{9.9}$$

But this just $\mathbb{P}(R_1 \leq x \cap R_2 \in \mathbb{R})$, which just equals

$$\int_{-\infty}^{x} \int_{\mathbb{R}} f_{12} \, \mathrm{d}v \, \mathrm{d}u \tag{9.10}$$

The inside can be written as

$$f_1(u) \equiv \int_{\mathbb{R}} f_{12} \, \mathrm{d}v \tag{9.11}$$

Thus, R_1 (and symmetrically, R_2) are absolutely continuous. □

Example 9.4 Suppose (R_1, R_2) has density

$$f_{12}(x, y) = \begin{cases} x + y & 0 \leq x, y \leq 1 \\ 0 & \text{else} \end{cases} \tag{9.12}$$

Then computing f_1 and f_2 are quite trivial using Proposition 9.3.

Proposition 9.5 The converse of Proposition 9.3 is *not* true. Given absolutely continuous random variables R_1, R_2, it is not necessarily true that the pair (R_1, R_2) is absolutely continuous. Moreover, there is not necessarily a unique f_{12} given f_1, f_2.

Proof. There can be integration constants that make the map $R_1, R_2 \to (R_1, R_2)$ not injective. □

§9.3 Independent Random Variables

Recall that events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{9.13}$$

Definition 9.6 Take R_1, R_2 random variables. We say R_1 and R_2 are independent if

$$\mathbb{P}(R_1 \in B_1, R_2 \in B_2) = \mathbb{P}(R_1 \in B_1)\mathbb{P}(R_2 \in B_2) \tag{9.14}$$

where $B_1, B_2 \in \mathcal{B}$ are Borel sets.

This can be extrapolated to a family of random variables, even uncountably many.

Definition 9.7 A family $\mathcal{R} = \{R_i\}_{i \in I}$ (where I is some index set) is independent if for every finite subset $\{R_{i_1}, \ldots, R_{i_k}\} \subset \mathcal{R}$,

$$\mathbb{P}(R_{i_1} \in B_{i_1}, \ldots, R_{i_k} \in B_{i_k}) = \prod_j \mathbb{P}(R_{i_j} \in B_{i_j}) \tag{9.15}$$

for all $B_{\ldots} \in \mathcal{B}$.

Proposition 9.8 Let $R_1, \ldots, R_n$ be independent and individually absolutely continuous with densities $f_1, \ldots, f_n$. Then, the joint random variable $(R_1, \ldots, R_n)$ is absolutely continuous with density

$$f_{1\cdots n} \equiv \prod_i^n f_i(x_i) \tag{9.16}$$

Proof. By definition

$$F_{12\cdots n}(x_1, \ldots, x_n) = \mathbb{P}(R_1 \le x_1 \cap \cdots \cap R_n \le x_n) \tag{9.17}$$

Independence tells us this equals

$$F_{12\cdots n}(x_1, \ldots, x_n) = \mathbb{P}(R_1 \le x_1) \cdots \mathbb{P}(R_n \le x_n) \tag{9.18}$$

and those can trivially be written as

$$F_{12\cdots n}(x_1, \ldots, x_n) = F_1(x_1) \cdots F_n(x_n) \tag{9.19}$$

Then,

$$\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} F_{12\cdots n} = f_{12\cdots n} = f_1(x_1) \cdots f_n(x_n) \tag{9.20}$$

because when taking $\frac{\partial}{\partial x_j}$, all $f_{k \neq j}$ are treated as constants. □

§9.4 Functions of a Random Variable

Definition 9.9 A function

$$g : \mathbb{R} \to \mathbb{R} \tag{9.21}$$

is Borel measurable if the inverse image

$$\forall B \in \mathcal{B}, g^{-1}(B) \in \mathcal{B} \tag{9.22}$$

Proposition 9.10 Let R be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be Borel measurable. Then, $g(R)$ is a random variable.

Proof. Recall a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ is some map

$$R : \Omega \to \mathbb{R} \tag{9.23}$$

such that for all $B \in \mathcal{B}$

$$R^{-1}(B) \in \mathcal{F} \tag{9.24}$$

Let $B \in \mathcal{B}$. Write

$$\widetilde{R} \equiv g(R) : \Omega \to \mathbb{R} \tag{9.25}$$

Then,

$$\begin{aligned} \widetilde{R}^{-1}(B) &= \{\omega \in \Omega \mid g(R(\omega)) \in B\} & (9.26) \\ &= \{\omega \in \Omega \mid R(\omega) \in g^{-1}(B)\} & (9.27) \\ &= R^{-1}(g^{-1}(B)) \in \mathcal{F} & (9.28) \end{aligned}$$

□

Proposition 9.11 Let $R_1, \ldots, R_n$ be independent random variables and $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ be Borel measurable functions[a]. Then,

$$g_1(R_1), \ldots, g_n(R_n) \tag{9.29}$$

are independent, i.e. functions of independent random variables are independent.

[a]anything continuous is Borel measurable

Proof. Let $\widetilde{R_i} \equiv g_i(R_i)$. Let $B_1, \ldots B_n \in \mathcal{B}$. Then

$$\begin{aligned} \mathbb{P}(\widetilde{R_1} \in B_1, \ldots, \widetilde{R_n} \in B_n) &= \mathbb{P}(g_1(R_1) \in B_1, \ldots, g_n(R_n) \in B_n) & (9.30) \\ &= \mathbb{P}(R_1 \in g_1^{-1}(B_1), \ldots, R_n \in g_n^{-1}(B_n)) & (9.31) \end{aligned}$$

By independence of $\{R_i\}$, this equals

$$\begin{aligned} &= \mathbb{P}(R_1 \in g_1^{-1}(B_1)) \cdots \mathbb{P}(R_n \in g_n^{-1}(B_n)) & (9.32) \\ &= \mathbb{P}(g_1(R_1) \in B_1) \cdots \mathbb{P}(g_n(R_n) \in B_n) & (9.33) \end{aligned}$$

which means $\widetilde{R_1}, \ldots, \widetilde{R_n}$ are independent.[2] □

[2]a LiTtLe BiT ThEoReTiCaL ToDaY

§10 October 14, 2024

§10.1 Functions of More Than One Random Variable

Example 10.1 Take R_1, R_2 as independent, uniformly distributed random variables between 0 and 1. Find the density of $R_3 = \frac{R_2}{R_1^2}$

