Basic properties of probability

Math 308

Definition: Let S be a sample space. A probability on S is a real valued function P,

$$P : \{\text{Events}\} \to \mathbb{R},$$

satisfying:

1. $P(A) \geq 0$ for any event A.
2. $P(S) = 1$.
3. If $A_1, A_2, \ldots$ are **mutually exclusive events** (m.e.e) then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i),$$

   where by **m.e.e.** we mean $A_i \cap A_j = \emptyset$ when $i \neq j$.

Basic properties of probability:

1. $P(\emptyset) = 0$.
2. Let $A_1, A_2, \ldots A_n$ be **m.e.e.**, then

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

3. $P(A') = 1 - P(A)$.
4. If $A \subset B$ then $P(A) \leq P(B)$.
   In particular, if B is S, we get $0 \leq P(A) \leq 1$ for any event A.
5. Let $B_1, B_2, \ldots$ be **m.e.e.** such that $S = \cup_{i=1}^{\infty} B_i$, that is, the B_i's form a **partition** of S. Then for any event A,

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

6. It follows from (5) that for any event B, since $S = B \cup B'$ is a **partition** of S, then

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

   In particular, since $A \cap B' = A \setminus B$ (draw the Venn diagram), we have

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

7. For any events A and B,

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

8. (probability of a finite union)

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} \cap A_{i_2}) + \sum_{i_1 < i_2 < i_3} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) - \cdots + (-1)^{n+1} P(\cap_{i=1}^{n} A_i).$$

In particular,

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

9. Using the basic properties (and Venn diagrams) you can find formulas for probabilities of other operations on sets. For example, if A and B are events, then the probability that event A occur **or** B occur, but **not both** is

$$P\left((A \cup B) \setminus (A \cap B)\right) = P\left((A \setminus B) \cup (B \setminus A)\right) = P(A) + P(B) - 2P(A \cap B).$$

Note that the last equality follows from property (2) since $(A \setminus B)$ and $(B \setminus A)$ are **m.e.e.** (we also used property (6)).

1. **QUESTION**:
Describe the sample space and all 16 events for a trial in which two coins are thrown and each shows either a *head* or a *tail*.

SOLUTION:
The sample space is $\mathcal{S} = \{hh,\ ht,\ th,\ tt\}$. As this has 4 elements there are $2^4 = 16$ subsets, namely ϕ, hh, ht, th, tt, $\{hh,\ ht\}$, $\{hh,\ th\}$, $\{hh,\ tt\}$, $\{ht,\ th\}$, $\{ht,\ tt\}$, $\{th,\ tt\}$, $\{hh,\ ht,\ th\}$, $\{hh,\ ht,\ tt\}$, $\{hh,\ th,\ tt\}$, $\{ht,\ th,\ tt\}$ and finally $\{hh,\ ht,\ th,\ tt\}$.

2. **QUESTION**:
A fair coin is tossed, and a fair die is thrown. Write down sample spaces for

(a) the toss of the coin;
(b) the throw of the die;
(c) the combination of these experiments.

Let A be the event that a head is tossed, and B be the event that an odd number is thrown. Directly from the sample space, calculate $\mathrm{P}(A \cap B)$ and $\mathrm{P}(A \cup B)$.

SOLUTION:

(a) $\{Head, Tail\}$
(b) $\{1, 2, 3, 4, 5, 6\}$
(c) $\{(1 \cap Head), (1 \cap Tail), \ldots, (6 \cap Head), (6 \cap Tail)\}$

Clearly $\mathrm{P}(A) = \frac{1}{2} = \mathrm{P}(B)$. We can assume that the two events are independent, so

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

Alternatively, we can examine the sample space above and deduce that three of the twelve equally likely events comprise $A \cap B$.
Also, $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = \frac{3}{4}$, where this probability can also be determined by noticing from the sample space that nine of twelve equally likely events comprise $A \cup B$.

3. **QUESTION**:
A bag contains fifteen balls distinguishable only by their colours; ten are blue and five are red. I reach into the bag with both hands and pull out two balls (one with each hand) and record their colours.

(a) What is the *random phenomenon*?
(b) What is the *sample space*?
(c) Express the *event* that the ball in my left hand is red as a subset of the sample space.

SOLUTION:

(a) The random phenomenon is (or rather the phenomena are) the colours of the two balls.
(b) The sample space is the set of all possible colours for the two balls, which is

$$\{(B, B), (B, R), (R, B), (R, R)\}.$$

(c) The event is the subset $\{(R, B), (R, R)\}$.

4. **QUESTION**:
M&M sweets are of varying colours and the different colours occur in different proportions. The table below gives the probability that a randomly chosen M&M has each colour, but the value for tan candies is missing.

Colour	Brown	Red	Yellow	Green	Orange	Tan
Probability	0.3	0.2	0.2	0.1	0.1	?

(a) What value must the missing probability be?

(b) You draw an M&M at random from a packet. What is the probability of each of the following events?

i. You get a brown one or a red one.
ii. You don't get a yellow one.
iii. You don't get either an orange one or a tan one.
iv. You get one that is brown or red or yellow or green or orange or tan.

