Week 2: Conditional probability, independent events, Bayes's theorem, discrete random variables

Armenak Petrosyan

Tuesday class

Example from last week

In a class of 100 students, 30 have been to France, 15 have been to Germany, and 7 have been to both countries. We assume everyone in class is equally probable to be selected.

$S =$ all students
$A =$ students that have been to France
$B =$ students that have been to Germany

$$
\begin{aligned}
P(A) &= \frac{\#\text{ been to France}}{\#\text{ all students}} &&= \frac{30}{100}\\
P(B) &= \frac{\#\text{ been to Germany}}{\#\text{ all students}} &&= \frac{15}{100}\\
P(A \cap B) &= \frac{\#\text{ been to Germany and France}}{\#\text{ all students}} &&= \frac{7}{100}
\end{aligned}
$$

Example (new)

In a class of 100 students, 30 have been to France, 15 have been to Germany, and 7 have been to both countries. Assuming everyone in class is equally probable to be selected, what is the probability that a randomly selected student that has been to Germany, has also been to France?

(i.e. we select a student, we disregard them if they haven't been to Germany and among those that have been, we want to know the probability that they have been to France).

Discussion

- Does this change the outcome space? **Yes! The new outcome space =B.**
- Probability $= \frac{\#\text{ been to Germany and France}}{\#\text{ been to Germany}} = \frac{7}{15}$.

This new probability is called "**conditional probability** of event A, given that event B occurred", denoted by

$$P(A|B).$$

Observe that

$$P(A|B) = \frac{\#\text{ been to Germany and France}}{\#\text{ been to Germany}} = \frac{\frac{\#\text{ been to Germany and France}}{\#\text{ all students}}}{\frac{\#\text{ been to Germany}}{\#\text{ all students}}} = \frac{P(A \cap B)}{P(B)}$$

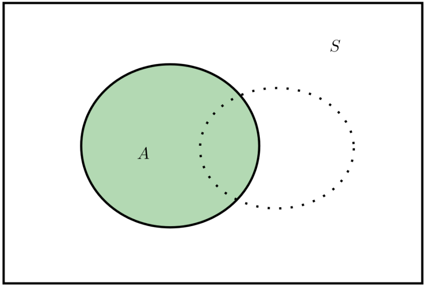
Using this observation, the concept can be extended to any probability space.

Definition

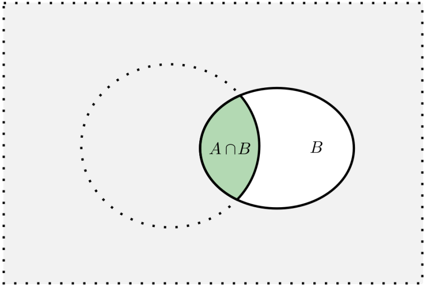
Let P be a probability on the set of outcomes S and $A, B \subset S$ be two events where $P(B) > 0$. The conditional probability of A, given that B occurred, is called the following number

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

$P(A) =$ "proportion of A inside S"



$P(A|B) =$ "proportion of A inside B"



- We introduced conditional probability as an "induced" probability if we restrict the set of outcomes from S to B.
- Notice that, it formally defines a set function on the events in S by the correspondence $A \mapsto P(A|B)$.

Theorem

Let P be a probability on the set of outcomes S, and $P(B) > 0$ for an event $B \subset S$. Then the set function $P(A|B)$ is also a probability on S.

Proof.

We need to check that it satisfies all three conditions in the definition of a probability.

1. $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$ as $P(A \cap B) \geq 0$ and $P(B) > 0$.
2. $P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

□

Proof (cont.)

3. Let $\{A_1, A_2, \ldots\}$ be any (finite or infinite) collection of mutually exclusive events,. Then (using the distributive law of sets; see page 4 in the book)

$$P((A_1 \cup A_2 \cup \cdots)|B) = \frac{P((A_1 \cup A_2 \cup \cdots) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B) \cup \cdots)}{P(B)}.$$

Notice that the events $(A_1 \cap B), (A_2 \cap B), \ldots$ are mutually exclusive hence

$$\begin{aligned} P((A_1 \cup A_2 \cup \cdots)|B) &= \frac{P((A_1 \cup A_2 \cup \cdots) \cap B)}{P(B)} \\ &= \frac{P(A_1 \cap B) + P(A_2 \cap B) + \cdots}{P(B)} \\ &= \frac{P(A_1 \cap B)}{P(B)} + \frac{P(A_2 \cap B)}{P(B)} + \cdots \\ &= P(A_1|B) + P(A_2|B) + \cdots . \end{aligned}$$

Multiplication rule

In some problems you will be given the $P(A|B)$ and $P(B)$ to find the $P(A \cap B)$.

