Probability and Statistics

UT Knoxville, Fall 2024, MATH 323

David White, Alex Zhang

September 27, 2023

Contents

1

Probability

1.1 Introduction

Definition 1.1.1 ▶ 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.

Definition 1.1.2 ▶ 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.

Definition 1.1.3 ▶ Probability

Intuitively, 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.

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 $P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$.

Example 1.1.4 ▶ 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:

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

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

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.

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:

$$\begin{aligned} P(E_1) &= 1/3 \\ P(E_2) &= 1/4 \\ P(E_3) &= 1/6 \\ P(E_4) &= 1/12 \\ P(E_5) &= 1/8 \\ P(E_6) &= 1/24 \end{aligned}$$

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 1.1.5 ▸ Laptop Refurbish

Suppose we had a refurbished laptop with the following probabilities:

$$\begin{aligned} P(\text{bad battery}) &= 0.32 \\ P(\text{bad screen}) &= 0.18 \\ P(\text{bad battery and bad screen}) &= 0.12 \end{aligned}$$

Find the probabilities for:

$$\begin{aligned} P(\text{bad battery OR bad screen}) \\ P(\text{neither defect}) \\ P(\text{bad screen but NOT bad battery}) \end{aligned}$$

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:

$$\begin{aligned} P(\text{bad battery OR bad screen}) &= 0.20 \\ P(\text{neither defect}) &= 0.62 \\ P(\text{bad screen but NOT bad battery}) &= 0.06 \end{aligned}$$

Example 1.1.6 ▸ 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:

$$S := \{W_1W_2, W_1D_1, W_1D_2, W_1D_3, W_2D_1, W_2D_2, W_2D_3, D_1D_2, D_1D_3, D_2D_3\}$$

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 $\{D_1D_2, D_1D_3, D_2D_3\}$ whose probability is $3/10$.

Technique 1.1.7 ▸ Calculating Probability: The Sample Point Method

The sample point method is a very straightforward approach to calculating the probability of an event in an experiment.

1. List all the simple events associated with an experiment. This defines the sample space S.
2. Assign reasonable probabilities to the sample points in S.
3. Define the event of interest A as a subset of S.
4. Find $P(A)$ by summing the probability of each sample point in A.

1.2 Combinatorics

Definition 1.2.1 ▸ Permutation

A ***permutation*** is an ordered arrangement of r distinct objects. The number of permutations of size n among r objects is defined as:

$$P_n^r := \frac{n!}{(n-r)!} = n(n-1)(n-2)\cdots(n-r+1)$$

Theorem 1.2.2 ▸ Number of partitions

The number of ways partitioning n distinct objects into k disjoint sets is:

$$\binom{n}{n_1\ n_2\ \cdots\ n_k} := \frac{n!}{n_1!n_2!\cdots n_k!}$$

The terms $\binom{n}{n_1\ n_2\ \cdots\ n_k}$ are often called ***multinomial coefficients*** because they occur in the expansion of $y_1 + y_2 + \cdots + y_k$ raised to the nth power:

$$(y_1 + y_2 + \cdots + y_k)^n = \sum \binom{n}{n_1\ n_2\ \cdots\ n_k} y_1^{n_1} y_2^{n_2} \cdots y_k^{n_k}$$

Definition 1.2.3 ▸ Combination

The number of combinations of n objects taken r at a time is given by

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

We can think of the above definition in terms of sets. Given a set A of size N, $\binom{n}{r}$ is the number of possible distinct subsets of A that are of size r.

Theorem 1.2.4 ▸ Binomial Theorem

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

Intuition: For example, consider $(t+h)^5$, which expands to:

$$(t+h)(t+h)(t+h)(t+h)(t+h)$$

If we were to fully distribute this out, we have:

$$t^5 + 5t^4h + \binom{5}{3}t^3h^2 + \binom{5}{2}t^2h^3 + 5th^4 + h^5$$

Example 1.2.5 ▸ Coin flips

Four fair coins are flipped. What is the most likely outcome?

(a) All H or all T
(b) 2H, 2T
(c) 3H 1T or 1H 3T

Let's consider how many ways there are to get each of the results.

