Mathematical Statistics I

Chapter 1: Probability

Jesse Wheeler

Outline

Introduction

Introduction

Formally, a random variable is a function from a sample space Ω to the real numbers[1].

That is, for any element $\omega \in \Omega$, a random variable X will map ω to a real number: $X(\omega) \in \mathbb{R}$.

Most often people think of random variables as random numbers rather than functions; in most instances in this class, this treatment will be sufficient.

[1] In this class, will assume real-valued spaces, though more generally a random variable can map to any measureable space

Example of a random variable

Consider the experiment of flipping three coins. The sample space is

$$\Omega = \{hhh, hht, hth, thh, htt, tht, tth, ttt\}.$$

Some possible random variables include (1) the number of heads, (2) the number of tails, (3) the number of heads minus the number of tails.

Importantly, a random variable must assign a value to all possible outcomes $\omega \in \Omega$.

Number of Heads

Let X be the random variable representing the number of heads. If the result of the outcome is the event hth, the $X(\{hth\}) = 2$.

A few comments on random variables

- Sometimes in this course I will use the abbreviation RV to mean "random variable", and you can do so as well.
- It is conventional to use uppercase letters (math text or italics) to denote random variables.
- While a random variable is a function, the outcome of an experiment $\omega \in \Omega$ is random (that's the point), and we only ever see a single outcome. Thus, the fact that X is a function is often dropped, and we just write X. The realized value of X is random, because the input is random.

Discrete Random Variables

Discrete Random Variables

Definition: Discrete random variable

A discrete random variable is a random variable that can take on only a finite or at most a countably infinite number of values.

Example: The number of heads in three coin flips can only be in the set $\{0, 1, 2, 3\}$. Alternatively, consider flipping a coin indefinitely until you achieve a heads. The possible outcomes are in the set $\{1, 2, 3, \ldots\}$, which is countably infinite.

Probabilities

The probability measure on the sample space determines the probability of the values of X. In our example, if a coin is fair, then we can assign a uniform probability measure on the sample set of flipping a coin three times. That is, all outcomes are equally likely, each with probability $1/8$. The probability that X takes on it's potential values is easily computed, by counting the number of outcomes that result in the particular value of X:

Probabilities II

$$P(X=0)=\frac{1}{8}$$
$$P(X=1)=\frac{3}{8}$$
$$P(X=2)=\frac{3}{8}$$
$$P(X=3)=\frac{1}{8}.$$

Probabilities III

More generally, let's assume that X is a discrete RV, and denote the possible values as $x_1, x_2, \ldots$. There exists a function p such that $p(x_i) = P(X = x_i)$ that satisfies $\sum_i p(x_i) = 1$. This function p is called the probability mass function (PMF) of the random variable X.

We may also be interested in calculating for all values $x \in \mathbb{R}$, the probability $F(x) = P(X \leq x)$; the function F is called the cumulative distribution function (CDF). The CDF plays a number of important roles in probability and statistics that we will see later on.

Probabilities IV

Some notes:

- The CDF is non-decreasing (see Theorem 1.2), and

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F(x) = 1.$$

- The PMF and CDF are connected: the CDF "jumps" at all values that the pdf $p(x) > 0$.
- Conventionally, the PMF is usually denoted with lower-case letters (e.g., p, f), whereas the CDF is usually denoted with upper-case letters (e.g., F).

Independence

Jumping ahead a little bit, we will define what it means for random variables to be independent (a chapter 3 topic), as it will be useful for our discussions in this chapter.

Definition: Independent random variables

Let X and Y be discrete random variables defined on the same probability space, taking values $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$, respectively. X and Y are said to be independent if, for all i, j,

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i).$$

This definition follows very similarly to that of independent events. Similar to that case, we can extend this definition to mutaul independence of many variables if the probabilities of all combinations of variables can be factored.

Bernoulli Random Variables

A Bernoulli RV only takes on two values[2], 0 and 1, with probabilities $1-p$ and p, respectively. The PMF is therefore

$$\begin{aligned} p(1) &= p \\ p(0) &= 1-p \\ p(x) &= 0, \quad \text{if } x \neq 0 \text{ and } x \neq 1. \end{aligned}$$

By using the output of 0 and 1, the PMF is usually written in a more compact form:

$$p(x) = \begin{cases} p^x(1-p)^x, & \text{if } x = 0 \text{ or } x = 1, \\ 0 & \text{otherwise} \end{cases}$$

[2]Sometimes you'll see the random variable take values -1 and 1.

Indicator functions

A common instance of a Bernoulli RV is an indicator random variable. Let I_A be the random variable that takes on the value of 1 if the event $A \subset \Omega$ occurs, and 0 otherwise:

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

Here, we see that $P(I_A = 1) = P(A)$.

Junk Citation

Rice (2007)

References and Acknowledgements

Rice JA (2007). *Mathematical statistics and data analysis*, volume 371. 3 edition. Thomson/Brooks/Cole Belmont, CA.

- Compiled on August 13, 2025 using R version 4.5.1.
-
- We acknowledge students and instructors for previous versions of this course / slides.