AMATH 561A: Probability and Random Processes

Marlin Figgins

Oct. 1, 2020

Introduction

This is a collection of my notes taken during AMATH 561A during fall quarter 2020 at the University of Washington.

Table of Contents

1. Probability Spaces and Random Variables

Definition 1.1 (Probability Space: Version A)**.** A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where Ω is the *sample space* or the space of possible outcomes, $\mathcal{F}$ is the set of events which are subsets of Ω and $\mathbb{P} \colon \mathcal{F} \to [0,1]$ is a probability measure that assigns probabilities to events.

Example 1.2 (Rolling a fair standard die)**.** In this case, we have six possible outcomes as our die is six-sided. Therefore, our sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}. \tag{1.1}$$

The set of events in this case are all possible subsets of Ω i.e. the power set $\mathcal{P}(\Omega)$. Finally, since we specified that the die is fair, each outcome is equally likely, so our probability measure $\mathbb{P}$ is uniform over Ω. Therefore, our probability of a given event is just the fraction of our possible out comes which are in the event.

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{6} \text{ for all } A \in \mathcal{F}. \tag{1.2}$$

For example, we would say that the probability of rolling an even number would be given by

$$\mathbb{P}(\{2, 4, 6\}) = \frac{3}{6} = 0.5. \tag{1.3}$$

In this example, it is simple enough to use all possible subets are our set of events, but as we attempt to deal with more complicated probability spaces, we'll need to develop a more rigorous notion of what constitutes an event, so that our probability measures will have certain properties of interest. This idea is formalized by the σ-algebra.

Definition 1.3 (σ-algebra)**.** A non-empty collection of subsets of Ω is a σ-*algebra* of Ω if it satistifies the following:

(i) For each event $A \in \mathcal{F}$, then its complement $A^c \in \mathcal{F}$

(ii) If we have a countable sequence of sets $A_i \in \mathcal{F}$, then their union $\bigcup_i A_i \in \mathcal{F}$.

These two conditions form the statement that a σ-algebra is closed under complements and countable unions.

Proposition 1.4. *The definition of the σ-algebra implies that σ-algebras are also closed under countable intersection because*

$$\bigcap_i A_i = \left(\bigcup_i A_i^c \right)^c. \tag{1.4}$$

Minimally, we know that a σ-algebra must contain the empty set and the space on which it lives.

Exercise 1.5. Show that any σ-algebra of Ω contains both the entire space Ω and the empty set.

This formalizes the notion of events and combinations of events espressed earlier. Now, we'll develop a formal notion of probabiltiy usinng measure theory.
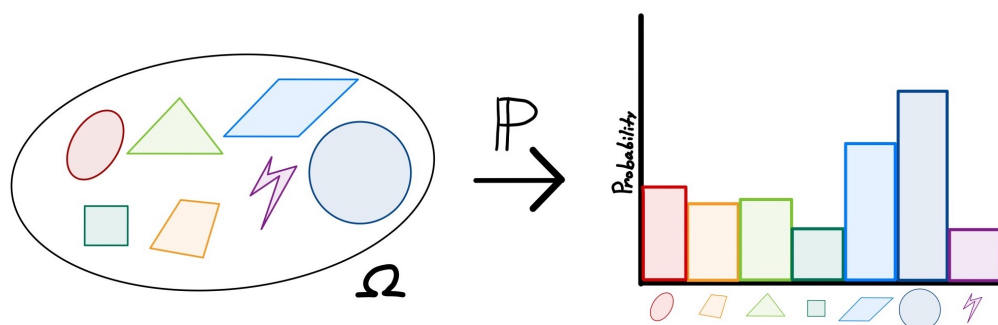
When we have a probability space Ω and a σ-algebra $\mathcal{F}$ on Ω, we can define a measure on the space $(\Omega, \mathcal{F})$. *Describe what a measure is intuitively.*

Definition 1.6 (Measure)**.** A *measure* is a function $\mu : \mathcal{F} \to [0, \infty)$ which satisfies:

(i) $\mu(A) \geq \mu(\emptyset) = 0$ for all $A \in \mathcal{F}$

(ii) if $A_i \in F$ is a countable sequence of disjoint sets, then

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i). \tag{1.5}$$

In the case that $\mu(\Omega)$, we will call μ a *probability measure* and denote it as $\mathbb{P}$.

