Introduction to Probability Theory

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota

August 27, 2020

Table of Contents

Table of Contents

Simple Experiment

The field of "probability theory" is a branch of mathematics that is concerned with describing the likelihood of different outcomes from uncertain processes.

A simple experiment is some action that leads to the occurrence of a single outcome s from a set of possible outcomes S.

- The single outcome s is referred to as a sample point
- The set of possible outcomes S is referred to as the sample space

Examples of Simple Experiments

Example. Suppose that you flip a coin $n \geq 2$ times and record the number of times you observe a "heads". The sample space is $S = \{0, 1, \ldots, n\}$, where $s = 0$ corresponds to observing no heads and $s = n$ corresponds to observing only heads.

Example. Suppose that you pick a card at random from a standard deck of 52 playing cards. The sample points are the individual cards in the deck (e.g., the Queen of Spades is one possible sample point), and the sample space is the collection of all 52 cards.

Example. Suppose that you roll two standard (six-sided) dice and sum the obtained numbers. The sample space is $S = \{2, 3, \ldots, 11, 12\}$, where $s = 2$ corresponds to rolling "snake eyes" (i.e., two 1's) and $s = 12$ corresponds to rolling "boxcars" (i.e., two 6's).

Definition of an Event

An <u>event</u> A refers to any possible subspace of the sample space S, i.e., $A \subseteq S$, and an <u>elementary event</u> is an event that contains a single sample point s.

For the coin flipping example, we could define the events

- $A = \{0\}$ (we observe no heads)
- $B = \{1, 2\}$ (we observe 1 or 2 heads)
- $C = \{c \mid c \text{ is an even number}\}$ (we observe an even # of heads)

Note that event A is an elementary event.

More Examples of Events

For the playing card example, we could define the events

- $A = \{$Queen of Spades$\}$ (i.e., we draw the Queen of Spades)
- $B = \{b \mid b$ is a Queen$\}$ (i.e., we draw a card that is a Queen)
- $C = \{c \mid c$ is a Spade$\}$ (i.e., we draw a card that is a Spade)

For the dice rolling example, we could define the events

- $A = \{2\}$ (i.e., we roll snake eyes)
- $B = \{7, 11\}$ (i.e., we roll natural or yo-leven)
- $C = \{c \mid c$ is an even number$\}$ (i.e., we roll dice that sum to an even number)

Note that event A is an elementary event in both of these examples.

Sure and Impossible Events

A <u>sure event</u> is an event that always occurs, and an <u>impossible event</u> (or null event) is an event that never occurs.

Example. For the coin flipping example,
$E = \{e \mid e$ is an integer satisfying $0 \leq e \leq n\}$ is a sure event and
$I = \{i \mid i > n\}$ is an impossible event.

Example. For the playing card example,
$E = \{e \mid e$ is a Club, Diamond, Heart, or Spade$\}$ is a sure event and
$I = \{\text{Joker}\}$ is an impossible event.

Example. For the dice rolling example,
$E = \{e \mid e$ is an integer satisfying $2 \leq e \leq 12\}$ is a sure event and
$I = \{i \mid i > 12\}$ is an impossible event.

Mutually Exclusive and Exhaustive Events

Two events A and B are said to be <u>mutually exclusive</u> if $A \cap B = \emptyset$, i.e., if one event occurs, then the other event can not occur. Two events A and B are said to be <u>exhaustive</u> if $A \cup B = S$, i.e., if one of the two events must occur.

Example. For the coin flipping example, the two events $A = \{0\}$ and $B = \{n\}$ are mutually exclusive events, whereas
$A = \{a \mid a$ is an even number between 0 and n$\}$ and
$B = \{b \mid b$ is an odd number between 1 and n$\}$ are exhaustive events.

Note that this is assuming that 0 is considered an even number.