SOLUTION:

(a) The probabilities must sum to 1.0 Therefore, the answer is $1-0.3-0.2-0.2-0.1-0.1 = 1-0.9 = .1$.

(b) Simply add and subtract the appropriate probabilities.

i. $0.3 + 0.2 = 0.5$ since it can't be brown and red simultaneously (the events are incompatible).
ii. $1 - \mathrm{P(yellow)} = 1 - 0.2 = 0.8$.
iii. $1 - \mathrm{P(orange\quad or\quad tan)} = 1 - \mathrm{P(orange)} - \mathrm{P(tan)} = 1 - 0.1 - 0.1 = 0.8$ (since orange and tan are incompatible events).
iv. This must happen; the probability is 1.0

5. **QUESTION**:
You consult Joe the bookie as to the form in the 2.30 at Ayr. He tells you that, of 16 runners, the favourite has probability 0.3 of winning, two other horses each have probability 0.20 of winning, and the remainder each have probability 0.05 of winning, excepting Desert Pansy, which has a worse than no chance of winning. What do you think of Joe's advice?

SOLUTION:
Assume that the sample space consists of a win for each of the 16 different horses. Joe's probabilities for these sum to 1.3 (rather than unity), so Joe is *incoherent*, albeit profitable! Additionally, even "Dobbin" has a non-negative probability of winning.

6. **QUESTION**:
Not all dice are fair. In order to describe an unfair die properly, we must specify the probability for each of the six possible outcomes. The following table gives answers for each of 4 different dice.

	Probabilities			
Outcome	Die 1	Die 2	Die 3	Die 4
1	1/3	1/6	1/7	1/3
2	0	1/6	1/7	1/3
3	1/6	1/6	1/7	-1/6
4	0	1/6	1/7	-1/6
5	1/6	1/6	1/7	1/3
6	1/3	1/7	2/7	1/3

Which of the four dice have validly specified probabilities and which do not? In the case of an invalidly described die, explain why the probabilities are invalid.

SOLUTION:

(a) Die 1 is valid.
(b) Die 2 is invalid; The probabilities do not sum to 1. In fact they sum to 41/42.
(c) Die 3 is valid.
(d) Die 4 is invalid. Two of the probabilities are negative.

7. **QUESTION**:
A six-sided die has four green and two red faces and is balanced so that each face is equally likely to come up. The die will be rolled several times. You must choose one of the following three sequences of colours; you will win £25 if the first rolls of the die give the sequence that you have chosen.

R G R R R
R G R R R G
G R R R R R

Without making any calculations, explain which sequence you choose. (In a psychological experiment, 63% of 260 students who had not studied probability chose the second sequence. This is evidence that our intuitive understanding of probability is not very accurate. This and other similar experiments are reported by A. Tversky and D. Kahneman, *Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment*, Psychological Review **90** (1983), pp. 293–315.)

SOLUTION:
Without making calculations, the sequences are identical except for order for the first five rolls. Consequently, these sequences have the same probability up to and including the first five rolls. The second and third sequences must now be less probable than the first, as an extra roll, with probability less than one, is involved. Hence the first sequence is the most probable.
Calculation requires the notion of independence. Two methods. Firstly, work out the probabilities for the sequences: The probability of a red on an individual roll is $\frac{2}{6} = \frac{1}{3}$ and the probability of a green is $\frac{2}{3}$. Hence, since successive rolls are independent, the probability of the first sequence is

$$\frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{243} = 0.0082.$$

Similarly the probabilities of the other two sequences are

$$\frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} = \frac{4}{729} = 0.0055,$$

and

$$\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{2}{729} = 0.0027.$$

The sequence with highest probability is the first one. For a second method, reason as follows. All three sequences begin with five rolls containing one green and four reds. The order in which these green and reds occur is irrelevent, because of independence. So, let H be the event that we obtain one green and four reds in the first five rolls. The three sequences are now H, HG, and HR, with probabilities $\mathrm{P}(H)$, $\mathrm{P}(H)\mathrm{P}(G) = \frac{2}{3}\mathrm{P}(H)$, and $\mathrm{P}(H)\mathrm{P}(R) = \frac{1}{3}\mathrm{P}(H)$. Clearly, the first sequence is more probable than the second, which is more probable than the third.

8. **QUESTION**:
Suppose that for three dice of the standard type all 216 outcomes of a throw are equally likely. Denote the scores obtained by X_1, X_2 and X_3. By counting outcomes in the events find (a) $P(X_1+X_2+X_3 \leq 5)$; (b) $P(\min(X_1, X_2, X_3) \geq i)$ for $i = 1, 2, \ldots 6$; (c) $P(X_1 + X_2 < (X_3)^2)$.

SOLUTION:

(a) There are 216 equally likely triples and of these only 10 have a sum ≤ 5 so $P(X_1 + X_2 + X_3 \leq 5) = 10/216$.

(b) The smallest of three numbers is bigger than i only when all three are so $P(\min(X_1, X_2, X_3) \geq i) = P(X_1 \geq i, X_2 \geq i, X_3 \geq i) = (7 - i)^3/216$ (picture this group as a cube within the bigger cube of all 216 states).