Solution. The density of the pair (R_1, R_2) is

$$f_{12}(x,y) = \begin{cases} 1 & 0 \leq x, y \leq 1 \\ 0 & \text{else} \end{cases} \tag{10.1}$$

which is the uniform square $[0,1] \times [0,1]$. We can compute the distribution function for R_3:

$$F_3(z) = \mathbb{P}(R_3 \leq z) \tag{10.2}$$

This just equals

$$F_3(z) = \mathbb{P}\big(R_2 \leq (R_1)^2 z\big) \tag{10.3}$$

Note that we do not care about $R_1 = 0$ because it has zero probability. This equals

$$\iint_{y \leq x^2 z} f_{12}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{10.4}$$

We can take $z \geq 0$ because negative quotients are not possible. Then, the cases $z \in [0,1]$ and $z > 1$ are separate in evaluating this double integral.

1. Take $z \in [0,1]$. Then, we are integrating

$$\int_0^1 x^2 z\,\mathrm{d}x = \frac{z}{3} \tag{10.5}$$

2. Take $z > 1$. Then, we are integrating

$$\int_0^{\sqrt{1/z}} x^2 z\,\mathrm{d}z + \int_{\sqrt{1/z}}^1 1\,\mathrm{d}z \tag{10.6}$$

which equals

$$\int_0^{\sqrt{1/z}} x^2 z\,\mathrm{d}z + \left(1 - \sqrt{1/z}\right) \tag{10.7}$$

which trivially equals

$$1 - \frac{2}{3\sqrt{z}} \tag{10.8}$$

Thus, the distribution function equals

$$F_3(z) = \begin{cases} 0 & z < 0 \\ \frac{z}{3} & z \in [0,1] \\ 1 - \frac{2}{3\sqrt{z}} & z > 1 \end{cases} \tag{10.9}$$

The density is the derivative

$$f_3(z) = \begin{cases} 0 & z < 0 \\ \frac{1}{3} & z \in [0,1] \\ \frac{1}{3}z^{-3/2} & z > 1 \end{cases} \tag{10.10}$$

□

Definition 10.2 A random variable R_1 with density

$$f(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{10.11}$$

where μ is the mean and $\sigma > 0$ is the standard deviation of a Gaussian/normal distribution.

Everybody knows what a normal distribution, its mean, and its standard deviation are.

Example 10.3 Let R_1, R_2 be independent normal distributions with $\mu = 0, \sigma = 1$. Find the density of $R_3 = \sqrt{R_1^2 + R_2^2}$

Solution. For $z > 0$,

$$F_3(z) = \mathbb{P}(R_3 \leq z) \tag{10.12}$$

This is the probability that $R_1^2 + R_2^2 \leq z^2$. The joint density of (R_1, R_2) is the product, i.e.

$$f_{12} = \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} \tag{10.13}$$

Then,

$$\begin{align} F_3(z) &= \iint_{x^2+y^2 \leq z^2} \frac{1}{2\pi} e^{\frac{-(x^2+y^2)}{2}} \, \mathrm{d}x \, \mathrm{d}y \tag{10.14} \\ &= \int_0^{2\pi} \int_0^z \frac{1}{2\pi} e^{\frac{-r^2}{2}} r \, \mathrm{d}r \, \mathrm{d}\theta \tag{10.15} \\ &= \int_0^z e^{\frac{-r^2}{2}} r \, \mathrm{d}r \tag{10.16} \end{align}$$

It's pretty trivial to compute this, but we want the density. The inside of this integral *is* the density. □

§10.2 Poisson Distribution

Recall the Binomial distribution. Suppose we have n Bernoulli trials, for each of which p is the probability of success. Let R be the discrete random variable where R is the number of successes after n trials. Obviously,

$$\mathbb{P}_R(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{10.17}$$

What happens when we have many, many trials? Suppose n is very large and p is very small. Also assume that $np \to \lambda$ where $\lambda > 0$ is some positive constant. ($p \sim \lambda/n$)

Proposition 10.4 This distribution converges to the discrete **Poisson Distribution** with probability function

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \tag{10.18}$$

Proof. Write R_n for this random variable. Using

$$\mathbb{P}(R_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{10.19}$$

As $n \to \infty$,

$$\begin{aligned}
\mathbb{P}(R_n = k) &= \frac{n(n-1)\cdots(n-k+1)}{k!}(np)^k \frac{1}{n^k}\left(1 - \frac{np}{n}\right)^{n-k} && (10.20)\\
&= \frac{1(1-1/n)\cdots(1-(k-1)/n)}{k!}\lambda^k \left(1 - \frac{np}{n}\right)^n \left(1 - \frac{np}{n}\right)^{-k} && (10.21)\\
&= \frac{1}{k!}\lambda^k \left(1 - \frac{np}{n}\right)^n \left(1 - \frac{np}{n}\right)^{-k} && (10.22)
\end{aligned}$$

But, $\left(1 - \frac{np}{n}\right)^{-k} \to 1$ because $np \to \lambda$, and $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$. Thus, our R_n approaches

$$e^{-\lambda} \frac{\lambda^k}{k!} \tag{10.23}$$

□

Proposition 10.5 The Poisson Distribution is well defined, i.e.

$$\sum_0^\infty e^{-\lambda} \frac{\lambda^k}{k!} \, \mathrm{d}k = 1 \tag{10.24}$$

Proof. This just equals

$$e^{-\lambda} \sum_{0}^{\infty} \frac{\lambda^k}{k!} \tag{10.25}$$

which equals

$$e^{-\lambda} e^{\lambda} \tag{10.26}$$

which is obviously 1. □

Example 10.6 Poisson Distributions can model the number of typos on the page of a printed book.