(a) There are 2 ways to get either all heads or all tails.
(b) There are 6 ways to get 2 heads and 2 tails.
(c) There are 8 ways to get 3 heads 1 tail, or 3 tails 1 head.

(TODO: choose notation and explanations)

Example 1.2.6 ▸ Poker hands

A standard deck of cards has 52 cards, with 4 suits and 13 ranks. The number of distinct 5 card hands (not accounting for order) can be calculated by $\binom{52}{5}$. If order does matter, then we count the number of permutations P_5^{52}. Most card games don't care about the order of the hand, so we look at the first option (which is called choices without replacement).

What is the probability of having a hand with exactly 2 aces? The total number of 2 ace hands is calculated by:

$$\binom{4}{2}\binom{48}{3}$$

Although there are 50 cards left over after selecting 2 aces, we do not want three or four-tuple aces, which explains 48 instead of 50. (TODO: wording) Thus, the probability of getting a 2 ace hand is:

$$\frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}}$$

(TODO: 48 instead of 50?)

The probability that we get a hand with two pairs is decided by the following choices:

- Which pairs?
- Which cards for those pairs?
- What's the left over card?

As such, we can calculate the number of two pair hands by:

$$\underbrace{\binom{13}{2}}_{\text{choose two ranks}} \cdot \binom{4}{2}\binom{4}{2}\binom{44}{1}$$

The probability that we get a full house (1 pair, 1 triple) is decided by (TODO). The total number of full house hands is calculated by:

$$\underbrace{\binom{13}{1}}_{\text{pair rank}} \underbrace{\binom{12}{1}}_{\text{triple rank}} \underbrace{\binom{4}{2}}_{\text{which pair}} \underbrace{\binom{4}{3}}_{\text{triple}}$$

(left to right, pair rank, triple rank, which pair, triple)

Example 1.2.7 ▶ Yahtzee

In Yahtzee, we roll 5 six-sided dice. There are 6^5 total different outcomes.
The number of two-pair rolls can be calculated by:

$$\binom{6}{2}\binom{4}{1}\binom{5}{2}\binom{3}{2}$$

(left to right: ranks of the pairs, rank of leftover, location of first pair, location of second pair)
The number of rolls where all five dice are different can be calculated by:

$$\underbrace{6}_{\text{which num. missing}} \cdot \underbrace{5!}_{\text{ways to rearrange}}$$

TODO :oisdjfojsdoisjf

1.3 Conditional Probability and Independence

Definition 1.3.1 ▶ Conditional probability

The ***conditional probability*** of an event A, given that B has occurred, is defined as:

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)} \quad \text{if} \quad P(B) > 0.$$

We read $P(A \mid B)$ as "probability of A given B."

For example, if we roll a six-sided die, and we already know that the die landed on an odd number, then the probability that it's 1 is:

$$P(1 \mid \text{odd}) = \frac{P(1 \text{ and odd})}{P(\text{odd})} = \frac{1/6}{3/6}$$

Example 1.3.2 ▶ Cards

Cards are dealt one at a time from a standard deck. If the first 2 are spades, what is the probability that the next 3 are also spades?

In this example, we consider the first two being spades to be the initial condition B and the next 3 being spades as A. We first calculate $P(A \cap B)$:

$$P(A \cap B) = \frac{\binom{13}{5}}{\binom{52}{5}}$$

$$P(B) = \frac{\binom{13}{2}}{\binom{52}{2}}$$

Thus:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\binom{13}{5}\binom{52}{2}}{\binom{52}{5}\binom{13}{2}} = \dots$$

(TODO: answer) To confirm our answer, we can asdofadsiofj

Definition 1.3.3 ▸ Independent, dependent

Intuitively, two events are called ***independent*** if the occurrence of one does not affect the probability of occurrence of the other.

More formally, two events A and B are ***independent*** any of the following are true:

- $P(A \mid B) = P(A)$,
- $P(B \mid A) = P(B)$, or
- $P(A \cap B) = P(A) \cdot P(B)$.

Otherwise, the events are ***dependent***.

Example 1.3.4 ▸ Independent dice rolls

Roll a six-sided die once. Let:

- $A \coloneqq \{\text{roll is odd}\}$,
- $B \coloneqq \{\text{roll is even}\}$,
- $C \coloneqq \{\text{roll is 1 or 2}\}$.