(c) Of 36 triples with $X_3 = 2$ only 3 have $X_1 + X_2 < 4$ and of 36 triples with $X_3 = 3$, 26 have $X_1 + X_2 < 9$ so that $P((X_3)^2 > X_1 + X_2) = \sum_j P(X_1 + X_2 < j^2, X_3 = j) = 137/216$.

9. **QUESTION**:
You play draughts against an opponent who is your equal. Which of the following is more likely: (a) winning three games out of four or winning five out of eight; (b) winning at least three out of four or at least five out of eight?

SOLUTION:
Let X and Y be the numbers of wins in 4 and 8 games respectively. For 4 games there are $2^4 = 16$ equally likely outcomes e.g. $WLWW$ which has 3 wins so $X = 3$. Using our basic counting principles there will be $\binom{4}{j}$ outcomes containing j wins and so $P(X = 3) = 4 \times 0.5^4 = 0.25$.

Similarly with 8 games there are $2^8 = 256$ equally likely outcomes and this time $P(Y = 5) = 56 \times 0.5^8 = 0.2188$ so the former is larger.

For part (b) remember that $X \geq 3$ means all the outcomes with at least 3 wins out of 4 etc and that we sum probabilities over mutually exclusive outcomes. Doing the calculations, $P(X \geq 3) = 0.25 + 0.0625 = 0.3125$ is less than $P(Y \geq 5) = 0.2188 + 0.1094 + 0.0313 + 0.0039 = 0.3633$ – *we deduce from this that the chance of a drawn series falls as the series gets longer.*

10. **QUESTION**:
Count the number of distinct ways of putting 3 balls into 4 boxes when:

- *MB* all boxes and balls are distinguishable;
- *BE* the boxes are different but the balls are indentical;
- *FD* the balls are identical, the boxes are different but hold at most a single ball.

See if you can do the counting when there are m balls and n boxes.

SOLUTION:
#(MB)= $4^3 = 64$, #(BE)= $\binom{6}{3} = 20$, #(FD)= $\binom{4}{3} = 4$. The general cases are n^m, $\binom{m+n-1}{m}$ (i.e. arrangements of balls and fences), $\binom{n}{m}$.

11. **QUESTION**:
A lucky dip at a school fête contains 100 packages of which 40 contain tickets for prizes. Let X denote the number of prizes you win when you draw out three of the packages. Find the probability density of X i.e. $P(X = i)$ for each appropriate i.

SOLUTION:
There are $\binom{100}{3}$ choices of three packages (in any ordering). There are $\binom{60}{3}$ choices of three packages without prizes. Hence $P(X = 0) = \binom{60}{3}/\binom{100}{3} \approx 0.2116$. If a single prize is won this can happen in $\binom{40}{1} \cdot \binom{60}{2}$ ways. Hence $P(X = 1) = \binom{40}{1} \cdot \binom{60}{2}/\binom{100}{3} \approx 0.4378$ and similarly $P(X = 2) = \binom{40}{2} \cdot \binom{60}{1}/\binom{100}{3} \approx 0.2894$ and $P(X = 3) = \binom{40}{3}/\binom{100}{3} \approx 0.0611$ (there is some small rounding error in the given values).

12. **QUESTION**:
Two sisters maintain that they can communicate telepathically. To test this assertion, you place the sisters in separate rooms and show sister A a series of cards. Each card is equally likely to depict either a circle or a star or a square. For each card presented to sister A, sister B writes down 'circle', or 'star' or 'square', depending on what she believes sister A to be looking at. If ten cards are shown, what is the probability that sister B correctly matches at least one?

SOLUTION:
We will calculate a probability under the assumption that the sisters are guessing. The probability of at least one correct match must be equal to one minus the probability of no correct matches. Let F_i be the event that the sisters fail to match for the ith card shown. The probability of no correct matches is $\mathrm{P}(F_1 \cap F_2 \cap \ldots \cap F_{10})$, where $\mathrm{P}(F_i) = \frac{2}{3}$ for each i. If we assume that successive attempts at matching cards are independent, we can multiply together the probabilities for these independent events, and so obtain

$$\mathrm{P}(F_1 \cap F_2 \cap \ldots \cap F_{10}) = \mathrm{P}(F_1)\mathrm{P}(F_2)\ldots\mathrm{P}(F_{10}) = (\frac{2}{3})^{10} = 0.0173.$$

Hence the probability of at least one match is $1 - 0.0173 = 0.9827$.

13. **QUESTION**:
An examination consists of multiple-choice questions, each having five possible answers. Suppose you are a student taking the exam. and that you reckon you have probability 0.75 of knowing the answer to any question that may be asked and that, if you do not know, you intend to guess an answer with probability 1/5 of being correct. What is the probability you will give the correct answer to a question?