Example 10.7 Poisson Distributions can model the number of radioactive emissions / alpha particles in an hour.

10.2.1 Sum of Poisson Distributions

Proposition 10.8 Let R_1, R_2 be two Poisson distributions with parameters λ_1, λ_2. Then, the random variable

$$R' = R_1 + R_2 \tag{10.27}$$

is a Poisson distribution with parameter $\lambda' = \lambda_1 + \lambda_2$.

Lemma 10.9 Let $R_1, R_2, \ldots, R_n$ be independent, discrete random variables with probability functions $P_1, P_2, \ldots, P_n$. Write $P_{12\cdots n}$ for the joint probability function of the n random variables, i.e.

$$P_{12\cdots n}(x_1, x_2, \ldots, x_n) = \mathbb{P}(R_1 = x_1, \ldots) \tag{10.28}$$

Then, $R_1, R_2, \ldots, R_n$ are independent if and only if

$$\mathbb{P}(R_1 = x_1, \ldots, R_n = x_n) = \mathbb{P}(R_1 = x_2) \cdots \mathbb{P}(R_n = x_n) \tag{10.29}$$

Proof of Proposition. The joint probability function $f_{12}(i, j)$ equals

$$\begin{aligned} f_{12}(i, j) &= \mathbb{P}(R_1 = i \quad \text{and} \quad R_2 = j) &(10.30)\\ &= \mathbb{P}(R_1 = i) \cdot \mathbb{P}(R_2 = j) &(10.31)\end{aligned}$$

That is just the product of the two Poisson distributions

$$e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^k}{k!} \tag{10.32}$$

Then,

$$\mathbb{P}(R_1 + R_2 = k) = \sum_{i=0}^{k} \mathbb{P}(R_1 = i)\mathbb{P}(R_2 = k - i) \tag{10.33}$$

which equals

$$\sum_{i=0}^{k} \frac{1}{i!}\lambda_1^i e^{-\lambda_1} \frac{1}{(k-i)!}\lambda_2^{k-i} e^{-\lambda_2} \tag{10.34}$$

After multiplying by $k!/k!$, we get

$$\sum_{i=0}^{k} \binom{k}{i} \frac{1}{k!}\lambda_1^i \lambda_2^{k-i} e^{-(\lambda_1+\lambda_2)} \tag{10.35}$$

which simplifies into

$$\frac{(\lambda_1 + \lambda_2)^k}{k!} e^{-(\lambda_1+\lambda_2)} \tag{10.36}$$

which is clearly a Poisson Distribution with parameter $\lambda_1 + \lambda_2$. $\square$

§11 October 16, 2024

'new' stuff lol

§11.1 Expectation

11.1.1 Discrete Expectation

Throw a fair six-sided die. Let R be the result. What is the expected value of R? Obviously, in this case, $R = 3.5$. What if R is any *simple* random variable?

Definition 11.1 A **simple random variable** takes at most *finitely* many values, such as rolling a dice with 6 possible values.

Definition 11.2 Let R be a simple random value. Then, the expectation of R equals

$$E(R) = \sum_x x \cdot \mathbb{P}(x) \tag{11.1}$$

the weighted average of the probabilities.

Example 11.3 Take a biased coin

$$\mathbb{P}(\text{heads}) = \frac{3}{4} \qquad \mathbb{P}(\text{tails}) = \frac{1}{4} \tag{11.2}$$

and flip it twice. Define $R \equiv$ number of heads. What is $E(R)$?

Solution. $E(R)$ equals

$$\sum_x x \cdot \mathbb{P}(x) \tag{11.3}$$

which equals

$$0 \cdot \mathbb{P}(0) + 1 \cdot (2(3/4)(1/4)) + 2 \cdot (9/16) = \frac{3}{2} \tag{11.4}$$

□

Definition 11.4 Let R be a simple random variable, and let

$$g : \mathbb{R} \to \mathbb{R} \tag{11.5}$$

Then, the expectation of this is

$$E(g(R)) = \sum_x g(x) \cdot \mathbb{P}(x) \tag{11.6}$$

Definition 11.5 Let R be a discrete random variable (it could have countably infinitely many values). Then, (again)

$$E(g(R)) = \sum_x g(x)\mathbb{P}(x) \tag{11.7}$$

This could be an infinite sum, which indicates a possibility of divergence. Thus, we define Equation 11.7 as long as

1. $g \geq 0$ (in this case, $E(R)$ could diverge)
2. or the sum is absolutely convergent

11.1.2 Absolutely Continuous Expectation

Definition 11.6 Let R be an absolutely continuous random variable with density $f_R(x)$. Then, define the expectation

$$E(R) = \int_{\mathbb{R}} x \cdot f_R(x) \, \mathrm{d}x \tag{11.8}$$

Additionally, if g is a Borel function, define

$$E(g(R)) = \int_{\mathbb{R}} g(x) f_R(x) \, \mathrm{d}x \tag{11.9}$$

as long as

1. $g(x) \geq 0$
2. or $\int \ldots$ is absolutely convergent

Lemma 11.7 This above definition can be extended to finitely many random variables $R_1, \ldots, R_n$. Take $R_1, \ldots, R_n$ discrete. Then,

$$E(g(R_1, \ldots, R_n)) = \sum_{x_1, \ldots, x_n} g(x_1, \ldots, x_n) \cdot \mathbb{P}(R_1 = x_1, \ldots, R_n = x_n) \tag{11.10}$$

This same thing can be done to absolutely continuous random variables:

$$E(g(R_1, \ldots, R_n)) = \int_{\mathbb{R}^n} g(x_1, \ldots, x_n) f_{12\cdots n}(x_1, \ldots x_n) \, \mathrm{d}\mathbf{x}_n \tag{11.11}$$

where g is some Borel function and we make similar assumptions as above.

Example 11.8 Let R be a random variable with density

$$f_R(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{else} \end{cases} \tag{11.12}$$

Fine $E(R)$.

