Discrete Probability

Applications

- Genetics – Inheritance Traits and other applications
- Complexity of algorithms…determining average case complexity
- Machine learning
  - Classification based on probability…such as whether an incoming email message is spam or not spam
- Probabilistic algorithms – solve problems not solvable by deterministic algorithms
  - Instead of following the same steps, a probabilistic algorithm makes one or more random choices, which lead to different outputs.
- Simulation and testing with randomized models
- Diagnostic test accuracy

Discrete vs Continuous Probability

- Discrete probability deals with **finite** or **countably infinite** sets

- Countably Infinite – There is a one-to-one correspondence between elements of the set and the integers.

Probability - Introduction

- Simplest definition of probability is when all outcomes are equally likely.

Experiment – A procedure that yields **one** outcome.
Sample Space – The set of **all possible** outcomes.
Event – A subset of the sample space. A set of outcomes.

Probability of event E given a sample space S is:

$$P(E) = \frac{|E|}{|S|}$$

The probability of an event is a number between 0 and 1

This is known as the Laplace definition of probability

- Assuming all outcomes in the sample space **are equally likely.**

Probability of a single outcome in S is:

$$\frac{1}{|S|}$$

Another way to think about it:

$$P(E) = \frac{\text{Number of successful outcomes}}{\text{Number of possible outcomes}}$$

Rosen, Kenneth. *Discrete Mathematics and Its Applications, 7th Edition*. McGraw-Hill, 2012

Example:

Suppose a jar contains 9 Skittles, 4 green and 5 red. What is the probability that a randomly chosen Skittle will be green?

Assume each skittle is equally likely to be chosen.

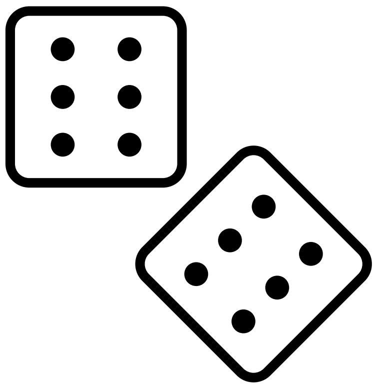
Sample space: $S = \{g1, g2, g3, g4, r1, r2, r3, r4, r5\}$ — All possible outcomes.

Event: $E = \{g1, g2, g3, g4\}$ — A subset of the sample space. Each represents the outcome of a green skittle.

$$P(E) = \frac{|E|}{|S|} = \frac{4}{9} = 0.444444\ldots$$

Rosen, Kenneth. *Discrete Mathematics and Its Applications, 7th Edition*. McGraw-Hill, 2012

What is the probability that when two dice are rolled, the sum of the dice is 7?



What is the sample space S?

$S = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} = \{(1,1), (1,2), (1,3), \ldots, (6,5), (6,6)\}$

$|S| = 36$

What is the event space?

$E = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$

$|E| = 6$

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = \frac{1}{6}$$

Python example

What is the probability of correctly guessing a 4-digit number?

How many possible choices?

$$\_\_ \ \_\_ \ \_\_ \ \_\_$$
$$10 \cdot 10 \cdot 10 \cdot 10$$

$10^4 = 10{,}000$

This is the size of the sample space.

The size of the event space is 1…the event that the correct number is guessed.

Probability of correctly choosing a single 4-digit number is $\frac{1}{10000} = .0001$

What is the probability of correctly choosing only 3 of the 4 digits?

$$9 \cdot 1 \cdot 1 \cdot 1 + 1 \cdot 9 \cdot 1 \cdot 1 + 1 \cdot 1 \cdot 9 \cdot 1 + 1 \cdot 1 \cdot 1 \cdot 9$$

$9 + 9 + 9 + 9 = 36$ ways to choose correctly. $P(E) = \frac{36}{10000}$

Imagine you are participating in a game show where you win **$1,000,000** if you correctly guess a **set** of six numbers between 1 and 40. What are your chances of winning?

There is only one winning set of six numbers, so the size of the event space is 1.