SOLUTION:
Let A be the event that you give the correct answer. Let B be the event that you knew the answer. We want to find $\mathrm{P}(A)$. But $\mathrm{P}(A) = \mathrm{P}(A \cap B) + \mathrm{P}(A \cap B^c)$ where $\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B) = 1 \times 0.75 = 0.75$ and $\mathrm{P}(A \cap B^c) = \mathrm{P}(A|B^c)\mathrm{P}(B^c) = \frac{1}{5} \times 0.25 = 0.05$. Hence $\mathrm{P}(A) = 0.75 + 0.05 = 0.8$.

14. **QUESTION**:
Consider the following experiment. You draw a square, of width 1 foot, on the floor. Inside the square, you inscribe a circle of diameter 1 foot. The circle will just fit inside the square.
You then throw a dart at the square in such a way that it is equally likely to fall on any point of the square. What is the probability that the dart falls inside the circle? (Think about area!)
How might this process be used to estimate the value of π?

SOLUTION:
All points in the square are equally likely so that probability is the ratio of the area of the circle to the area of the square. The area of the square is 1 and the area of the circle is $\pi/4$ (since the radius is 1/2). If you don't know π you can estimate it by repeating the experiment a very large number of times. Then π will be approximately the same as the proportion of times the dart fall in the circle multiplied by 4.

15. **QUESTION**:
I have in my pocket ten coins. Nine of them are ordinary coins with equal chances of coming up head and tail when tossed and the tenth has two heads.

   (a) If I take one of the coins at random from my pocket, what is the probability that it is the coin with two heads ?
   (b) If I toss the coin and it comes up heads, what is the probability that it is the coin with two heads ?
   (c) If I toss the coin one further time and it comes up tails, what is the probability that it is one of the nine ordinary coins ?

SOLUTION:
Denote by D the event that the coin is the one with two heads.

   (a) $\mathrm{P}(D) = 1/10$.

(b) Denote by H the event that we get a head when we toss the coin. Then we want to find $P(D|H)$. By Bayes theorem, we have

$$P(D|H) = \frac{P(H|D)P(D)}{P(H)}.$$

We have $P(H|D) = 1$ and $P(D) = \frac{1}{10}$. Now, we need to think about H, getting a head, in terms of getting a head with either a double headed or single headed coin. Using the idea of a partition,

$$\begin{aligned} P(H) &= P(H \cap D) + P(H \cap D^c) \\ &= P(H|D)P(D) + P(H|D^c)P(D^c) \\ &= (1)(\frac{1}{10}) + (\frac{1}{2})(\frac{9}{10}) \\ &= \frac{11}{20}. \end{aligned}$$

Finally, here is another way of calculating $P(H)$: think of the bag as containing the possible tosses. As the bag contains 9 fair coins and one double-headed coin, it must contains 11 heads and 9 tails, so that the probability of choosing a head is $11/(11+9) = 11/20$.
To return to the original question, we now obtain the answer

$$P(D|H) = \frac{\frac{1}{10}}{\frac{11}{20}} = \frac{2}{11}.$$

(c) 1. If it comes up tails, it can't be the coin with two heads. Therefore it must be one of the other nine.

16. **QUESTION**:
Let A, B and C be any three events. Draw Venn diagrams to deduce that

(a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$;
(b) $(A \cap B)' = A' \cup B'$;
(c) $(A \cup B)' = A' \cap B'$.

SOLUTION:
Just draw the picture!

17. **QUESTION**:
A certain person considers that he can drink and drive: usually he believes he has a negligible chance of being involved in an accident, whereas he believes that if he drinks two pints of beer, his chance of being involved in an accident on the way home is only one in five hundred. Assuming that he drives home from the same pub every night, having drunk two pints of beer, what is the chance that he is involved in at least one accident in one year? Are there any assumptions that you make in answering the question?

SOLUTION:
We must assume that each drive home is *independent* of any other drive home. Write $A_i; i = 1, \ldots, 365;$ to be the event that our driver is *not* involved in an accident on day i, with $P(A_i) = 0.998$. We find the probability of at least one accident in a year as unity minus the the probability of no accidents at all, i.e.

$$\begin{aligned} P(\textit{At least one accident}) &= 1 - P(\textit{No accidents}) \\ &= 1 - P(\bigcap_{i=1}^{365} A_i) \\ &= 1 - \prod_{i=1}^{365} P(A_i) \quad (\textit{by independence}) \\ &= 1 - (0.998)^{365} \\ &= 0.5184. \end{aligned}$$

18. **QUESTION**:
Two events A and B are such that $P(A) = 0.5$, $P(B) = 0.3$ and $P(A \cap B) = 0.1$. Calculate

(a) $P(A|B)$;
(b) $P(B|A)$;
(c) $P(A|A \cup B)$;
(d) $P(A|A \cap B)$;
(e) $P(A \cap B|A \cup B)$.

SOLUTION:
(Venn diagrams are helpful in understanding some of the events that arise below.)