Multiplication rule

For an event B, with $P(B) > 0$,

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

Follows from the definition.

Do not confuse with the Multiplication Principle in Combinatorics

Exercise 1

Problem

There are 7 blue and 3 red balls in an urn. Two are drawn without replacement. What is the probability that the first is blue, the second is red.

Solution

Let us solve using the multiplication rule.

- $P(\textit{first is blue}) = \frac{7}{10}$.
- $P(\textit{second is red} \mid \textit{first is blue}) = \frac{3}{9}$
- *Thus* $P(\textit{first is blue and second is red}) = \frac{7}{10}\frac{3}{9} = \frac{7}{30}$.

Exercise 2

Problem (1.3-7 in the textbook)

A researcher finds that, of 982 men who died in 2002, 221 died from some heart disease. Also, of the 982 men, 334 had at least one parent who had some heart disease. Of the latter 334 men, 111 died from some heart disease. A man is selected from the group of 982. Given that neither of his parents had some heart disease, find the conditional probability that this man died of some heart disease.

Solution

- A = *died from heart disease:* $P(A) = \frac{221}{982}$.
- B = *at least one parent with heart disease:* $P(B) = \frac{334}{982}$.
- $P(A|B) = \frac{111}{334}$.
- *We want to find* $P(A|B')$.

Solution (cont.)

- $$P(A|B') = \frac{P(A \cap B')}{P(B')} = \frac{P(A \cap B')}{1 - P(B)}.$$
- $A \cap B$ and $A \cap B'$ are mutually exclusive and their union is equal to A, hence

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

  or

  $$P(A \cap B') = P(A) - P(A \cap B).$$
- Using the multiplication rule

  $$P(A \cap B) = P(B)P(A|B) = \frac{334}{982}\frac{111}{334} = \frac{111}{982}.$$

Therefore $P(A \cap B') = \frac{221}{982} - \frac{111}{982} = \frac{110}{982}$ which implies

$$P(A|B') = \frac{\frac{110}{982}}{1 - \frac{334}{982}} = \frac{110}{648}.$$

Combining multiple experiments

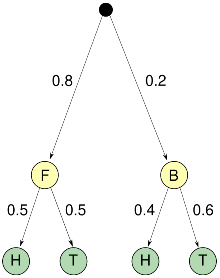
- We have two experiments exp_1, exp_2 with corresponding sets of all outcomes S_1, S_2.
- Combine them into a single experiment.
- If the order of the experiments matters, the outcomes of the new experiment will consist of pairs (s_1, s_2), $s_1 \in S_1, s_2 \in S_2$.
- Let A be an event in S_1 and B be an event in S_2.
- If the outcome of the first experiment affects the outcome of the second in this new experiment, with an *abuse of notation* we can talk about conditional probability $P(B|A)$.

Example

We have five coins in the pocket. 4 of them are fair, one of them is biased with $P(H) = 0.4$, $P(T) = 0.6$. We randomly pick a coin from the pocket, toss it. Find the probability that the outcome was a head.

Solution

- $P(\textit{biased and } H) = P(\textit{biased})P(H|\textit{biased}) = 0.2 \cdot 0.4 = 0.08.$
- $P(\textit{fair and } H) = P(\textit{fair})P(H|\textit{fair}) = 0.8 \cdot 0.5 = 0.4.$
- $P(H) = P(\textit{biased and } H) + P(\textit{fair and } H) = 0.48.$



- Possible to combine multiple experiments with different pairings (some with order, some not) into a single experiment.
- After specifying dependencies, this leads to so called *Bayesian networks* (is beyond the scope of our class).

Definition

Events A and B are called independent if

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

Observe that events A, B being independent is equivalent to any of these two statements (when $P(A), P(B) > 0$):

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

In other words, if A and B are independent then

1. knowing that the event B occurred does not change the probability of A, and
2. knowing that the event A occurred does not change the probability of B.

Theorem

If events A and B are independent then so are the pair A and B'.

Proof.

- We saw that $P(A \cap B) + P(A \cap B') = P(A)$.
- Hence,

$$\begin{aligned} P(A \cap B') &= P(A) - P(A \cap B) = P(A) - P(A)P(B) \\ &= P(A)(1 - P(B)) = P(A)P(B'). \end{aligned}$$

$\square$

Consecutively, the pairs of events $\{A', B\}$ and $\{A', B'\}$ will also be independent.