We can see A and B are **not** independent by the following calculation:

$$0 = P(A \cap B) \neq P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

However, we can see A and C are independent by the following:

$$\frac{1}{6} = P(A \cap C) = P(A) \cdot P(B) = \frac{3}{6} \cdot \frac{2}{6} = \frac{1}{6}$$

Example 1.3.5 ▸ Independent coffee brands

Three brands of coffee, X, Y, and Z, are ranked according to taste by a judge.

- A: X is better than Y
- B: X is the best
- C: X is second best
- D: X is last

If the ranking is truly random, is A independent of B, C, and/or D?

First, we see that there are 6 possible rankings, which we will denote by S:

$$S \coloneqq \{XYZ, XZY, YXZ, YZX, ZXY, ZYX\}$$

We have the following probabilities:

- $P(A) = 1/2$
- $P(B) = 1/3$
- $P(C) = 1/3$
- $P(D) = 1/3$

1.4 Two Laws of Probability

Theorem 1.4.1 ▸ Multiplicative Law of Probability

The probability of the intersection of two events A and B is:

$$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

If A and B are independent, then:

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

Theorem 1.4.2 ▸ Additive Law of Probability

The probability of the union of two events A and B is:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A and B are disjoint, then:

$$P(A \cap B) = 0$$

Theorem 1.4.3 ▸ Probability of complementary event

Let A be an event. Then $P(A) = 1 - P(\overline{A})$.

1.5 The Law of Total Probability and Bayes' Theorem

Definition 1.5.1 ▸ Partition

Let S be a set. If $S = B_1 \cup B_2 \cup \cdots \cup B_k$, and these sets $B_1, \ldots, B_k$ are disjoint, then the set $\{B_1, \ldots, B_k\}$ is called a ***partition*** of S. Moreover, for any $A \subseteq S$:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$$

where $A \cap B_1, \ldots, A \cap B_k$ are disjoint.

Partitions are especially useful to us in computing probability. For example:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{k} P(A \mid B_i)P(B_i)$$

Example 1.5.2 ▸ Probability using partition

A diagnostic test for a disease is 95% accurate. Let E_d denote the event where a person has a disease, and let E_+ denote a positive test. Then:

$$P(E_+ \mid E_d) = 0.95$$
$$P(\overline{E_+} \mid \overline{E_d}) = 0.95$$

If 1% of the population has the disease, what is the probability that a randomly selected person tests positive?

Since E_d and $\overline{E_d}$ are disjoint, we can partition the event E_+ into two sets, $E_+ \cap E_d$ and $E_+ \cap \overline{E_d}$. As such, we can calculate $P(E_+)$ by:

$$\begin{aligned} P(E_+) &= P(E_+ \cap E_d) + P(E_+ \cap \overline{E_d}) \\ &= P(E_+ \mid E_d)P(E_d) + P(E_+ \mid \overline{E_d})P(\overline{E_d}) \\ &= 0.95 \cdot 0.01 + 0.05 \cdot 0.99 \\ &= 0.059 \end{aligned}$$

Theorem 1.5.3 ▸ Bayes' Theorem

Let S be a set. If $\{B_1, \ldots, B_k\}$ is a partition of S, then for any $j \in \{1, \ldots, k\}$:

$$\begin{aligned} P(B_j \mid A) &= \frac{P(B_j \cap A)}{P(A)} \\ &= \frac{P(A \mid B_j)P(B_j)}{\sum_{i=0}^{k} P(A \mid B_i)P(B_i)} \end{aligned}$$

Example 1.5.4 ▸ Probability using Baye's theorem

Revisiting Example 1.5.2, we have:

$$\begin{aligned} P(E_d \mid E_+) &= \frac{P(E_+ \mid E_d)P(E_d)}{P(E_+ \mid E_d)P(E_d) + P(E_+ \mid \overline{E_d})P(\overline{E_d})} \\ &= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \\ &= 0.16 \end{aligned}$$

Example 1.5.5

Five identical bowls. Bowl i contains i white marbles and $5 - i$ black marbles. A bowl is randomly selected, and two marbles are removed without replacement. Determine the probability of $P(\text{both white})$ and $P(\text{bowl } 3 \mid \text{both white})$.