(a) $P(A|B) = P(A \cap B)/P(B) = \frac{1}{3}$
(b) $P(B|A) = P(A \cap B)/P(A) = \frac{1}{5}$
(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7$, and the event $A \cap (A \cup B) = A$, so

$$P(A|A \cup B) = P(A)/P(A \cup B) = \frac{5}{7}.$$

(d) $P(A|A \cap B) = P(A \cap B)/P(A \cap B) = 1$, since $A \cap (A \cap B) = A \cap B$.
(e) $P(A \cap B|A \cup B) = P(A \cap B)/P(A \cup B) = \frac{1}{7}$, since $A \cap B \cap (A \cup B) = A \cap B$.

19. **QUESTION**:
An urn contains r red balls and b blue balls, $r \geq 1$, $b \geq 3$. Three balls are selected, without replacement, from the urn. Using the notion of conditional probability to simplify the problem, find the probability of the sequence Blue, Red, Blue.

SOLUTION:
Let B_i be the event that a blue ball is drawn on the ith draw, and define R_i similarly. We require

$$\begin{aligned} P(B_1 R_2 B_3) &= P(B_3|R_2 B_1)P(R_2|B_1)P(B_1) \\ &= (\frac{b-1}{r+b-2})(\frac{r}{r+b-1})(\frac{b}{r+b}). \end{aligned}$$

20. **QUESTION**:
Three babies are given a weekly health check at a clinic, and then returned randomly to their mothers. What is the probability that at least one baby goes to the right mother?

SOLUTION:
Let E_i be the event that baby i is reunited with its mother. We need $P(E_1 \cup E_2 \cup E_3)$, where we can use the result

$$Pr(A \cup B \cup C) = Pr(A) + Pr(B) + Pr(C) - Pr(A \cap B) - Pr(A \cap C) - Pr(B \cap C) + Pr(A \cap B \cap C).$$

for any A, B, C. The individual probabilities are $P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$. The pairwise joint probabilities are equal to $\frac{1}{6}$, since $P(E_1 E_2) = P(E_2|E_1)P(E_1) = (\frac{1}{2})(\frac{1}{3})$, and the triplet $P(E_1 E_2 E_3) = \frac{1}{6}$ similarly. Hence our final answer is

$$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \frac{1}{6} - \frac{1}{6} - \frac{1}{6} + \frac{1}{6} = \frac{2}{3}.$$

21. **QUESTION**:
In a certain town, 30% of the people are Conservatives; 50% Socialists; and 20% Liberals. In this town at the last election, 65% of Conservatives voted, as did 82% of the Socialists and 50% of the Liberals. A person from the town is selected at random, and states that she voted at the last election. What is the probability that she is a Socialist?

SOLUTION:
We organise the problem as follows: let C, S and L be the events that a person is Conservative, Socialist, or Liberal respectively. Let V be the event that a person voted in the last election. We require to find $\mathrm{P}(S|V)$, where the information we are given can be summarised as:

$$\mathrm{P}(C)=0.3,\ \ \mathrm{P}(S)=0.5,\ \ \mathrm{P}(L)=0.2,$$

$$\mathrm{P}(V|C)=0.65\ \ \mathrm{P}(V|S)=0.82,\ \ \mathrm{P}(V|L)=0.5.$$

Now, by Bayes theorem,

$$\mathrm{P}(S|V)=\frac{\mathrm{P}(V|S)\mathrm{P}(S)}{\mathrm{P}(V)}.$$

Each term is known, excepting $\mathrm{P}(V)$ which we calculate using the idea of a partition. We can calculate $\mathrm{P}(V)$ by associating V with the certain partition $C\cup S\cup L$:

$$\begin{aligned}\mathrm{P}(V) &= \mathrm{P}(V\cap(C\cup S\cup L))\\ &= \mathrm{P}(VC)+\mathrm{P}(VS)+\mathrm{P}(VL)\\ &= \mathrm{P}(V|C)\mathrm{P}(C)+\mathrm{P}(V|S)\mathrm{P}(S)+\mathrm{P}(V|L)\mathrm{P}(L)\\ &= (0.65)(0.3)+(0.82)(0.5)+(0.5)(0.2)\\ &= 0.705.\end{aligned}$$

Hence

$$\begin{aligned}\mathrm{P}(S|V) &= \frac{(0.82)(0.5)}{0.705}\\ &= 0.5816.\end{aligned}$$

22. **QUESTION**:
Three prisoners, A, B, and C, are held in separate cells. Two are to be executed. The warder knows specifically who is to be executed, and who is to be freed, whereas the prisoners know only that two are to be executed. Prisoner A reasons as follows: my probability of being freed is clearly $\frac{1}{3}$ until I receive further information. However, it is clear that at least one of B and C will be executed, so I will ask the warder to name one prisoner *other than myself* who is to be executed. Once I know which of B and C is to be executed, either I will go free or the other, unnamed, prisoner will go free, with equal probability. Hence, by asking the name of another prisoner to be executed, I raise my chances of survival from $\frac{1}{3}$ to $\frac{1}{2}$. Investigate A's reasoning. [Hint: find the conditional probability that A is freed, given that the warder names B to be executed.]