Exercise 3

Problem

Consider the equiprobability space for the 2 coin toss experiment

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

Then the value of the first toss is independent of the value of the second toss.

Solution

Let $A =$\{*second toss is head*\} *and* $B =$\{*first toss is tail*\}.

- *Notice that* $A' =$\{*second toss is tail*\} *and* $B' =$\{*first toss is head*\}.
- *So we only need to show the events* A *and* B *are independent; the independence for all other pairs of events will follow from previous theorem.*
- *We have*

$$A = \{HH, TH\} \qquad B = \{TH, TT\} \qquad A \cap B = \{TH\}$$
$$P(A) = \frac{2}{4} \qquad P(B) = \frac{2}{4} \qquad P(A \cap B) = \frac{1}{4}.$$

- *Hence* $P(A \cap B) = P(A)P(B)$.

Exercise 4

Problem (1.4-11 from the textbook)

Let A and B be two events.

(a) *If A and B are mutually exclusive, are A and B always independent? If the answer is no, can they ever be independent?*

(b) *If $A \supseteq B$, can A and B ever be independent events?*

Solution

(a)
- $A \cap B = \emptyset$, *so if* $P(A) > 0$ *and* $P(B) > 0$ *then impossible.*
- *If* $P(A) = 0$ *or* $P(B) = 0$ *then* $P(A)P(B) = P(A \cap B) = P(\emptyset) = 0$.

(b)
- *If* $A \supseteq B$ *then* $P(A \cap B) = P(B)$.
- *Independent if* $P(B) = P(A)P(B)$.
- *Happens only if* $P(B) = 0$ *or* $P(A) = 1$.

Pairwise and mutual independence

Definition

Events A, B, C are said to be pairwise independent if

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

Definition

Events A, B, C are said to be mutually independent if they are pairwise independent and

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

Exercise 5

Problem

A fair coin is flipped three times. Let

$$\begin{aligned} A &= \{\textit{tosses 1 and 2 are equal}\}, \\ B &= \{\textit{tosses 2 and 3 are equal}\}, \\ C &= \{\textit{tosses 1 and 3 are equal}\}. \end{aligned}$$

Show that events A, B, C are pairwise independent but not mutually independent.

Solution

- $P(A) = P(B) = P(C) = \frac{2^2}{2^3} = \frac{1}{2}$.
- $A \cap B = \{HHH, TTT\}$ *and so* $P(A \cap B) = \frac{2}{2^3} = \frac{1}{4}$.
- *Hence* $P(A \cap B) = P(A)P(B)$.
- *The independence of* $\{B, C\}$ *and* $\{A, C\}$ *follows the same way.*
- *However* $A \cap B \cap C = \{HHH, TTT\}$ *therefore*

$$P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C).$$

Extension to multiple events

Definition

Events $A_1, \ldots, A_n$ are called mutually independent if for any $1 \le k \le n$ and any $1 \le i_1 < \cdots i_k \le n$,

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_n}).$$

The space of independent trials

Let S be a set of outcomes and P be probability on S.

- Denote by S^n the set of all ordered samples from S of size n with replacement (i.e. combining n-duplicates of the same experiment).
- Introduce S^n as a new set of outcomes.
- For $s = (s_1, \ldots, s_n) \in S^n$, define

$$P_n(s) = P(s_1) \cdots P(s_n).$$

- The probability of any event $A \subset S^n$ is defined as the sum of probabilities of the outcomes in A.

For an outcome $s = (s_1, \ldots, s_n) \in S^n$, the value s_i is called the i-th trial, $i = 1, \ldots, n$.

It can be checked that the values of all n trials as events are mutually independent by design.

The equiprobability space of n fair coin toss experiment is the same as the probability space of n independent trials of the single coin toss experiment.

Exercise 6

Problem (1.4-13 from the textbook)

An urn contains two red balls and four white balls. Sample successively five times at random and with replacement, so that the trials are independent. Compute the probability of each of the two sequences WWRWR and RWWWR.

Solution

From the definition above,

- $P(WWRWR) = P(W)P(W)P(R)P(W)P(R) = \frac{4}{6}\frac{4}{6}\frac{2}{6}\frac{4}{6}\frac{2}{6} = \frac{8}{3^5}$.
- $P(RWWWR) = P(R)P(W)P(W)P(W)P(R) = \frac{2}{6}\frac{4}{6}\frac{4}{6}\frac{4}{6}\frac{2}{6} = \frac{8}{3^5}$.