Solution. Since f is zero on $x < 0$, the expectation equals

$$E(R) = \int_0^\infty x \cdot e^{-x} \, \mathrm{d}x \tag{11.13}$$

which is a fairly trivial integration by parts. It equals

$$E(R) = \cdots = 1 \tag{11.14}$$

□

We were then shown *another* example.

Example 11.9 Let R_1, R_2 be independent random variables each with density

$$f_R(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{else} \end{cases} \tag{11.15}$$

Find $E(\max(R_1, R_2))$

Solution. First, the joint density is just the product, i.e.

$$f_{12}(x, y) = \begin{cases} e^{-x-y} & (x \geq 0) \cap (y \geq 0) \\ 0 & \text{else} \end{cases} \tag{11.16}$$

Using Equation 11.11, we get

$$E(\max(R_1, R_2)) = \iint_{\mathbb{R}_+^2} \max(x, y) \cdot e^{-x-y} \, \mathrm{d}\mathbf{x}_2 \tag{11.17}$$

$$= \int_0^\infty \int_0^\infty \max(x, y) \cdot e^{-x-y} \, \mathrm{d}x \, \mathrm{d}y \tag{11.18}$$

Using some inequalities, the $\max(\cdots)$ becomes

$$\max(x, y) = \begin{cases} x & x \geq y \\ y & x \leq y \end{cases} \tag{11.19}$$

So we can write our integral as

$$\underbrace{\int_0^\infty \int_0^x x \cdot e^{-x-y} \, \mathrm{d}y \, \mathrm{d}x}_{1} + \underbrace{\int_0^\infty \int_x^\infty y \cdot e^{-x-y} \, \mathrm{d}y \, \mathrm{d}x}_{2} \tag{11.20}$$

which can be quite trivially integrated by parts. The answer is

$$E(\max(\cdots)) = \frac{3}{2} \tag{11.21}$$

$\square$

§11.2 Moments

Definition 11.10 Let R be a random variable. Then, for $k > 0$ (k is not necessarily an integer). Then, the k-th moment of R is defined as

$$\alpha_k = E(R^k) \tag{11.22}$$

Note that

$$\alpha_1 = E(R) = \mu \qquad \text{mean} \tag{11.23}$$

Let R be an absolutely continuous random variable with density $f(x)$. Consider the centroid of the mass defined by $0 \leq y \leq f(x)$ (the "center of mass"), located at (x_c, y_c). Then

$$x_c = \alpha_1(R) = E(R) = \mu_R \tag{11.24}$$

Given N points

$$(x_1, y_1), \cdots, (x_N, y_N) \tag{11.25}$$

the centroid is the point

$$\frac{1}{N}\sum_{i=1}^{N}(x_i, y_i) \tag{11.26}$$

The same thing is true for infinitely many points:

$$\frac{\iint (x, y)\,\mathrm{d}x\,\mathrm{d}y}{\iint_{\mathbb{R}^2} 1\,\mathrm{d}x\,\mathrm{d}y} \tag{11.27}$$

In the case of some random mass (possibly a nose, which we can assume to be spherical), the denominator may be complicated and annoying. In the case of an absolutely continuous random variable, that denominator just equals 1. So, for our random variable,

$$\iint (x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{11.28}$$

In the specific case of a shaded region under some function $f(x)$, this is

$$\int_{\mathbb{R}}\int_0^{f(x)} (x, y)\,\mathrm{d}y\,\mathrm{d}x \tag{11.29}$$

11.2.1 Central Moments

Definition 11.11 Let R be a random variable; let $k > 0$ be positive. Then, the k-th central moment of R is

$$\beta_k = E\big[(R-\mu)^k\big] \tag{11.30}$$

such that the mean is zero. In particular,

$$\beta_1 = E(R-\mu) = 0 \tag{11.31}$$

Then, β_2 is the variance, which is the square of standard deviation.

Example 11.12 Take $R = N(\cdots)$, so

$$f_R(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{11.32}$$

Then, using the first moment and second central moment,

$$\alpha_1 = \mu \qquad\qquad \beta_2 = \sigma^2 \tag{11.33}$$

Solution. I've done this twice already and will not do it a third time. □

§12 October 18, 2024

§12.1 Recap: Random Variables

Recall that a random variable R on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function

$$R : \Omega \to \mathbb{R} \tag{12.1}$$

Example 12.1 Flip one coin. Then

$$R = \begin{cases} 1 & \text{heads} \\ 0 & \text{tails} \end{cases} \tag{12.2}$$

Then, the probability distribution is

$$F_R(x) = \mathbb{P}(R \le x) = \begin{cases} 0 & x < 0 \\ 1/2 & x \in [0, 1) \\ 1 & x \ge 1 \end{cases} \tag{12.3}$$

This can also be thought of as $\Omega = \{H, T\}$, and

$$R(\Omega) = \{H \to 1, T \to 0\} \tag{12.4}$$

Example 12.2 Roll a dice. Let

$$R = \begin{cases} 1 & \text{even} \\ 0 & \text{odd} \end{cases} \tag{12.5}$$

Then, R can also be expressed as

$$R = \{1 \to 0, 2 \to 1, \cdots\} \tag{12.6}$$

The probability distribution is the same as in Equation 12.3.

Take R as a random variable with a density $f(x)$. Note that the probability space here is not specified, because there are many random variables with the same density f. **Choose** the probability space $(\mathbb{R}, \mathcal{B}, \mathbb{P})$ such that

$$\mathbb{P}(B \in \mathcal{B}) = \int_B f(x)\,\mathrm{d}x \tag{12.7}$$

Then we can define

$$R : \mathbb{R} \to \mathbb{R} \equiv R(x) = x \tag{12.8}$$

Then the probability

$$\mathbb{P}(R \leq z) = \int_{x \leq z} f(x)\,\mathrm{d}x \tag{12.9}$$

> **Note** If a random variable has a density function, it is assumed to be absolutely continuous.