SOLUTION:
A's reasoning is unsound. It does not take into account the latitude that the warder has in naming another prisoner to be executed. To see this, let A_F be the event that A goes free, and let W_B be the event that the warder names B. We need to calculate

$$\mathrm{P}(A_F|W_B)=\frac{\mathrm{P}(W_B|A_F)\mathrm{P}(A_F)}{\mathrm{P}(W_B)}.$$

We have

$$\begin{aligned}\mathrm{P}(W_B) &= \mathrm{P}(W_B|A_F)\mathrm{P}(A_F)+\mathrm{P}(W_B|B_F)\mathrm{P}(B_F)+\mathrm{P}(W_B|C_F)\mathrm{P}(C_F)\\ &= (\frac{1}{2})(\frac{1}{3})+(0)(\frac{1}{3})+(1)(\frac{1}{3})\\ &= \frac{1}{2}\end{aligned}$$

and we find thereby that $\mathrm{P}(A_F|W_B)=\frac{1}{3}$. (This analysis supposes that the warder is equally likely to name either B or C in the situation that both are to be executed.)

23. **QUESTION**:
You're playing duplicate bridge. Your partner has bid two spades, and you have to decide whether to pass or to bid game in spades, namely to bid four spades. You reckon that there is a good chance, 40%, that four spades will make. Otherwise, you think three spades will make about 40% of time, and two spades the rest of the time. Suppose there are no doubles (by the opposition, for penalties). The gains and losses depend on whether you are *vulnerable* or not. The possible outcomes and scores are as follows:

	Not vulnerable Score if you make			Vulnerable Score if you make		
You bid	2 spades	3 spades	4 spades	2 spades	3 spades	4 spades
2 spades	110	140	170	110	140	170
3 spades	-50	140	170	-100	140	170
4 spades	-100	-50	420	-200	-100	620

What should you bid when not vulnerable? What should you bid when vulnerable? Calculate the variation in score for one of the bids.

SOLUTION:
This involves calculating expected values and variances for each bid separately, given the different possible outcomes. Suppose you bid two spades. Let X be the score you obtain. X is a random variable with probability distribution as shown below. The expected value is $\mathrm{E}(X) = \sum_x x\mathrm{P}(X = x)$. The calculations are shown below. For the variance, we also need to calculate $\mathrm{E}(X^2)$.

x	110	140	170	Sum
$\mathrm{P}(X = x)$	0.2	0.4	0.4	1
$x\mathrm{P}(X = x)$	22	56	68	146
$x^2\mathrm{P}(X = x)$	2420	7840	11560	21820

It follows that $\mathrm{E}(X) = 146$ and $\mathrm{E}(X^2) = 21820$ so that

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = 21820 - (146)^2 = 504,$$

so that $SD(X) = \sqrt{504} = 22.45$.
The complete set of expected values and standard deviations, for each case, is as follows.

	Not vulnerable		Vulnerable	
You bid	$\mathrm{E}(X)$	SD(X)	$\mathrm{E}(X)$	SD(X)
2 spades	146	22	146	22
3 spades	114	83	104	103
4 spades	128	240	168	371

If you are not vulnerable, you maximise your expected score by bidding two spades. If you are vulnerable, you maximise your expected score by bidding four spades. There is substantial variation amongst the scores, particularly for the higher bids.

24. **QUESTION**:
Tay-Sachs disease is a rare fatal genetic disease occurring chiefly in children, especially of Jewish or Slavic extraction. Suppose that we limit ourselves to families which have (a) exactly three children, and (b) which have both parents carrying the Tay-Sachs disease. For such parents, each child has independent probability $\frac{1}{4}$ of getting the disease.
Write X to be the random variable representing the number of children that will have the disease.

(a) Show **(without using any knowledge you might have about the binomial distribution!)** that the probability distribution for X is as follows:

k	0	1	2	3
$\mathrm{P}(X=k)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

(b) Show that $\mathrm{E}(X) = \frac{3}{4}$ and that $\mathrm{Var}(X) = \frac{9}{16}$.

SOLUTION:
To answer this question, it is necessary both to think about *how* certain events can occur, and the *probability* that they occcur. Let H be the event that a child is healthy, and D be the event that a child has the disease. If there are three children, there are 8 possiblities (including different orderings) as shown in the next table, where by the sequence H, H, D we mean that the first two children were born healthy, and the third was born with the disease.

Sequence	Probability	X
H,H,H	$\frac{3}{4}\frac{3}{4}\frac{3}{4} = \frac{27}{64}$	0
H,H,D	$\frac{3}{4}\frac{3}{4}\frac{1}{4} = \frac{9}{64}$	1
H,D,H	$\frac{3}{4}\frac{1}{4}\frac{3}{4} = \frac{9}{64}$	1
D,H,H	$\frac{1}{4}\frac{3}{4}\frac{3}{4} = \frac{9}{64}$	1
H,D,D	$\frac{3}{4}\frac{1}{4}\frac{1}{4} = \frac{3}{64}$	2
D,H,D	$\frac{1}{4}\frac{3}{4}\frac{1}{4} = \frac{3}{64}$	2
D,D,H	$\frac{1}{4}\frac{1}{4}\frac{3}{4} = \frac{3}{64}$	2
D,D,D	$\frac{1}{4}\frac{1}{4}\frac{1}{4} = \frac{1}{64}$	3