Examples of Mutually Exclusive and Exhaustive Events

Example. For the playing card example, the two events
$A = \{a \mid a$ is a Spade$\}$ and $B = \{b \mid b$ is a Club$\}$ are mutually
exclusive events, whereas $A = \{a \mid a$ is a Club or Spade$\}$ and
$B = \{b \mid b$ is a Diamond or Heart$\}$ are exhaustive events.

Example. For the dice rolling example, the two events $A = \{2\}$ and
$B = \{12\}$ are mutually exclusive events, whereas
$A = \{a \mid a$ is an even number between 2 and 12$\}$ and
$B = \{b \mid b$ is an odd number between 3 and 11$\}$ are exhaustive events.

Table of Contents

Definition of a Probability

A <u>probability</u> is a real number (between 0 and 1) that we assign to events in a sample space to represent their likelihood of occurrence.

The notation $P(A)$ denotes the probability of the event $A \subseteq S$.

Two common interpretations of a probability:

- Physical interpretation views $P(A)$ as the relative frequency of events that would occur in the *long run*, i.e., if the experiment was repeated a very large number of times. (Frequentist)
- Evidential interpretation views $P(A)$ as a means of representing the subjective plausibility of a statement, regardless of whether any random process is involved. (Bayesian)

Axioms of Probability

Regardless of which interpretation you prefer, a probability must satisfy the three axioms of probability (Kolmogorov, 1933), which are the building blocks of all probability theory.

The three probability axioms

1. $P(A) \geq 0$ (non-negativity)
2. $P(S) = 1$ (unit measure)
3. $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$ (additivity)

define a <u>probability measure</u> that makes it possible to calculate the probability of events in a sample space.

Table of Contents

Definition of Probability Distribution

A probability distribution $F(\cdot)$ is a mathematical function that assigns probabilities to outcomes of a simple experiment.

Note that a probability distribution is a function from the sample space S to the interval $[0, 1]$, which can be denoted as $F : S \rightarrow [0, 1]$.

Since $F : S \rightarrow [0, 1]$, we have that $F(s) \geq 0$ and $F(s) \leq 1$ for any $s \in S$.

Probability Distribution Example 1

Consider the coin flipping example with $n = 3$ coin flips. The sample space is $S = \{0, 1, 2, 3\}$.

Assume that the coin is fair, i.e., $P(H) = P(T) = 1/2$, and that the n flips are independent, i.e., unrelated to one another.

Although there are only four elements in the sample space, i.e., $|S| = 4$, there are a total of $2^n = 8$ possible sequences that we could observe when flipping two coins.

Each of the 8 possible sequences is equally likely. Thus, to compute the probability of each $s \in S$, we simply need to count all of the relevant sequences and divide by the total number of possible sequences.

Probability Distribution Example 1 (continued)

Then the probability of each elementary event is as follows:

s	$P(\{s\})$	Observed flip sequence
0	1/8	(T, T, T)
1	3/8	$(H, T, T), (T, H, T), (T, T, H)$
2	3/8	$(H, H, T), (H, T, H), (T, H, H)$
3	1/8	(H, H, H)

Some example probability calculations:

- $P(\{0\} \cap \{3\}) = 0$
- $P(\{0\} \cup \{3\}) = P(\{0\}) + P(\{3\}) = 2/8$
- $P(\{a \mid a \text{ is less than } 2\}) = P(\{0\}) + P(\{1\}) = 4/8$
- $P(\{a \mid a \text{ is less than or equal to } 2\}) = \sum_{s=0}^{2} P(\{s\}) = 1 - P(\{3\}) = 7/8$

Probability Distribution Example 2

Consider the dice rolling example where we sum the numbers of dots on two rolled dice. The sample space is $S = \{2, 3, \ldots, 11, 12\}$.

Assume that the dice are fair, i.e., equal chance of observing each outcome $\{1, \ldots, 6\}$ on a single roll, and that the two rolls are independent, i.e., unrelated to one another.

Although there are only 11 elements in the sample space, i.e., $|S| = 11$, there are a total of $6^2 = 36$ possible sequences that we could observe when rolling two dice.