Marlin Figgins (2020)
Twitter: @marlinfiggins

Figure 1: A probability measure $\mathbb{P}$ takes measurable subsets $E \in \mathcal{F}$ to a numerical value in $[0, 1]$.

Our definition of measure allows us to ensure that our notion of probabilty satisifies some of the intuitive properties one might expect of probabilities.
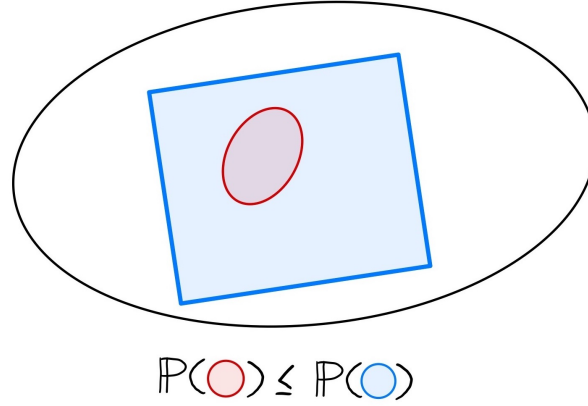
Theorem 1.7 (Properties of measure)**.**

(i) Monotonicity. *If $A \subset B$, then $\mu(A) \leq \mu(B)$.*

(ii) Subadditivity. *If $A \subset \bigcup_{j \in \mathbb{N}} A_j$, then*

$$\mu(A) \leq \sum_{j \in \mathbb{N}} \mu(A_m). \tag{1.6}$$

Proof of monotonicity. Finish proof. □

Proof of subadditivity. Finish proof. □

Figure 2: If a probability measure is consistent with our formulation, it ought to be monotonic. Appending additional outcomes to an event E should never decrease its probability of occuring.

Theorem 1.8 (Continuity of measure)**.**

(i) Continuity from above. *If we have a sequence of increasing subsets $A_1 \subset A_2 \subset \cdots$ such that $\bigcup_i A_i = A$, then*

$$\mu(A_i) \uparrow \mu(A) \text{ as } i \to \infty. \tag{1.7}$$

(ii) Continuity from below. *If we have a sequence of decreasing subsets $A_1 \supset A_2 \supset \cdots$ such that $\bigcup_i A_i = A$, then*

$$\mu(A_i) \downarrow \mu(A) \text{ as } i \to \infty. \tag{1.8}$$

Exercise 1.9. Prove that in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \text{ for any event } A \in \mathcal{F}. \tag{1.9}$$

1.1 Discrete Probability Spaces

Constructing a discrete probability space. We'll now focus on the case of discrete probability spaces. Let Ω be a countable set. That is, either finite or countably infinite. Also, let $\mathcal{F}$ be the set of all subsets of Ω which we will call the power set $\mathcal{P}(\Omega)$. We can know endow $(\Omega, \mathcal{F})$ with a probability measure.

Theorem 1.10. *Suppose we have a countable set Ω with a σ-algebra $\mathcal{F} = \mathcal{P}(\Omega)$. Then any function $p\colon \Omega \to [0, 1]$ such that*

$$\sum_{\omega \in \Omega} p(\omega) = 1 \tag{1.10}$$

induces a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ as follows. For any event $A \in \mathcal{F}$, we define the probability of A as

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega). \tag{1.11}$$

We call the function p the probability mass function *of $\mathbb{P}$.*

Example 1.11 (Repeated fair coins)**.** Consider the following experiment: We have a fair coin and we flip it until it lands on heads. We can write the possible outcomes as sequences of heads and tails, so that

$$\Omega = \{H, TH, TTH, TTH, \ldots\}. \tag{1.12}$$

As before, we can let $\mathcal{F} = \mathcal{P}(\Omega)$. Let's motivate our choice of probability mass function p. Assuming that the probability of heads and tails at every flip is $\frac{1}{2}$, then we have a probability of $\left(\frac{1}{2}\right)^n$ for having n consequective flips. This tells us that

$$p(\underbrace{T \cdots T}_{n \text{ tails}} H) = \left(\frac{1}{2}\right)^n. \tag{1.13}$$

We can check that this indeed sums to one:

$$\sum_{\omega \in \Omega} p(\omega) = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^n. \tag{1.14}$$

As shown above, this probability mass function induces a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. We can now use this measure the compute the probability of events like:

$$A = \{\text{Heads appears before third toss.}\} = \{H, TH\}, \tag{1.15}$$
$$B = \{\text{There are an even number of tails.}\} = \{H, TTH, TTTTH, \ldots\}. \tag{1.16}$$

We can then compute these probabilities as:

$$\mathbb{P}(A) = \frac{1}{2} + \frac{1}{4}, \tag{1.17}$$
$$\mathbb{P}(B) = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \cdots = \frac{2}{3}. \tag{1.18}$$

1.2   Uncountable probability spaces

We'll begin by further describing σ-algebra, so that we can understand what an uncountable σ-algebra must satisft.