Let $R_1, \ldots, R_n$ be independent, absolutely continuous random variables with densities $f_1(x), \ldots, f_n(x)$. Take the probability space $(\mathbb{R}^n, \mathcal{B}, \mathbb{P})$ where

$$\mathbb{P}(B) = \int \cdots \int_B f_1(x_1) \cdots f_n(x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{12.10}$$

Then,

$$R_1(x_1, \ldots, x_n) = x_1 \tag{12.11}$$

$$R_n(x_1, \ldots, x_n) = x_n \tag{12.12}$$

§12.2 Median

Definition 12.3 Let R be a random variable with distribution function $F_R(x)$. Assume absolute continuity. Then, $F_R(x)$ is continuous. Then, the median is defined as

$$m \in \mathbb{R} \mid f_R(m) = \frac{1}{2} \tag{12.13}$$

which is equivalent to

$$\mathbb{P}(R \leq m) = \frac{1}{2} = \mathbb{P}(R > m) \tag{12.14}$$

For general, not-necessarily-absolutely-continuous R,

Definition 12.4 The median $m \in \mathbb{R}$ of a random variable R satisfies

$$F_R(m) \geq 1/2 \qquad F_R(m^-) \leq 1/2 \tag{12.15}$$

Example 12.5 Take Example 12.1. The median is any number $m \in [0, 1]$.

Proof. Let $x \in [0, 1]$. Then,

$$F_R(x) \geq 1/2 \quad \text{and} \quad F_R(x^-) \leq 1/2$$

□

§12.3 Properties of Expectation

Note Assume expectations all exist and are finite, i.e. do not diverge.

1. Closure under addition. Let $R_1, \ldots, R_n$ be random variables. Then,

$$E\left(\sum_i R_i\right) = \sum_i E(R_i) \tag{12.16}$$

2. For $a \in \mathbb{R}$,

$$E(aR) = a \cdot E(R) \tag{12.17}$$

3. If $R_1 \leq R_2$ for all $\omega \in \Omega$, then $E(R_1) \leq E(R_2)$.
4. If $R \geq 0$ and $E(R) = 0$, then $\mathbb{P}(R = 0) = 1$, i.e. R is essentially zero.

Proof. It's in the textbook. $\square$

Note If for some random variable R, $\mathrm{Var}(R)$, then R is essentially constant.

5. Let $R_1, \ldots, R_n$ be independent random variables. Then,

$$E(R_1 \cdots R_n) = E(R_1) \cdots E(R_n) \tag{12.18}$$

6. Let R be a random variable with mean μ and variance σ^2. Then,

$$\mathrm{var}(aR + b) = a^2\sigma^2 \tag{12.19}$$

Proof. It's in the textbook. $\square$

7. Variance is closed under addition, i.e. for independent random variables $R_1, \ldots, R_n$,

$$\mathrm{var}\left(\sum_i R_i\right) = \sum_i \mathrm{var}(R_i) \tag{12.20}$$

Proof. It's in the textbook. $\square$

§13 October 23, 2024

§13.1 Review of Moments

Let X be a random variable. Recall that

$$\alpha_k = E(X^k) \qquad k > 0 \tag{13.1}$$

is the k-th moment. It follows that $E(X) = \text{mean}(X)$, so recall that

$$\beta_k = E((X-m)^k) \qquad k > 0 \tag{13.2}$$

is the k-th central moment (for $m < \infty$).

§13.2 Properties of Expectation Continued

8. The central moments $\beta_1, \beta_2, \ldots$ of a random variable can be obtained from the moments $\alpha_1, \alpha_2, \ldots$, assuming $\alpha_i < \infty$ for $i < n$ and α_n exists. The result is

$$\beta_n = E[(R-m)^n] = E\left[\sum_{k=0}^{n} \binom{n}{k} (-m)^{n-k} R^k\right] \tag{13.3}$$

assuming $m < \infty$. This simplifies into

$$\sum_{k=0}^{n} \binom{n}{k} (-m)^{n-k} \alpha_k \tag{13.4}$$

because $E(R^k) \equiv \alpha_k$. It follows that for $n = 2$,

$$\text{var}(R) = E(R^2) - 2mE(R) + m^2 \implies \boxed{\sigma_R^2 = E(R^2) - [E(R)]^2} \tag{13.5}$$

9. If $0 < j < k$, then

$$E(|R|^j) \le 1 + E(|R|^k) \tag{13.6}$$

Proof. For $\omega \in \Omega$,

$$|R(\omega)|^j \le \begin{cases} |R(\omega)|^k & |R(\omega)| \ge 1 \\ 1 & \text{else} \end{cases} \tag{13.7}$$

i.e. finiteness is a logical consequence of $0 < j < k$. Thus,

$$|R(\omega)|^j \le 1 + |R(\omega)|^k \tag{13.8}$$

for all $\omega \in \Omega$. This extends to the expectation:

$$E(|R|^j) \le 1 + E(|R|^k) \tag{13.9}$$

□

It follows that if some higher order expectation is finite, then lower order expectations are also finite.

Example 13.1 Let $R \sim \exp(\lambda)$ be

$$f_R(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases} \tag{13.10}$$

Find α_i and $\operatorname{var}(R)$.

Solution. By definition,

$$\alpha_j = E(R^j) = \int_0^\infty x^j \lambda e^{-\lambda x} \, \mathrm{d}x \tag{13.11}$$

This integration could clearly be done by parts. But this can also be done through a substitution. Set $y = \lambda x$. Then, $\mathrm{d}y = \lambda \, \mathrm{d}x$. We get

$$\frac{1}{\lambda^j} \cdot \int_0^\infty y^j \cdot e^{-y} \, \mathrm{d}y \tag{13.12}$$

Note Define the Gamma function

$$\Gamma(j) \equiv \int_0^\infty y^{j-1} \cdot e^{-y} \, \mathrm{d}y \tag{13.13}$$

We can do some induction on j:

$$\Gamma(1) = \int_0^\infty e^{-x} = 1 \tag{13.14}$$

For $\Gamma(r)$,

$$\int_0^\infty x^{r-1} e^{-x} \, \mathrm{d}x \tag{13.15}$$

By parts, this becomes

$$\cancel{\frac{x^r}{r} e^{-x} \Big|_0^\infty} + \frac{1}{r} \int_0^\infty x^r e^{-x} \, \mathrm{d}x \tag{13.16}$$

This indicates

$$\Gamma(r) = \frac{1}{r} \Gamma(r+1) \tag{13.17}$$

Inductively, for $n \in \mathbb{N}$, $\Gamma(n+1) = n!$.