The probabilities for each sequence are shown in the second column; successive births are independent so that we can multiply probabilities. Notice that the sum of the probabilities is 1. The random variable X is the number of children having the disease. We see that only one sequence leads to $X = 0$, and this sequence has probability $\frac{27}{64}$. Hence $\mathrm{P}(X = 0) = \frac{27}{64}$. There are three sequences leading to $X = 1$, each with probability $\frac{9}{64}$. Hence $\mathrm{P}(X = 1) = \frac{9}{64} + \frac{9}{64} + \frac{9}{64} = \frac{27}{64}$. The other probabilities are found similarly. It is easy to show that $\mathrm{E}(X) = \frac{3}{4}$ and that $\mathrm{E}(X^2) = \frac{18}{16}$, so that $\mathrm{Var}(X) = \frac{9}{16}$.
Remark. This is an example of a *binomial* distribution with parameters $n = 3$ and $p = \frac{1}{4}$. That is,

$$\mathrm{P}(X = k) = \frac{3!}{k!(3-k)!}(\frac{1}{4})^k(\frac{3}{4})^{3-k}, \quad k = 0, 1, 2, 3.$$

For such distributions it is well known that $\mathrm{E}(X) = np$ and that $\mathrm{Var}(X) = np(1-p)$.

25. **QUESTION**:
A six-sided die has four green and two red faces and is balanced so that each face is equally likely to come up. The die will be rolled several times. Suppose that we score 4 if the die is rolled and comes up green, and 1 if it comes up red. Define the random variable X to be this score. Write down the distribution of probability for X and calculate the expectation and variance for X.

SOLUTION:
The distribution of X is as follows.

k	1	4
$\mathrm{P}(X = k)$	$\frac{1}{3}$	$\frac{2}{3}$

$$
\begin{aligned}
E(X) &= (1)\frac{1}{3} + (4)\frac{2}{3} = 3. \\
E(X^2) &= (1^2)\frac{1}{3} + (4^2)\frac{2}{3} = 11. \\
Var(X) &= E(X^2) - (E(X))^2 = 11 - 9 = 2.
\end{aligned}
$$

26. **QUESTION**:
For two standard dice all 36 outcomes of a throw are equally likely. Find $P(X_1 + X_2 = j)$ for all j and calculate $E(X_1 + X_2)$. Confirm that $E(X_1) + E(X_2) = E(X_1 + X_2)$.

SOLUTION:
The possible totals are $j = 2, 3, \ldots, 12$ and $P(X_1 + X_2 = j) = (j-1)/36$, $j = 2, \ldots, 7$ and $P(X_1 + X_2 = j) = (13 - j)/36$, $j = 8, \ldots, 12$. For each of the dice $E(X_i) = 21/6 = 7/2$ while for the total

$$E(X_1 + X_2) = \tfrac{1}{36}(2 \cdot 1 + 3 \cdot 2 + \cdots + 7 \cdot 6 + \cdots + 12 \cdot 1) = 252/36 = 7$$

27. **QUESTION**:
X takes values 1, 2, 3, 4 each with probability 1/4 and Y takes values 1, 2, 4, 8 with probabilities 1/2, 1/4, 1/8 and 1/8 respectively. Write out a table of probabilities for the 16 paired outcomes which is consistent with the distributions of X and Y. From this find the possible values and matching probabilities for the total $X + Y$ and confirm that $E(X + Y) = E(X) + E(Y)$.

SOLUTION:
There are 16 pairs and infinitely many ways to allocate the probabilities. Selecting one, say $p(1,4) = p(1,8) = 1/8$, $p(2,2) = 1/4$, $p(3,1) = p(4,1) = 1/4$ we see this satisfies $\sum_j p(i,j) = P(X = i)$ and $\sum_i (i,j) = P(Y = j)$. The possible values and probabilities are

t	2	3	4	5	6	7	8	9	10	11	12
p_t	0	0	1/2	3/8	0	0	0	1/8	0	0	0

where for instance $P(X + Y = 4) = p(2,2) + p(3,1) = 1/2$. From the table, $E(X + Y) = (4 \cdot 4 + 5 \cdot 3 + 9 \cdot 1)/8 = 5$. As $E(X) = 5/2$ and $E(Y) = 5/2$ the required equality holds.

28. **QUESTION**:
Calculation practice for the binomial distribution. Find $P(X = 2)$, $P(X < 2)$, $P(X > 2)$ when

(a) $n = 4$, $p = 0.2$; (b) $n = 8$, $p = 0.1$;
(c) $n = 16$, $p = 0.05$; (d) $n = 64$, $p = 0.0125$.

SOLUTION:
(a) $P(X = 2) = 6 \cdot 0.2^2 \cdot 0.8^2 = 0.1536$, $P(X < 2) = P(X = 0) + P(X = 1) = 0.4096 + 0.4096 = 0.8192$, $P(X > 2) = 1 - P(X \leq 2) = 1 - 0.8192 - 0.1536 = 0.0272$.
(b) $P(X = 2) = 0.1488$, $P(X < 2) = 0.8131$, $P(X > 2) = 0.0381$.
(c) $P(X = 2) = 0.1463$, $P(X < 2) = 0.8108$, $P(X > 2) = 0.0429$.
(d) $P(X = 2) = 0.1444$, $P(X < 2) = 0.8093$, $P(X > 2) = 0.0463$.

