

Probability and Statistics

UT Knoxville, Fall 2024, MATH 323

David White, Alex Zhang

August 26, 2023

Contents

1	Introduction	2
----------	---------------------	----------

Introduction

As an introduction, let's consider some problems we might want to answer. We may use probability to test models against data. For example, imagine we rolled a six-sided die 100 times and rolled five each time. It's certainly a possibility albeit an unlikely one. Now imagine instead we rolled another die 100 times and got 17 1s, 17 2s, 17 3s, 17 4s, 16 5s, and 16 6s. Our intuition leads us to believe that the second die is more "fair" than the first. The first die certainly seems more suspicious than the second. The problem is: how do we quantify our suspicion with the first die?

Definition 1.0.1 ► Experiment, event, simple event, sample point

An *experiment* is the process by which an observation is made. Something can be an experiment if it has a measurable outcome. An *event* is the set of outcome(s). An event is *simple* if it only contains one outcome, in which case it cannot be decomposed.

For example, rolling dice, flipping a coin, or drawing a card from a deck can all be experiments. Let's consider an experiment in which we flip a coin three times. There are eight possible outcomes, which we will denote by a set named S :

$$S := \{HHH, TTT, HHT, HTH, HTT, THH, THT, TTH\}$$

In this case, any subset of S can be considered an event. Suppose we wanted to focus on all the outcomes where heads appeared more than tails. The set of these outcomes can be simply defined as:

$$E := \{HHH, HHT, HTH, THH\}$$

If we only focus on a single specific outcome, say HHH, we denote this by a singleton set containing only that outcome:

$$H := \{HHH\}$$

This event H is simple as it only contains one possible outcome: HHH.

Definition 1.0.2 ▶ Probability

The **probability** of an event is the likelihood that the event occurs. Given an event E , we write $P(E)$ to denote the probability that event E occurs.

In our coin flip example (and other naive probability examples), we can deduce the probability of an event by the following formula:

$$P(E) = \frac{\text{num. outcomes in } E}{\text{num. possible outcomes}}$$

The probability that more heads occur than tails (denoted by E) is calculated as:

$$P(E) = \frac{4}{8} = \frac{1}{2}$$

Example 1.0.3 ▶ Betting

Question: Suppose that two people, B and P , have placed equal bets on winning the best of 5 fair coin flips. B is betting on heads, and T is betting on tails. They are interrupted after 3 flips and have to stop the game short, with B ahead 2 heads to 1 tails. How should they fairly divide pot?

It's clear that B is more likely to win this game than P . The question is: how probable is B winning in this scenario? If we replicated this experiment 44 times, we may see that B wins 39 times and P wins 15 times. Denoting S as the set of all possible outcomes, we have:

$$S := \{HH, HT, TH, TT\}$$

From these 4 outcomes, B will win 3 of the 4 outcomes, and P will win 1 of the outcomes.

TODO: stuff below should be mixed with above stuff?

Definition 1.0.4 ▶ Sample point, sample space, discrete sample space

A **sample point** is any single outcome from an experiment. The **sample space** is the set of all possible sample points. A sample space is **discrete** if it contains a countable amount of distinct sample points.

In this context, we can think of an event as a collection of sample points. Note that a simple event is always a singleton set whereas a sample point does not have to be; we will not worry

over these differences and simply use the two terms interchangeably.

Definition 1.0.5 ► Probability (Formal Definition)

Let S be the sample space associated with an experiment. To each event $A \subseteq S$, we assign a number $P(A)$ called the **probability** of A , satisfying:

1. $P(A) \geq 0$ for all $A \subseteq S$,
2. $P(S) = 1$, and
3. if A_1, \dots, A_n are disjoint, then we have $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$.

For example, let's consider a homemade, six-sided die. Using S to denote our sample space, we have

$$S := \{E_1, E_2, E_3, E_4, E_5, E_6\}$$

where E_j is the event that j is rolled. Suppose that we had the following probabilities for each of the events:

$$P(E_1) = 1/3$$

$$P(E_2) = 1/4$$

$$P(E_3) = 1/6$$

$$P(E_4) = 1/12$$

$$P(E_5) = 1/8$$

$$P(E_6) = 1/24$$

To find the probability that we roll an even number, we look at the event $\{E_2, E_4, E_6\}$. The probability that this event occurs is $1/4 + 1/12 + 1/24 = 3/8$.

Example 1.0.6 ► Laptop Refurbish

Suppose we had a refurbished laptop with the following probabilities:

$$P(\text{bad battery}) = 0.32$$

$$P(\text{bad screen}) = 0.18$$

$$P(\text{bad battery and bad screen}) = 0.12$$

Find the probabilities for:

$$P(\text{bad battery OR bad screen})$$

$$P(\text{neither defect})$$

$$P(\text{bad screen but NOT bad battery})$$

In this example, we have only two simple events: having a bad battery but not a bad screen, and having a bad screen but not a bad battery.

TODO: venndiagram

From this, we have:

$$P(\text{bad battery OR bad screen}) = 0.20$$

$$P(\text{neither defect}) = 0.62$$

$$P(\text{bad screen but NOT bad battery}) = 0.06$$

Example 1.0.7 ► Markers

Five seemingly identical markers are left in a classroom. Only two of have enough ink to write well. The instructor selects two of these markers at random.

1. What is the sample space?
2. Assign probabilities to each sample point in the sample space.
3. What is the probability that neither marker has enough ink to write?

To determine the sample space, we consider all the possibilities for the instructor. We denote each marker as:

$$W_1, W_2, D_1, D_2, D_3$$

If we consider picking W_1 then W_2 to be the same event as picking W_2 then W_1 , we can simply enumerate all possible pairs of markers for our sample space:

$$S := \{W_1 W_2, W_1 D_1, W_1 D_2, W_1 D_3, W_2 D_1, W_2 D_2, W_2 D_3, D_1 D_2, D_1 D_3, D_2 D_3\}$$

The probability of any sample point occurring is $1/10$. From this, we can simply add up the sample points for each of the events. The probability that the instructor selects two dead markers is given by the event $\{D_1 D_2, D_1 D_3, D_2 D_3\}$ whose probability is $3/10$.