It follows that Equation 13.12 simplifies into

$$a_j = \frac{j!}{\lambda^j} \tag{13.18}$$

So it follows that

$$\text{var}(\exp(\lambda)) = \alpha_2 - \alpha_1^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \boxed{\frac{1}{\lambda^2}} \qquad \square$$

§13.3 Covariance

With multiple random variables, we can define covariance for two variables at a time. Let R_1, R_2 be two random variables.

Definition 13.2 The (i, j)-th joint moment for the pair of random variables is

$$\alpha_{i,j} = E\left[R_1^i \cdot R_2^j\right] \qquad i, j \geq 1 \tag{13.19}$$

Assume that the moments are finite, i.e. $m_1 = E(R_1) < \infty$ and $m_2 = E(R_2) < \infty$.

Definition 13.3 The (i, j)-th joint central moment for the pair of random variables is

$$\beta_{i,j} = E\left[(R_1 - m_1)^i \cdot (R_2 - m_2)^j\right] \tag{13.20}$$

Definition 13.4 The **covariance** of a pair of random variables R_1, R_2 is defined as

$$\beta_{1,1} = \text{cov} \tag{13.21}$$

Note that if $R_1 = R_2$, then

$$\text{cov}(R_1, R_2) = \text{var}(R_1) \tag{13.22}$$

Proposition 13.5 An easier way to compute covariance is given by

$$\begin{aligned}
\text{cov}(R_1, R_2) &= E(R_1R_2 - R_2m_1 - R_1m_2 + m_1m_2) && (13.23)\\
&= E(R_1R_2) - m_1 \cdot E(R_2) - m_2 \cdot E(R_1) + m_1m_2 && (13.24)\\
&= E(R_1R_2) - m_1m_2 - m_1m_2 + m_1m_2 && (13.25)\\
&= E(R_1R_2) - m_1m_2 && (13.26)
\end{aligned}$$

§13.4 Preview of Correlation

Covariance is not normalized and can take any value $r \in \mathbb{R}$. However, that may be unintuitive, so we want a normalized correlation constant $R \in [-1, 1]$

Proposition 13.6 If R_1, R_2 are independent, then $\text{cov}(R_1, R_2) = 0$ and the two variables are uncorrelated.

Proof. R_1, R_2 are independent, so

$$\begin{aligned} E[(R_1 - m_1)(R_2 - m_2)] &= E[(R_1 - m_1)] \cdot E[(R_2 - m_2)] \qquad (13.27) \\ &= 0 \cdot 0 = 0 \end{aligned}$$ □

This is not an iff; functions may be uncorrelated, but not independent.

This will be completed in Lecture 14.

§14 October 25, 2024

§14.1 Cauchy Schwarz

Theorem 14.1 For an inner product space V, let $u, v \in V$; then

$$\|\langle u, v\rangle\| \leq \langle u, u\rangle \cdot \langle v, v\rangle \tag{14.1}$$

Proof. Straightforward from the definition of inner product. □

Proposition 14.2 For random variables R_1, R_2 (assume $E(R_1^2), E(R_2^2)$ are finite), then

$$[E(R_1 R_2)]^2 \leq E(R_1^2) \cdot E(R_2^2) \tag{14.2}$$

with equality if and only if there exist $a_1, a_2 \in \mathbb{R} \neq 0$ such that

$$a_1 R_1 + a_2 R_2 = 0 \tag{14.3}$$

i.e. R_1 and R_2 are linearly dependent, which is equivalent to saying that the probability that this linearly combined random variable equalling 0 is 1.

Proof. Suppose then that R_1, R_2 are linearly dependent, which implies

$$a_1 R_1 + a_2 R_2 = 0 \tag{14.4}$$

Then,

$$[E(R_1 R_2)]^2 = \left(E\left(-\frac{a_2}{a_1} R_2^2\right)\right)^2 = \left(\frac{a_2}{a_1}\right)^2 \left(E\left(R_2^2\right)\right)^2 \tag{14.5}$$

$$E(R_1^2) \cdot E(R_2^2) = E\left(\left(-\frac{a_2}{a_1} R_2^2\right)^2\right) \cdot E(R_2^2) = \left(\frac{a_2}{a_1}\right)^2 \left(E\left(R_2^2\right)\right)^2 \tag{14.6}$$

Now, define

$$S = aR_1 - R_2 \tag{14.7}$$

Then,

$$0 \leq E(S^2) = E(a^2 R_1^2 + R_2^2 - 2aR_1 R_2) \tag{14.8}$$

$$\leq a^2 E(R_1^2) + E(R_2^2) - 2aE(R_1 R_2) \tag{14.9}$$

Pick $a = E(R_1 R_2)/E(R_1^2)$. Then, plugging in a and rearranging yields

$$[E(R_1 R_2)]^2 \leq E(R_1^2) \cdot E(R_2^2)$$ □

Tracing back, when Equation 14.7 is zero, R_1 and R_2 are linearly dependent, and S is essentially zero.

§14.2 Correlation Coefficient

Definition 14.3 Recall that

$$\text{cov}(R_1 R_2) = E((R_1 - E(R_1))(R_2 - E(R_2))) \tag{14.10}$$

Define the correlation coefficient of R_1 with R_2 to be

$$\rho(R_1, R_2) = \text{corr}(R_1, R_2) \equiv \frac{\text{cov}(R_1, R_2)}{\sigma_1 \sigma_2} \tag{14.11}$$

where $\sigma_1, \sigma_2 \in \mathbb{R} > 0$ are the standard deviation of R_1 and R_2, and $\text{cov}(\cdots)$ is assumed to exist.