On paper notes; branching thing!!!

2

Random Variables

Definition 2.0.1 ▶ Random variable

A ***random variable*** is a real-valued function whose domain is a sample space. A random variable is ***discrete*** if its domain is finite or countably infinite.

For example, consider an experiment where we flip three coins. Let Y be the number of heads. Then there are four possible values for Y: 0, 1, 2, and 3. The probability of each value is:

$$P(Y=0) = 1/8$$
$$P(Y=1) = 3/8$$
$$P(Y=2) = 3/8$$
$$P(Y=3) = 3/8$$

Examples from notes

The probability that Y takes on the value y, $P(Y = y)$, is sometimes written as $p(y)$. The probability distribution of Y must have:

1. $0 \leq p(y) \leq 1$ for any value y; and
2. $\sum_y p(y) = 1$, where the summation is over all values y in the domain of Y.

Definition 2.0.2 ▶ Expected Value

Let Y be a discrete random variable with probability function $p(y)$. Then the ***expected value*** of Y is defined as:

$$E(Y) \coloneqq \sum_y y p(y)$$

If the sum diverges, then no such expected value exists.

Intuitively, it's a weighted average of all the possible values of y. If Y is an accurate characterization of population frequency distribution, then $E(Y)$ is the population mean, in which case we write $E(Y) = \mu$.

If we have a real-valued function g of Y, the expected value of $g(Y)$ is given by:

$$E\left(g(y)\right) = \sum_{y} g(y)p(y)$$

where $p(y)$ is the probability function associated with Y. Note that $g(Y)$ is itself a random variable.

Definition 2.0.3 ▶ Variance, standard deviation

Let Y be a random variable with mean $E(Y) = \mu$. The ***variance*** of Y is defined as:

$$V(Y) := E\left((Y-\mu)^2\right)$$

The ***standard deviation*** of Y is defined as:

$$\sigma(Y) := \sqrt{V(Y)} \quad \text{where } \sigma(Y) \geq 0$$

Intuitively, variance tells us how "spread out" the probabilities are from the mean.

Theorem 2.0.4

$E(cg(Y)) = cE(g(Y))$
$E(g_1(Y) + g_2(Y) + \cdots + g_k(Y)) = E(g_1(Y)) + E(g_2(Y)) + \cdots = E(g_k(Y))$
$V(Y) = E((Y-\mu)^2) = E(Y^2) - \mu^2$

2.1 The Binomial Probability Distribution

Definition 2.1.1 ▶ Binomial experiment, Bernoulli(p) trial

A ***Bernoulli (p) trial*** is one experiment either a success S or failure F. In it:

- $P(S) = p$
- $P(F) = 1 - p = q$
- $X = \begin{cases} 1, & \text{with probability } p \\ 2, & \text{with probability } 1-p \end{cases}$

We write Binomial(n, p) to be the sum of n independent Bernoulli(p) trials.

Thus, we have the following expected values:

$$E(X) = 0(1-p) + 1(p) = p$$
$$E(X^2) = 0^2(1-p) + 1^2 p = p$$
$$V(X) = E(X^2) - (E(X))^2 = p - p^2 = pq$$

$Y = \text{binomial}(n, p)$, where n identical independent trials are each a success or failure. $P(S) = p$, and $P(F) = 1 - p = q$. In this, Y denotes the number of successes in n trials.

Example 2.1.2

40% of voters in a large population support candidate J. If we ask 10 randomly selected voters, how many will say they support J?

$X = \text{binomial}(10, 0.4)$. We should have:

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

This is related to $(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$.