The sample space is the number of ways to choose six numbers out of 40.

$$\binom{40}{6} = \frac{40!}{6!34!} = \frac{40 \cdot 39 \cdot 38 \cdot 37 \cdot 36 \cdot 35}{6!} = 3{,}838{,}380$$

The probability of correctly choosing the six numbers is $\frac{1}{3{,}838{,}380} \approx 0.00000026$

What if you have to guess the six numbers **in order**? (Assume no number will be chosen twice).

$$\frac{1}{P(40,6)} = \frac{1}{40 * 39 * 38 * 37 * 36 * 35} \approx 0.0000000036$$

Python example

Probability distribution

If we sum up the probability of each outcome in the sample space, the sum is 1.

$$\sum_{s \in S} p(s) = 1$$

If each outcome is equally likely, it is a **uniform distribution**.

$$\frac{1}{|S|} + \frac{1}{|S|} + \cdots + \frac{1}{|S|} = \frac{|S|}{|S|}$$

Sometimes the probability distribution is not uniform. This happens when some outcomes are more likely than other outcomes.

For example, if a loaded die is crafted in such a way that rolling a 6 is twice as likely as rolling any other number, then the probability distribution is:

$$p(E) = \sum_{s \in E} p(s)$$

1 2 3 4 5 6

$$\frac{1}{7} + \frac{1}{7} + \frac{1}{7} + \frac{1}{7} + \frac{1}{7} + \frac{2}{7} = 1$$

It is $\frac{1}{7}$ for each of the numbers 1-5 and $\frac{2}{7}$ for the 6 because the sum of the probabilities of each outcome must be equal to 1.

What is the probability of rolling a 6?
$E = \{6\}$,

$$p(E) = \sum_{s \in E} p(s) = \frac{2}{7}$$

What is the probability of rolling an even number?
$E = \{2,4,6\}$,

$$p(E) = \sum_{s \in E} p(s) = \frac{1}{7} + \frac{1}{7} + \frac{2}{7} = \frac{4}{7}$$

Try it:

If we flip a fair coin, what probabilities should we assign to the outcomes H and T?

Since each outcome is equally likely, and there are two possible outcomes:

$p(H) = p(T) = \frac{1}{2}$

$\frac{1}{2} + \frac{1}{2} = 1$ The probability of all outcomes sums to 1.

What if heads is twice as likely as tails?

$p(H) = 2 \cdot p(T)$

What is the probability of each?

$p(H) + p(T) = 1$

$2 \cdot p(T) + p(T) = 3 \cdot p(T) = 1$ Thus, $p(T) = \frac{1}{3}$ and $p(H) = 2 \cdot p(T) = \frac{2}{3}$

Using the complement of a probability.

$P(\bar{E}) = 1 - P(E)$

Suppose we randomly generate a sequence of 10 bits. What is the probability that at least one of these bits is a 1?

$2^{10} = 1024$

0000000000
0000000001
0000000010
...
1111111110
1111111111

Note it is easier to find the probability that none of the bits is a 1.
There is only one case where none of the bits is a 1: 0000000000

$E = \{0000000000\}$ and $P(E) = \frac{|E|}{|S|} = \frac{1}{2^{10}} = \frac{1}{1024}$

$\bar{E} = \{\text{all bitstrings containing at least one 1}\}$ (the opposite of E)

$P(\bar{E}) = 1 - P(E) = 1 - \frac{1}{1024} = \frac{1023}{1024}$

Probabilities for unions of events

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

What is the probability that a randomly chosen number between 1 and 100 (inclusive) is divisible by either 3 or 5?

$E_1 = \{\text{all integers divisible by 3}\} = \{3,6,9,12,15, \dots, 99\}$

$E_2 = \{\text{all integers divisible by 5}\} = \{5,10,15, \dots, 100\}$

$$P(E_1) = \frac{33}{100} \qquad P(E_2) = \frac{20}{100}$$

Note there is some overlap, so we need to remove those outcomes that are counted twice.

$E_1 \cap E_2 = \{15, 30, 45, 60, 75, 90\}$ $\qquad P(E_1 \cap E_2) = \frac{6}{100}$

$$P(E_1 \cup E_2) = \frac{33}{100} + \frac{20}{100} - \frac{6}{100} = \frac{47}{100} = .47$$

Python example

Solve the following using Python:

A coin is flipped four times. For each of the events described below, express the event as a set using roster notation. Each outcome in the sample space can be written as a string of length 4 from {H, T}, such as HHTH. Assuming the coin is a fair coin, give the probability of each event.

a. The first and last flips come up heads.

b. There are at least two consecutive flips that come up heads.

c. The first flip comes up tails and there are at least two consecutive flips that come up heads.

Python example

Additional Exercises:

12.1.1

12.1.2

12.1.4

12.2.1