Thursday class

Bayes's theorem, Bayesian probabilities

Bayes's formula

Theorem (Bayes's formula)

Let A, B be two events in the outcome set S with $P(A) > 0$ and $P(B) > 0$. Then

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

Proof.

- From the multiplication rule

$$P(A \cap B) = P(B)P(A|B) \text{ and, similarly, } P(B \cap A) = P(A)P(B|A).$$

- Equating both sides, we have

$$P(B)P(A|B) = P(B)P(B|A)$$

or

$$P(A|B) = \frac{P(A)P(A|B)}{P(B)}.$$

□

Let $B_1, \dots, B_n$ be mutually exclusive, exhaustive collection of events in S, i.e.

$$B_i \cap B_j = \emptyset \text{ for } i \neq j \text{ and } B_1 \cap \cdots \cap B_n = S.$$

In this case we call this collection a **partition** of S.

Theorem (Law of total probability)

Let $B_1, \dots, B_n$ be a partition of S such that $P(B_i) > 0$ for every $i = 1, \dots, n$. Then, for any event $A \subset S$,

$$P(A) = P(B_1)P(A|B_1) + \cdots + P(B_n)P(A|B_n) = \sum_{i=1}^{n} P(B_i)P(A|B_i).$$

Proof.

- Notice that $A \cap B_1, \dots, A \cap B_n$ are mutually exclusive and

$$(A \cap B_1) \cup \cdots \cup (A \cap B_n) = A \cap (B_1 \cup \cdots \cup B_n) = A \cap S = A.$$

- Therefore, from the definition of probability and the multiplication rule,

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

□

Theorem (Bayes's theorem)

Let $B_1, \ldots, B_n$ be a partition of S such that $P(B_i) > 0$ for every $i = 1, \ldots, n$. Then, for any event $A \subset S$ and any $k = 1, \ldots, n$

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}.$$

Proof.

- From Bayes's formula

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{P(A)}.$$

- Plugging the value of $P(A)$ from the law of total probability, we get

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)}.$$

$\square$

Bayesian probabilities

In Bayesian theory, the events $B_1, \ldots, B_n$ in the partition are usually called **hypothesis** and the set A is called **data**.

- Probabilities $P(B_k)$ are called **prior** probabilities.
- Probabilities $P(B_k|A)$ are called **posterior** probabilities.
- Probabilities $P(A|B_k)$ are called **likelihoods**.

The transition from priors to posteriors via Bayes's theorem is called **Bayesian update**.

Exercise 7

Problem (1.5-5 from the textbook)

At a hospital's emergency room, patients are classified and 20% of them are critical, 30% are serious, and 50% are stable. Of the critical ones, 30% die; of the serious, 10% die; and of the stable, 1% die. Given that a patient dies, what is the conditional probability that the patient was classified as critical?

Solution

- $A = \{$*patient dies*$\}$, $B_1 =\{$*critical*$\}$, $B_2 =\{$*serious*$\}$, $B_1 =\{$*stable*$\}$.
- $P(B_1) = 0.2$, $P(B_2) = 0.3$, $P(B_3) = 0.5$.
- $P(A|B_1) = 0.3$, $P(A|B_2) = 0.1$, $P(A|B_3) = 0.01$.
- *We want to compute* $P(B_1|A)$.
- *Using Bayes's theorem,*

$$\begin{aligned} P(B_1|A) &= \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} \\ &= \frac{0.2 \cdot 0.3}{02 \cdot 0.3 + 0.3 \cdot 0.1 + 0.5 \cdot 0.01} = \frac{0.06}{0.095} \approx 0.63. \end{aligned}$$

Exercise 8

Problem (1.5-9 in the textbook)

There is a new diagnostic test for a disease that occurs in about 0.05% of the population. The test is not perfect, but will detect a person with the disease 99% of the time. It will, however, say that a person without the disease has the disease about 3% of the time. A person is selected at random from the population, and the test indicates that this person has the disease. What are the conditional probabilities that

(a) *the person has the disease?*

(b) *the person does not have the disease?*

Solution

- $A = \{$*test positive*$\}$ $B = \{$*has disease*$\}$.
- $P(B) = 0.0005$, $P(A|B) = 0.99$, $P(A|B') = 0.03$.
- *We are asked to compute* $P(B|A)$ *in (a) and* $P(B'|A)$ *in (b).*
- *From Bayes's theorem*

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B')P(A|B')} = \frac{0.0005 \cdot 0.99}{0.0005 \cdot 0.99 + (1 - 0.0005)0.03} \approx 0.016$$

$$P(B'|A) = 1 - P(B|A) \approx 0.984.$$

Discussion

In the previous exercise

- True positives: $P(B|A) = 0.016$.
- False positives: $P(B'|A) = 0.984$.