Definition 2.1.3 ▶ Binomial distribution

A random variable Y has ***binomial distribution*** based on n trials with probability for success p if and only if:

$$P(Y = y) = \binom{n}{y} p^y q^{n-y} \quad y \in \{0, 1, \dots, n\} \text{ and } 0 \le p \le 1$$

For $Y = \text{binomial}(n, p)$, we have:

$$p(y) = \binom{n}{x} p^x (1-p)^{n-x}$$
$$E(Y) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

For $Y = X_1 + X_2 + \cdots + X_n$ where each $X_i = \text{Beronulli}(p)$:

$$E(Y) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = np$$

Similarly:

$$V(Y) = V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n)$$

Example 2.1.4 ▶ Calculators

10 calculators to sell, $80 each. But, double your money back guarantee if a calculator is defective. 0.08 probability that a calculator is defective. If all 10 are sold, what is the expected revenue?

Let X be the number of defective calculators. Then $X = \text{binomial}(n = 10, p = 0.08)$. Then our revenue will be:

$$\underbrace{80 \cdot 10}_{\text{revenue for 10 calcs}} - \underbrace{160 \cdot X}_{\text{double money back for defective}}$$

Then our expected revenue is:

$$E(800 - 160X) = 800 - 160E(X) = 800 - 160 \cdot 0.8$$

Example 2.1.5

What is more likely: at least one 6 in 4 dice rolls, or at least one double 6 in 24 double dice rolls?

We let $X = \text{binomial}(n = 4, p = 1/6)$. Then $P(X \geq 1) = 1 - (5/6)^4$. Also, let $Y = \text{binomial}(n = 24, p = 1/36)$. Then $P(Y \geq 1) = 1 - (35/36)^{24}$.

Definition 2.1.6 ▶ Geometric random variable

Independent Bernoulli(p) trials until the first success. This is written as:

$$Y = \text{geometric}(p)$$

In general, $P(Y = y)$ for geometric random variables is given by:

$$p(y) = \underbrace{(1 - p)^{y-1}}_{\text{must fail first } y-1 \text{ times}} p$$

Also, note that there can be infinitely many failed trials before a success trial. Thus, y can be any value in $\{1, 2, 3, \ldots\}$. The probability distribution for Y must still sum to 1. That is:

$$\sum_{y=1}^{\infty} P(Y = y) = \sum_{y=1}^{\infty} (1-p)^{y-1} p = 1$$

Example 2.1.7 ▸ Dice rolls

Suppose we roll a dice until a 6 appears. What is the probability it takes more than 3 rolls?

We have $X = \text{geometric}(p = 1/6)$. The probability it takes more than 3 rolls is given by:

$$P(X > 3) = \sum_{k=4}^{\infty} P(X = k)$$

These kinds of infinite sums are hard to deal with. We can instead take an approach considering the complement. $P(X > 3) = 1 - P(X \leq 3) = \sum_{k=1}^{3} P(X = k) = \ldots = \frac{125}{216}$.

Definition 2.1.8 ▸ Negative binomial

Independent Bernoulli(p) trials until the rth success. This is written as:

$$Z = \text{negativebinomial}(r, p)$$

Note that a negative binomial can be created by taking the sum of r independent geometric random variables. For example, if $Z = \text{negativebinomial}(r, p)$, then we have:

$$E(Z) = \frac{r}{p} \quad \text{and} \quad V(Z) = \frac{r(1-p)}{p^2}$$

The possible values for Z can be $r, r+1, r+2$, and so on. For any $k \geq r$:

$$p(k) = \binom{k-1}{r-1} p^r q^{k-r}$$

Example 2.1.9 ▸ Roblox

Each battle in a video game gives a reward of one Robux with probability 0.07. What is the probability it will take eight or fewer battles to collect five Robux?

We can construct a negative binomial as follows:

$$Z := \text{negativebinomial}(r = 5, p = 0.07)$$

We are concerned with needing eight or fewer battles, so we want:

$$P(Z \leq 8)$$

However, note that we would need at least five battles to collect 5 Robux. Thus:

$$\begin{aligned} P(Z \leq 8) &= P(Z = 5) + P(Z = 6) + P(Z = 7) + P(Z = 8) \\ &= \underbrace{(0.07)^5}_{P(Z=5)} + \underbrace{\binom{5}{4}(0.07)^5(0.93)}_{P(Z=6)} + \underbrace{\binom{6}{4}(0.07)^5(0.93)^2}_{P(Z=7)} + \underbrace{\binom{7}{4}(0.07)^5(0.93)^3}_{P(Z=8)} \end{aligned}$$