Each of the 36 possible sequences is equally likely. Thus, to compute the probability of each $s \in S$, we need to count all of the relevant sequences and divide by the total number of possible sequences.

Probability Distribution Example 2 (continued)

Then the probability of each elementary event is as follows:

s	$P(\{s\})$	Observed roll sequence
2	1/36	$(1,1)$
3	2/36	$(1,2)$, $(2,1)$
4	3/36	$(1,3)$, $(2,2)$, $(3,1)$
5	4/36	$(1,4)$, $(2,3)$, $(3,2)$, $(4,1)$
6	5/36	$(1,5)$, $(2,4)$, $(3,3)$, $(4,2)$, $(5,1)$
7	6/36	$(1,6)$, $(2,5)$, $(3,4)$, $(4,3)$, $(5,2)$, $(6,1)$
8	5/36	$(2,6)$, $(3,5)$, $(4,4)$, $(5,3)$, $(6,2)$
9	4/36	$(3,6)$, $(4,5)$, $(5,4)$, $(6,3)$
10	3/36	$(4,6)$, $(5,5)$, $(6,4)$
11	2/36	$(5,6)$, $(6,5)$
12	1/36	$(6,6)$

Probability Distribution Example 2 (continued)

Some example probability calculations:

- $P(\{2\} \cap \{12\}) = 0$
- $P(\{2\} \cup \{12\}) = P(\{2\}) + P(\{12\}) = 2/36$
- $P(\{7\} \cup \{11\}) = P(\{7\}) + P(\{11\}) = 8/36$
- $P(\{a \mid a \text{ is less than } 7\}) = \sum_{s=2}^{6} P(\{s\}) = 15/36$
- $P(\{a \mid a \text{ is an even number}\}) =$
  $P(\{2\}) + P(\{4\}) + P(\{6\}) + P(\{8\}) + P(\{10\}) + P(\{12\}) = 18/36$

Table of Contents

Definition of Joint Event

A joint event refers to an outcome of a simple experiment where the sample point is two-dimensional.

In this case, the sample points have the form $s = (a, b)$, where a and b are the two events that combine to form the joint event.

You can think of a joint event as either:

- a single experiment that produces two outcomes
- a combination of two experiments that each produce an outcome

Joint Event Example 1

Suppose that you flip a coin $n = 2$ times and record the outcome of each coin flip (instead of recording the number of heads).

In this case, the sample space is $S = \{(a, b) \mid a \in \{H, T\}, b \in \{H, T\}\}$, where a and b denote the outcomes of the first and second coin flip.

Note that the sample space has size $|S| = 4$ and the elementary events are defined as $S = \{(T, T), (H, T), (T, H), (H, H)\}$.

Joint Event Example 2

Suppose that you pick a card at random from a standard deck of 52 playing cards and record both the value and suit of the card separately.

In this case, the sample space is
$S = \{(a, b) \mid a \in \{2, 3, \ldots, 9, 10, J, Q, K, A\}, b \in \{C, D, H, S\}\}$.
- C = Club, D = Diamond, H = Heart, S = Spade

Note that the sample space has size $|S| = 52$, given that a could take 13 different values and b could take 4 different values (and $13 \times 4 = 52$).

Joint Event Example 3

Suppose that we roll two dice and record the value of each dice (instead of summing the values).

In this case, the sample space is $S = \{(a, b) \mid 1 \leq a \leq 6, 1 \leq b \leq 6\}$, where a and b denote the outcomes of the first and second dice roll.

Note that the sample space has size $|S| = 36$. See the example on Slide 19 for the 36 elementary events.

Independent Events and Conditional Probability

Two events are independent of one another if the probability of the joint event is the product of the probabilities of the separate events, i.e., if $P(A \cap B) = P(A)P(B)$.

The conditional probability of A given B, denoted as $P(A|B)$, is the probability that A and B occur given that B has occurred, i.e., $P(A|B) = P(A \cap B)/P(B)$.