Proposition 1.12. The intersection of σ-algebras is again a σ-algebra. *Let $I \neq \emptyset$ be an arbitrary set. If we have a collection of σ-algebras $\mathcal{F}_i$ for $i \in I$ defined on Ω, then their intersection*

$$\bigcap_{i \in I} \mathcal{F}_i \text{ is a } \sigma\text{-algebra on } \Omega. \tag{1.19}$$

Proof. Per an earlier exercise, all σ-algebras in the collection $\mathcal{F}_i$ must contain $\emptyset$ and Ω. Therefore, these sets are also in the intersection. For any event A that is in every $\mathcal{F}_i$, A^c must be in $\mathcal{F}_i$ since they are closed under complements. A similar argument holds for countable collections of events A_j and their unions. $\qquad\square$

Thus, if we are given a sample space Ω and $\mathcal{A}$ a collection of subsets of Ω, then there is a smallest σ-algebra which contains $\mathcal{A}$.

Definition 1.13 (σ-algebra generated by $\mathcal{A}$)**.** Given a sample space Ω and a collection of its subsets $\mathcal{A}$, the σ-algebra generated by $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}$. We denote this as $\sigma(\mathcal{A})$.

Example 1.14. Given $\Omega = \{1,2,3,4\}$ and $\mathcal{A} = \{\{1\}.\{1,2\}\}$, the smallest σ-algebra containing $\mathcal{A}$ is

$$\sigma\mathcal{A} = \{\emptyset, \Omega, \{1\}, \{2,3,4\}, \{1,2\}, \{3,4\}, \{1,3,4\}, \{2\}\} \tag{1.20}$$

Constructing uncountable probability spaces   Let's consider the experiment of drawing a random number between $[0,1]$. In this case, our sample space is clearly $\Omega = [0,1]$. In this case, we would like each number to be equally likely, but we can't simply set each number to occur with probability $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ as with the case of the fair die because our space is uncountable! Ideally, we want to find some function $p(x) = p$ which describes the probability that we choose $x \in [0,1]$. What is the probabiltiy that our chosen number is rational (in $\mathbb{Q}$). By the countable additivity of our probability measure,

$$\mathbb{P}(\mathbb{Q} \cap [0,1]) = \mathbb{P}\left(\bigcup_{x \in \mathbb{Q} \cap [0,1]} \{x\}\right) \tag{1.21}$$

$$= \sum_{x \in Q \cap [0,1]} p(x) = \sum_{x \in Q \cap [0,1]} p. \tag{1.22}$$

For any positive real value of p, the rightmost sum is infinite which is nonsensical as the probability of any event cannot exceed one. Thus the probability of an single value must be exactly 0. By the countable additivity of measure, this means that any countable subset of $[0,1]$ must have probablity 0.

The Lebesgue measure on $[0,1]$ We'll try an alternative approach to constructing probability spaces. We begin by defining the probabilities of certain subsets of interest and we extend this to the σ-algebra containing that collection of subsets. For example, we can specify that the probability of choosing a number between a and b is

$$\mathbb{P}((a,b]) = b - a \text{ for } a \text{ and } b \text{ with } 0 \le a < b \le 1. \tag{1.23}$$

This is a uniform probability measure on $[0,1]$ and it is called the *Lebesgue measure on* $[0,1]$.

Borel σ-algebra   We can extend the Lebesgue measure described below by extending it to the σ-algebra generated by intervals of the form $(a,b]$ by using the countable additivity property of measures. This is called the *Borel σ-algebra* on $[0,1]$ is denoted by $\mathcal{B}([0,1])$. $\mathcal{B}([0,1])$ is the smallest σ-algebra containing all open sets in $[0,1]$. Additionally, the extension of our measure to the Borel σ-algebra is unique by the Caratheodory Extension theorem. That being said, the Lebesgue measure cannot be extended to all subsets of $[0,1]$. That being said, we will restrict our attention to study $([0,1], \mathcal{B}([0,1]))$.