Proportion of false positives is overwhelmingly large. Is something wrong? This is called **base rate fallacy**. Happens when

- There is a large population.
- Small part of the population has the disease.
- There is a large scale testing of mixed population of healthy and sick individuals.

Even for a very accurate test the number of false positives can potentially be large as was the case in the exercise.

WARNING

Please, do not make wrong conclusions from here; keep social distance and see a doctor when feeling sick.

Discrete random variables

Definition

Let S be the set of all outcomes. Any function $X : S \to \mathbb{R}$ is called a **random variable**.

Random variable assigns a numerical value (measurement) to every outcome.

Example

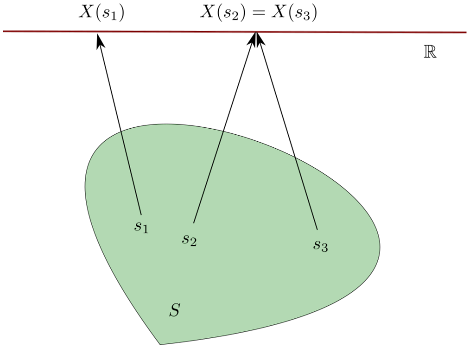
- Consider the double coin flip experiment $S = \{HH, HT, TH, TT\}$.
- Define $X(s) :=$ number of heads in s.
- Then
$$X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0.$$

Definition

The set
$$\mathrm{Range}(X) := \{X(s) :\ s \in S\}$$
is called the **range (or space)** of the random variable X.

In the above example
$$\mathrm{Range}(X) = \{0, 1, 2\}.$$

Different outcomes can have the same measurement.



Definition

If $\mathrm{Range}(X)$ is finite or countably infinite, then the random variable is called **discrete.**

- A set is called countably infinite, if its elements can be arranged into a sequence.
- The set of all real numbers or the interval $[0, 1]$ are not countably infinite (Cantor's theorem).

Example

Let S be the set of all living creatures. The duration of their lives is a continuous random variable because it can be any non-negative number (in seconds).

Probability mass function (pmf)

Definition

For every $x \in \mathrm{Range}(X)$, define

$$f(x) = P(\{s \in S : \ X(s) = s\}).$$

$f(x)$ is called **probability mass function** (pmf) or **probability density function** (pdf).

$P(\{s \in S : \ X(s) = s\})$ is often denoted as $P(X = x)$.

Example

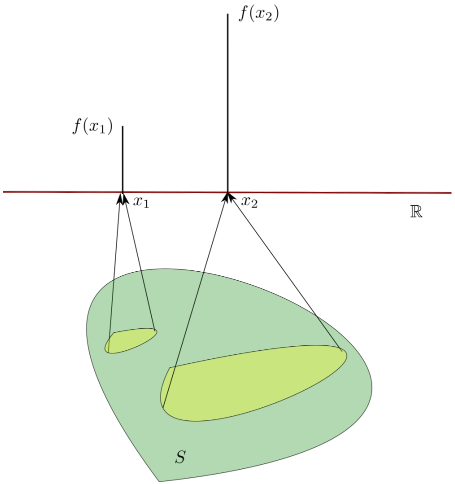
In the double coin flip example $\mathrm{Range}(X) = \{0, 1, 2\}$, and

$$f(0) = P(TT) = \frac{1}{4}$$

$$f(1) = P(\{HT, TH\}) = \frac{2}{4}$$

$$f(2) = P(HH) = \frac{1}{4}$$

- Two different random variables can have the same pmf.
- In the previous example take $X(s)$ to be the number of tails instead.


$f(x_2)$
$f(x_1)$
x_1
x_2
$\mathbb{R}$
S

- pmf is a function defined on $\mathrm{Range}(X)$.
- We often extend f as a function on $\mathbb{R}$ by defining

$$f(x) = 0, \; x \notin \mathrm{Range}(X).$$

- $\mathrm{Range}(X)$ is called the **support** of f.