Proposition 14.4 Assume $E(R_1^2), E(R_2^2) < \infty$. Then,

$$|\rho| \leq 1 \tag{14.12}$$

with equality if and only if $R_1 - E(R_1)$ and $R_2 - E(R_2)$ are linearly dependent.

Proof. By definition,

$$(\text{cov}(R_1, R_2))^2 = [E((R_1 - E(R_1))(R_2 - E(R_2)))]^2 \tag{14.13}$$

Cauchy Schwarz tells us that

$$[E((R_1 - E(R_1))(R_2 - E(R_2)))]^2 \leq E(R_1 - E(R_1)) \cdot E(R_2 - E(R_2))$$

Thus,

$$\left| \frac{(\text{cov}(R_1, R_2))^2}{\sigma_1 \sigma_2} \right| \leq 1 \qquad \square$$

The equality holds when $R_1 - E(R_1)$ and $R_2 - E(R_2)$ are linearly dependent as a consequence of Proposition 14.2.

§14.3 Method of Indicators

Take some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and some $A \in \mathcal{F}$.

Definition 14.5 The *indicator of* A is the discrete random variable I_A given by

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases} \tag{14.14}$$

which are analogous to characteristic functions in analysis.

Proposition 14.6 The expectation of an indicator is

$$E(I_A) = \mathbb{P}(I_A = 1) = \mathbb{P}(A) = \mathbb{P}(\{\omega \in \Omega \mid I_A(\omega) = 1\}) \tag{14.15}$$

The expectation of an indicator equals the probability of the event.

Example 14.7 Take R is the number of successes in n Bernoulli trials where the probability of success is p. Find $E(R)$ and $\mathrm{var}(R)$.

Solution. Define A_i as the event where the i-th trial is successful. Then,

$$R = \sum_j (I_{A_j}) = \text{number of successes} \tag{14.16}$$

The expectation of R is the sum of expectations of each indicator, i.e.

$$E(R) = \sum_j E(I_{A_j}) = \sum_j \mathbb{P}(A_j) = np \qquad \square$$

The variance then equals

$$E(R^2) - E(R)^2 \tag{14.17}$$

which equals $E(R^2) - (np)^2$. That first term equals

$$E\Big[(I_{A_1} + \cdots I_{A_n})^2\Big] \tag{14.18}$$

which equals

$$E\left[\sum_j I_{A_j}^2 + \sum_{j \neq k} I_{A_j} I_{A_k}\right] \tag{14.19}$$

Note For random variable A,

$$I_A^2 = I_A \tag{14.20}$$

Proof. $1^2 = 1$ and $0^2 = 0$ $\square$

Note If A and B are independent events, then I_A and I_B are independent.

Thus, $E(R^2)$ becomes

$$= \sum_j E\big[I_{A_j}\big] + \sum_{j \neq k} E\big[I_{A_j} I_{A_k}\big] \tag{14.21}$$

$$= \sum_j E\big[I_{A_j}\big] + \sum_{j \neq k} E\big[I_{A_j \cap A_k}\big] \tag{14.22}$$

$$= np + (n^2 - n)p^2 \tag{14.23}$$

Thus,

$$\operatorname{var}(R) = \left[np + (n^2 - n)p^2\right] - (np)^2 = np(1-p) \tag{14.24}$$

Example 14.8 Suppose N people throw their hat into the middle of a room then randomly each pick a hat. Find $E(R)$ where R is the number of people who pick their own hat.

Solution. Define A_i as the i-th person getting their own hat. Then,

$$R = I_{A_1} + \cdots + I_{A_N} \tag{14.25}$$

Note that these are not independent events! Thus,

$$E(R) = E(I_{A_1}) + \cdots + E(I_{A_N}) \tag{14.26}$$

For each person,

$$P(A_i) = 1/N \tag{14.27}$$

Then,

$$E(R) = N \cdot 1/N = 1$$

□

§15 October 28, 2024

§15.1 Chebyshev's Inequality

Stating this theorem is harder than proving it.

Theorem 15.1

a. $R \geq 0$ r.v.; $b > 0$ random number

$$\mathbb{P}(R \geq b) \leq E(R)/b \qquad \text{assume expectation exists} \tag{15.1}$$

b. R r.v.; c constant, $l, \varepsilon > 0$ constants

$$\mathbb{P}(|R - c| \geq \varepsilon) \leq \frac{E(|R - c|)^l}{\varepsilon^l} \tag{15.2}$$

c. R r.v. finite mean m, finite variance $\sigma^2 > 0$, $k > 0$

$$\mathbb{P}(|R - m| \geq k\sigma) \leq 1/k^2 \tag{15.3}$$

15.1.1 Proof of (a)

[a] First, assuming R is absolutely continuous with density $f(x)$. Then,

$$E(R) = \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x \tag{15.4}$$

Since $R \geq 0$, this is equivalent to

$$E(R) = \int_0^{\infty} x f(x) \, \mathrm{d}x \tag{15.5}$$

We can shove an inequality at this

$$E(R) = \int_0^{\infty} x f(x) \, \mathrm{d}x \geq \int_b^{\infty} x f(x) \, \mathrm{d}x \tag{15.6}$$

Note that we are throwing away $\int_0^b$, so this is not a very sharp estimate. Note that within this integral, $x \geq b$, so this previous line is greater than

$$E(R) \geq \int_b^{\infty} x f(x) \, \mathrm{d}x \geq b \int_b^{\infty} f(x) \, \mathrm{d}x = b\mathbb{P}(R \geq b) \tag{15.7}$$

Rearranging,

$$\mathbb{P}(R \geq b) \leq E(R)/b \tag{15.8}$$

For the general case, we claim

$$R \geq bI_{A_b} \tag{15.9}$$

where $A_b \equiv \{R \geq b\} \subset \Omega$. Recall that I is the indicator function, i.e.

$$I_{A_b}(\omega) = \begin{cases} 1 & \omega \in A_b \\ 0 & \omega \notin A_b \end{cases} \tag{15.10}$$

To prove this, let $\omega \in \Omega$. We want $R(\omega) \geq bI_{A_b}(\omega)$. We can plsplit into cases

1. Case 1: $\omega \notin A_b$. Then, we want to show that $R(\omega) \geq 0$, which is true by assumption that $R \geq 0$.
2. Case 2: $\omega \in A_b$. Then $bI_{A_b}(\omega) = b$. We need to show that $R(\omega) \geq b$. This is true by how A_b is defined, i.e. as $\{R \geq b\}$.