Example 1.15 (Infinite coin tosses). Suppose that we conduct an infinite sequence of coin tosses. Let the probability of each coin showing heads be p and the probability of tails be $q = 1 - p$. In this case, our sample space is all possible countable sequences of heads and tails. There is a one-to-one correspondence between Ω and $[0,1]$ by considering the base two expansion of numbers between $[0,1]$. As Ω is uncountable, we begin by specifying the probability on some reasonable subset of events which we can expand to some larger σ-algebra. We'll represent $w \in \Omega$ as a sequence $\omega_1 \omega_2 \ldots$ where each w_i is the result of the ith coin toss. Let $A_H = \{\omega \in \Omega \mid \omega_1 = H\}$. That is, A_H is the event that the first coin toss is heads. Naturally, the probability of A_H should be the probability of flipping heads, so that

$$\mathbb{P}(A_H) = p \text{ and similarly } \mathbb{P}(A_T) = q = 1 - p. \tag{1.24}$$

These events generate a σ-algebra $\mathcal{F}_1 = \sigma(\{A_H, A_T\})$. In fact, we can define a series of σ-algebras $\mathcal{F}_n$ based on the events of the first n coin tosses. For example, consider $\mathcal{F}_2$. This would contain events like:

$$A_{HH} = \{\omega \in \Omega \mid \omega_1 = H, \omega_2 = H\}, \quad A_{HT} = \{\omega \in \Omega \mid \omega_1 = H, \omega_2 = T\}, \tag{1.25}$$
$$A_{TH} = \{\omega \in \Omega \mid \omega_1 = T, \omega_2 = H\}, \quad A_{TT} = \{\omega \in \Omega \mid \omega_1 = T, \omega_2 = T\}. \tag{1.26}$$

These events would have probabilities:

$$\mathbb{P}(A_{HH}) = p^2, \quad \mathbb{P}(A_{HT}) = pq, \tag{1.27}$$
$$\mathbb{P}(A_{TH}) = pq, \quad \mathbb{P}(A_{TT}) = q^2. \tag{1.28}$$

We can similarly extend this to the events in $\mathcal{F}_n$. Due to our construction, we have an increasing sequence of σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \ldots$ which describe the results of finite

coin tosses. We'll then define $\mathcal{F}_\infty = \cup_{n\in\mathbb{N}}\mathcal{F}_n$. This is **not** a σ-algebra because it is not closed under countable intersection. Consider the event w_H which is the infinite sequence of outcomes with all H. We also consider the outcomes ω_{nH} which is where we count n heads in a row. As you can see,

$$\{\omega_H\} = \bigcap_{n\in\mathbb{N}}\{\omega_{mH}\}. \tag{1.29}$$

However, $\{w_H\}$ is not in $\mathcal{F}_n$ for any n since it depends on the results of all coin tosses. This means that $\mathcal{F}_\infty$ cannot be a σ-algebra as it does not contain the countable union of $\{\omega_{nH} \mid n \in \mathbb{N}\} \in \mathcal{F}_\infty$. To remedy this, we can just use $\mathcal{F} = \sigma(\mathcal{F}_\infty)$. Once again, we can extend our probabilities to events in $\mathcal{F}$ that are in $\mathcal{F}_\infty$ uniquely. Unlike $\mathcal{F}_\infty$, $\mathcal{F}$ contains events which cannot be specified by any finite number of coin tosses like ω_H which we discussed before. We can assign a probability to this event using the continuity of measure from above. Since $\{\omega_{nH}\} \supset \{\omega_{(n+1)H}\}$ and $\{\omega_H\} = \cap_{n\in\mathbb{N}}\{\omega_{mH}\}$, it follows that

$$\mathbb{P}(\{\omega_{nH}\}) = p^n \downarrow \mathbb{P}(\{\omega_H\}) = 0. \tag{1.30}$$

1.3 Random variables and distributions

Defining random variables. Though probability spaces are plenty interesting themselves, we're often interested in evaluating functions of probability spaces to pick apart to answer the questions which interest us. For this purpose, we'll rely on random variables.