Theorem

Notice that the pmf has the following properties

1. $f(x) \geq 0$ *for every* $x \in \mathbb{R}$,
2. $\sum_{x \in \mathrm{Range}(X)} f(x) = 1$,
3. *For any* $A \subset \mathrm{Range}(X)$,

$$P(X \in A) = \sum_{x \in A} f(x).$$

Proof.

1. Follows from condition 1 in the definition of probability.
2. From condition 2 in the definition of probability as $\sum_{x \in \mathrm{Range}(X)} f(x) = P(S)$.
3. From condition 3 in the definition of probability.

$\square$

Exercise 9

Problem

Consider the equiprobability space on $S = \{(i,j):\ 1 \le i \le n, 1 \le j \le n\}$ and for any $s = (i,j)$ let

$$X(i,j) = \max\{i,j\}.$$

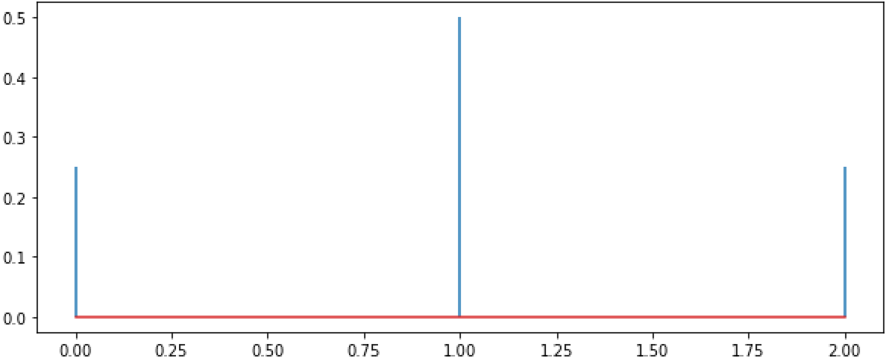
(tossing two dice each with n faces and choosing the largest value). Find the pmf of X.

Solution

- $\mathrm{Range}(X) = \{1, \ldots, n\}$.
- *Notice $X(i,j) = k$ if and only if $i = k$ or $j = k$ and $i \le k$, $j \le k$.*
- *If $i = k$, j can take values $1, \ldots k$ so number of such pairs is k.*
- *When $j = k$, similarly there are k values of i.*
- *We counted (k,k) twice so total number of (i,j) with $\max\{i,j\} = k$ is $2k - 1$.*
- *The size sample set S is n^2 (sample with replacement).*
- *Hence $f(k) = \frac{2k-1}{n^2}$*

Graphical representations of pmf: line graph

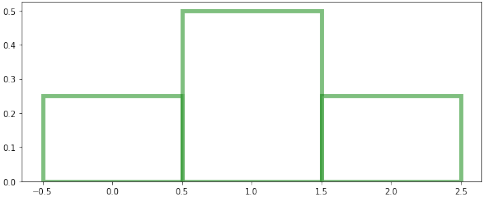
Line graph for the double coin flip experiment with the random variable defined as the number of heads.



Graphical representations of pmf: histogram

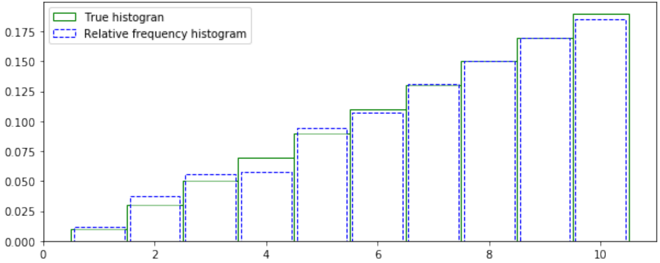
- Histogram is very similar to the line graph, but the lines are replaced with boxes.
- When $\mathrm{Range}(X)$ is the set $\{0, 1, \dots, k\}$, the bottom of the box is chosen to be of size 1.
- In that case the total area of boxes is equal to $\displaystyle\sum_{x \in \mathrm{Range}(X)} f(x) = 1.$

Histogram for the double coin flip experiment with the random variable defined as the number of heads.



Frequency histogram of data

- As mentioned earlier, probability represents how frequent a specific outcome comes up in large number of repetitive trials.
- If we do many trials and compute the relative frequency of a specific value in $\mathrm{Range}(X)$, we will have an approximation to the histogram of the pmf.



- 10-faced die double toss example for maximum value random variable.
- Data from 1000 pairs of tosses.
- relative frequency at $k := \frac{\#\text{ times X}=k}{\#\text{data}}$