This proves the claim, i.e. for all ω, this statement is true:

$$R \geq bI_{A_b} \tag{15.11}$$

This implies that $E(R) \geq E(bI_{A_b})$. By linearity,

$$E(R) \geq bE(I_{A_b}) = b\mathbb{P}(A_b) = b\mathbb{P}(R \geq b) \tag{15.12}$$

which can be rearranged to complete the proof for the general, non-absolutely-continuous case. □

15.1.2 Proof of (b)

This follows from A. The probability

$$\mathbb{P}(|R-c| \geq \varepsilon) = \mathbb{P}(|R-c|^l \geq \varepsilon^l) \tag{15.13}$$

Then we apply (a) where $|R-c|$ is our new, non-negative random variable. Thus,

$$\mathbb{P}(|R-c|^l \geq \varepsilon^l) \leq \frac{E\Big(|R-c|^l\Big)}{\varepsilon^l} \tag{15.14}$$

□

15.1.3 Proof of (c)

The probability $\mathbb{P}(|R-m| \geq k\sigma)$, by (b),

$$\mathbb{P}(|R-m| \geq k\sigma) \leq \frac{E(|R-m|^2)}{(k\sigma)^2} = \frac{\sigma^2}{k^2\sigma^2} = 1/k^2 \tag{15.15}$$

□

Note This is not a very sharp inequality. Take $R \sim \exp\{1\}$ (mean and variance are both 1). By (c), (assume k large)

$$\mathbb{P}(|R-m| \geq k\sigma) \leq 1/k^2 \tag{15.16}$$

For our specific distribution, $m = \sigma = 1$, so we get

$$\mathbb{P}(|R-m| \geq k\sigma) = \mathbb{P}(|R-1| \geq k) = \mathbb{P}(R \geq k+1) = e^{-(k+1)} \tag{15.17}$$

For k large,

$$e^{-(k+1)} \ll 1/k^2 \tag{15.18}$$

so the inequality is not sharp, and quite wasteful.

§15.2 Weak Law of Large Numbers

Theorem 15.2 Let $R_1, R_2, \ldots$ be a whole bunch of independent random variables on a given probability space. Assume the means and variances are finite, and shared (equal) across all $R_{1\cdots N}$. And assume $\sigma_i^2 \leq M$, i.e. the variances are bounded by some constant $M \in \mathbb{R}$ for all i. Define

$$S \equiv \sum_i R_i \tag{15.19}$$

and let n be the number of random variables being summed over. Then, for all $\varepsilon > 0$,

$$\mathbb{P}\left[\frac{|S - E(S)|}{n} \geq \varepsilon\right] \xrightarrow{n \to \infty} 0 \tag{15.20}$$

or

$$\mathbb{P}\left[\left|\frac{R_1 + \cdots + R_n}{n} - m\right| \geq \varepsilon\right] \xrightarrow{n \to \infty} 0 \tag{15.21}$$

i.e. the mean of the random variables tends towards the actual mean as $n \to \infty$

Example 15.3 Flip N coins. As $N \to \infty$, the expected number of heads is $N/2$. As $N \to \infty$, the probability of deviation from the expected value tends towards zero.

Solution. Obvious. □

Why is this law of large numbers *weak*? Because it is stating that the probability of being greater than ε goes to zero, as opposed to an actual probability going to zero.

15.2.1 Proof

Take

$$\mathbb{P}\left(\frac{|S - E(S)|}{n} \geq \varepsilon\right) \tag{15.22}$$

By Part (b) of Chebyshev, we can write

$$\begin{aligned}
\mathbb{P}\left(\frac{|S - E(S)|}{n} \geq \varepsilon\right) &\leq \frac{E\left(\left|\frac{S-E(S)}{n}\right|^2\right)}{\varepsilon^2} && (15.23)\\
&\leq \frac{1}{n^2\varepsilon^2}E\left[(S - E(S))^2\right] && (15.24)\\
&\leq \frac{\mathrm{var}(S)}{n^2\varepsilon^2} && (15.25)
\end{aligned}$$

By linearity on the independent variables,

$$\begin{aligned}
\mathbb{P}\left(\frac{|S - E(S)|}{n} \geq \varepsilon\right) &\leq \frac{\sum_i^n \mathrm{var}(R_i)}{n^2\varepsilon^2} && (15.26)\\
&\leq \frac{Mn}{n^2\varepsilon^2} = \frac{M}{n\varepsilon^2} && (15.27)\\
&\leq 0 && (15.28)\\
\implies \mathbb{P}\left(\frac{|S - E(S)|}{n} \geq \varepsilon\right) &= 0 && (15.29)
\end{aligned}$$

§15.3 Conditional Probability

Let R_1, R_2 be discrete r.v. with probability functions $P_{R_1}(x)$ and $P_{R_2}(x)$ which are 'probability mass functions'. What is

$$\mathbb{P}(R_1 = x \mid R_2 = y) \tag{15.30}$$

The probability $p(x \mid y)$ equals

$$\frac{\mathbb{P}(R_1 = x, R_2 = y)}{\mathbb{P}(R_2 = y)} = p_{12}(x, y)/p_2(y) \tag{15.31}$$

which is the quotient of the joint probability function and the individual probability function.

Note For a given, fixed y, if $p_2(y) > 0$, then

$$p(x \mid y) \tag{15.32}$$

is a probability function for x, i.e.

$$\sum_x p(x \mid y) = 1 \tag{15.33}$$

which is true by extending the law of total probability.