Definition 1.16. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X \colon \Omega \to \mathbb{R}$ is called a random variable if for every Borel set $B \in \mathcal{B}(\mathbb{R})$, we have that

$$X^{-1}(B) = \{\omega \mid X(\omega) \in B\} \in \mathcal{F}. \tag{1.31}$$

To be precise about the choice of σ-algebra, we would say that X is $\mathcal{F}$-measurable and write $X \in \mathcal{F}$.

In the case where Ω is discrete, then any function $X \colon \Omega \to \mathbb{R}$ is a random variable because the σ-algebra $\mathcal{F} = \mathcal{P}$ which contains all subsets of Ω.

Indicator functions One of the most simple examples of random variables are the indicator function. An indicator function is a random variable of the form

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A, \end{cases} \tag{1.32}$$

where is some event $A \in \mathcal{F}$. We can prove directly that $\mathbb{1}_A$ is indeed a random variable through a bit of case work. Suppose that we have a Borel set B, then

$$\mathbb{1}_A(B) = \begin{cases} A, & 1 \in B \text{ and } 0 \notin B \\ A^c, & 0 \in B \text{ and } 1 \notin B \\ \Omega, & 0, 1 \in B \\ \emptyset, & 0, 1 \notin B. \end{cases} \tag{1.33}$$

Distributions of random variables. We can use random variables to create entirely new probability distributions.

Proposition 1.17. *If X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, then X induces a new probability measure μ on $\mathbb{R}$ by*

$$\mu(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)). \tag{1.34}$$

We call this new probablity measure the *distribution of X*. We won't prove proposition entirely, but note that countable additivity follows from the fact that $\mathbb{P}$ is countably additive and

$$X^{-1}\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \bigcup_{i \in \mathbb{N}} X^{-1}(B_i) \tag{1.35}$$

when the B_i are disjoint.

Distributions functions The distribution of a random variable X is usually described by its *distribution function*

$$F(x) = \mathbb{P}(X \le x) = \mathbb{P}(X \in (-\infty, x)). \tag{1.36}$$

There are certain properties that distribution function must follow in order to properly characterize a probability distribution.

Proposition 1.18 (Properties of distribution functions.)**.** *A distribution function F must satisfy the following:*

(i) F must be non-decreasing.

(ii) $F(x) \xrightarrow[x \to \infty]{} 1$ and $F(x) \xrightarrow[x \to -\infty]{} 0$.

(iii) F is right continuous i.e. $\lim_{y \to x^+} F(y) = F(x)$.

(iv) If $F(x-) = \lim_{y \to x^-} F(y)$, then $F(x-) = \mathbb{P}(X < x)$.

(v) $\mathbb{P}(X = x) = F(x) - F(x-)$.

Proof. Insert proof. $\qquad\square$

Theorem 1.19. *If a function F satisfies $(i), (ii),$ and (iii) of the previous proposition, then it is the distribution function of some random variable,*

Sketch of proof. $\qquad\square$

Generating random variables Though the distribution function may not actually have a proper incerse, we will call the random variable X in the above theorem the inverse of F and denote it as F^{-1}. Using the sketch outlined above, we can actually generate random variables using a bit of algebra and a computer. In computer systems, standard algorithms generate random variables with uniform distributions U. We can use this fact to find the proper inverse function taking U to desired distribution of a random variable X

$$F_X = F^{-1}(U). \tag{1.37}$$

Density functions   When the distribution function $F(x)$ of a random variable X can be written in the form

$$F(x) = \int_{-\infty}^{x} f(y)dy, \tag{1.38}$$

we say that the random variable X has density f.

Example 1.20 (Uniform distribution on (a, b))**.** Suppose we're given a density function

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{otherwise.} \end{cases} \tag{1.39}$$

Here a and b are real numbers such that $a < b$. We can integrate this function directly to find the distribution function of this random variable to find

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0, & x < a \\ \frac{x}{b-a}, & a \le x \le b \\ 1, & x > b. \end{cases} \tag{1.40}$$

This distribution is called the uniform distribution on $[a, b]$.

Example 1.21 (Exponential distribution with rate parameter λ)**.** Suppose we're given the density funciton

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0. \end{cases} \tag{1.41}$$

Once again, we can integrate to find that

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0. \end{cases} \tag{1.42}$$

This distribution is called the exponential distribution with rate parameter λ.