If A and B are independent of one another, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Knowing that one of the events has occurred tells us nothing about the likelihood of the other event occurring.

Conditional Probability Example 1

For the coin flipping example, if we assume that the coin is fair and the two flips are independent, then $P(s) = (1/2)(1/2) = 1/4$ for any $s \in S$. The sample space is $S = \{(T, T), (H, T), (T, H), (H, H)\}$ and each of the possible outcomes in the sample space is equally likely to occur.

Define the events $A = \{$first flip is heads$\}$, $B = \{$second flip is heads$\}$, and $C = \{$both flips are heads$\}$

Then we have the following probabilities:

- $P(A \cap C) = P(B \cap C) = 1/4$
- $P(A^c \cap C) = P(B^c \cap C) = 0$
- $P(C|A) = P(C|B) = (1/4)/(1/2) = 1/2$
- $P(C|A^c) = P(C|B^c) = 0/(1/2) = 0$

Conditional Probability Example 2

For the card drawing example, note that $P(s) = 1/52$ for any $s \in S$, given that we have equal probability of drawing any card in the deck.

Define $A = \{$the card is a King$\}$ and $B = \{$the card is a face card$\}$.

- $P(A) = 4/52$ given that there are four Kings in a deck
- $P(B) = 12/52$ given that there are 12 face cards in a deck
- $P(A \cap B) = 4/52$ given that $A \subset B$

Then we have the following conditional probabilities:

- $P(A|B) = (4/52)/(12/52) = 4/12 \quad \longrightarrow \quad$ if we draw a face card, then the probability of it being a King is $1/3$
- $P(B|A) = (4/52)/(4/52) = 1 \quad \longrightarrow \quad$ if we draw a King, then it must be a face card

Conditional Probability Example 3

For the dice example, if we assume that the dice are fair and the two rolls are independent, then $P(s) = (1/6)(1/6) = 1/36$ for any $s \in S$.

Define the events $A = \{$the sum of the dice is equal to 7$\}$ and $B = \{$the first dice is a 1 or 2$\}$.

- $P(A) = 6/36$ (see Slide 19)
- $P(B) = 2/6$
- $P(A \cap B) = 2/36$ (see Slide 19)

Then we have the following probabilities:

- $P(A|B) = (2/36)/(2/6) = 2/12$ $\longrightarrow$ if the first roll is 1 or 2, then the probability of the sum being 7 is equal to 1/6
- $P(B|A) = (2/36)/(6/36) = 2/6$ $\longrightarrow$ if the sum of the dice is 7, then the probability of the first roll being 1 or 2 is equal to 1/3

Table of Contents

Bayes' Theorem (due to Reverend Thomas Bayes, 1763)

Bayes' theorem states that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{and} \qquad P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

which is due to the fact that $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$.

This theorem has important consequences because it allows us to derive unknown conditional probabilities from known quantities.

This theorem is the foundation of Bayesian statistics, where the goal is to derive the posterior distribution $P(A|B)$ given the assumed

- distribution for the data given the parameters $P(B|A)$
- prior distribution $P(A)$ of the parameters

Table of Contents

Some Helpful Probability Theory Rules

1. $0 \leq P(A) \leq 1$
2. $P(A^c) = 1 - P(A)$
3. $P(A \cup A^c) = 1$
4. $P(S) = 1$
5. $P(\emptyset) = 1 - P(S) = 0$
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
7. $P(A \cup B) \leq P(A) + P(B)$
8. $P(A \cap B) \leq P(A \cup B)$
9. If $A \subseteq B$, then $P(A) \leq P(B)$
10. If $A \subseteq B$, then $P(B \backslash A) = P(B) - P(A)$
11. $P(A|B) = P(A \cap B)/P(B) = P(B|A)P(A)/P(B)$
12. $P(A|B) = P(A)P(B)$ if A and B are independent