Constructing the exponential distribution   Following the sketch proof of INSERT REFERENCE, we can constuct an exponential random variable X from the Uniform random variable on $[0, 1]$. Starting with FORMULA REFERENCE

$$X(\omega) = \sup\{y \mid F(y) > \omega\} = \sup\{\omega \mid 1 - e^{\lambda y} < \omega\}. \tag{1.43}$$

Since the distribution function is continuous, we can compute the above supremum as the value x which satisfies $1 - e^{-\lambda x} = \omega$, Solving for x, we obtain that

$$X(\omega) = x = \frac{-\log(1 - \omega)}{\lambda}. \tag{1.44}$$

Therefore, if we wanted to generate random samples from an expoenetial random variable,, we could generate random numbers ω from the uniform distribution on $[0, 1]$ and apply X to them.

Discrete probability distributions   A probability measure or distribution is called discrete if it only obtains positive value on a countable set.

Example 1.22 (Poisson Random Variable)**.** The Poisson random is a random variable defined on $\Omega = \{0, 1, 2, \ldots\}$ with σ-algebra $\mathcal{F} = \mathcal{P}(\Omega)$. It has probability mass function

$$\mathbb{P}(\omega = k) = \frac{\lambda^k}{k!}e^{\lambda}. \tag{1.45}$$

1.4   Expectation

Definition 1.23 (Expectation of a random variable)**.** The *expctation of a random variable* $X \colon \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the Lebesgue integral

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P}. \tag{1.46}$$

Computing expectations   Moving forward, we will show several examples and techniques which can be used to compute the expectation of random variables.

Example 1.24 (Expectation of an indicator function.)**.** The simplest case is that of indicator functions. Consider the indicator function $\mathbb{1}_A$ for set $A \in \mathcal{F}$. Fromt the definition of the expectation, we can write this as

$$\mathbb{E}[\mathbb{1}_A] = \int_{\Omega} \mathbb{1}_A d\mathbb{P} = 1 \cdot \mathbb{P}(A) + 0 \cdot \mathbb{P}(A^c) = \mathbb{P}(A). \tag{1.47}$$

This example shows that we can easily compute the expectation by treating the indicator function as a simple function, then computing the expectation accordingly.

The previous example gives some necessary insight to the kinds of tricks we'll have to use in order to compute some expectations, but there are some simplifications we can make based on our calculus intuition for computing the expectation for random variables. In the more general setting, we can use a change of variables to do this integration over the real numbers, so that the expectation of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is given by

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} x \mathbb{P}_X(dx), \tag{1.48}$$

where $\mathbb{P}_X$ is the distribution of X on $\mathbb{R}$. In the case X is continuous and has a density function f_X, we have that

$$\mathbb{P}_X(dx) = f_X(x)dx. \tag{1.49}$$

Therefore, the full expectation can be computed as

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} x f_X(x)dx, \tag{1.50}$$

for continuous random variables X with density f_X.

When X is instead discrete, then

$$\mathbb{P}_X(dx) = \mathbb{P}(X^{-1}(\{x\})) = \mathbb{P}(X = x). \tag{1.51}$$

This allows us to write the expectation as

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \sum_{\omega \in \Omega} \omega \mathbb{P}(X = \omega). \tag{1.52}$$

We can think of this as a change of variables from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$.

Let's use this technique to compute the expectation of several of the random variables we've previously discussed.

Example 1.25 (Expectation of a Poisson Random Variable)**.** Using the previous definition of the Poisson variable X with rate parameter λ, we can compute the expectation as

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \lambda. \tag{1.53}$$

Example 1.26 (Expectation of an Exponential Random Variable)**.**

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} = \int_{\mathbb{R}} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}. \tag{1.54}$$

We can also use this technique to compute the expectation of measurable functions g of a random variable X. Since $g(X)$ itself is a random variable due to the measurability of g, we can compute the expecation of $g(X)$ as

$$\mathbb{E}[g(X)] = \int_\Omega g(X) d\mathbb{P} = \int_{\mathbb{R}} g(x) \mathbb{P}_X(dx). \tag{1.55}$$

This result allows us to prove some interesting consequences of the expectation. The most important one being that the expectation is linear.

Theorem 1.27 (Linearity of expectation)**.** *If X and Y are random variables and a and b are constants, then*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]. \tag{1.56}$$

2. Independence, Martingales, and Conditioning

3. Characteristic Functions

4. Markov Chains

5. Generating Functions and Branching Processes

6. Convergence of Random Variables