29. **QUESTION**:
A wholesaler supplies products to 10 retail stores, each of which will independently make an order on a given day with chance 0.35. What is the probability of getting exactly 2 orders? Find the most probable number of orders per day and the probability of this number of orders. Find the expected number of orders per day.

SOLUTION:
Using the independence of orders the chance that only the first two stores place orders is $0.35^2 \cdot 0.65^8$. As there are $10 \times 9/2 = 45$ distinct pairs of stores that could order we have

$$P(X = 2) = 450.35^2 0.65^8 = 0.1757$$

A similar argument works for any number of orders. We say that the number of orders placed has the $\text{Bin}(10, 0.35)$ distribution. The formula for x orders is

$$P(X = x) = \binom{10}{x} 0.35^x 0.65^{10-x}$$

The most probable number of orders is 3 (either calculate $P(X = x)$ for a few different x values or look at binomial tables in a textbook) and $P(X = 3) = 120(0.35)^3(0.65)^7 \approx 0.2522$. The expected number of orders is

$$\sum_0^{10} x \cdot P(X = x) = 1 \cdot 0.0725 + 2 \cdot 0.1757 + 3 \cdot 0.2522 + 4 \cdot 0.2377 + \cdots$$

which (barring numericals errors) will give the same answer as the formula $E(X) = np = 10 \times 0.35 = 3.5$.

(The problem of which number is most likely for general n and p was not set but is not all that hard – show that $P(X = x + 1) < P(X = x) \Leftrightarrow x + 1 > (n + 1)p$ and think about what that means)

30. **QUESTION**:
A machine produces items of which 1% at random are defective. How many items can be packed in a box while keeping the chance of one or more defectives in the box to be no more than 0.5? What are the expected value and standard deviation of the number of defectives in a box of that size?

SOLUTION:
Let X be the number of defectives when n items are packed into a box. $P(X = 0) = 0.99^n$ so that $P(X \geq 1) = 1 - 0.99^n$. To ensure $1 - 0.99^n < 0.5$ we must take $n < \log 0.5/\log 0.99 = 68.97$ so $n = 68$. The expected value and standard deviation of X when $n = 68$ are 0.68 and $\sqrt{0.68 \times 0.99} = 0.8205$.

31. **QUESTION**:
Suppose that 0.3% of bolts made by a machine are defective, the defectives occurring at random during production. If the bolts are packaged in boxes of 100, what is the Poisson approximation that a given box will contain x defectives? Suppose you buy 8 boxes of bolts. What is the distribution of the number of boxes with no defective bolts? What is the expected number of boxes with no defective bolts?

SOLUTION:
D is $\text{Bin}(100, 0.003)$ which is approximately Poisson with parameter $100 \times 0.003 = 0.3$. Hence $P(D = x) \approx e^{-0.3}(0.3)^x/x!\,,\ x = 0, 1, \ldots$. In particular, $P(D = 0) \approx 0.7408$. Finally, N, the no. of boxes with no defectives is $\text{Bin}(8, 0.7408)$ and so $E(N) = 8 \times 0.7408 = 5.926$.

32. **QUESTION**:
Events which occur randomly at rate r are counted over a time period of length s so the event count X is Poisson. Find $P(X = 2)$, $P(X < 2)$ and $P(X > 2)$ when

(a) $r = 0.8$, $s = 1$; (b) $r = 0.1$, $s = 8$; (c) $r = 0.01$, $s = 200$; (d) $r = 0.05$, $s = 200$.

SOLUTION:
(a) and (b) $\lambda = rs = 0.8$ so that $P(X = 2) = e^{-0.8}0.8^2/2 = 0.1438$, $P(X < 2) = 0.4493 + 0.3695 = 0.8088$ and $P(X > 2) = 1 - 0.8088 - 0.1438 = 0.0474$. (c) $\lambda = 2$ so that $P(X = 2) = 0.2707$, $P(X < 2) = 0.4060$, $P(X > 2) = 0.3233$. (d) $\lambda = 10$ so that $P(X = 2) = 0.00227$, $P(X < 2) = 0.00050$, $P(X > 2) = 0.9972$.

33. **QUESTION**:
Given that 0.04% of vehicles break down when driving through a certain tunnel find the probability of (a) no (b) at least two breakdowns in an hour when 2,000 vehicles enter the tunnel.

SOLUTION:
The number of breakdowns X has a binomial distribution which can be approximated by the $\text{Pn}(\lambda)$ distribution with $\lambda = 2000 \times 0.0004 = 0.8$. Hence $P(X = 0) \approx 0.4493$ and $P(X \geq 2) = 1 - P(X \leq 1) \approx 1 - 0.8088 = 0.1912$.

34. **QUESTION**:
Experiments by Rutherford and Geiger in 1910 showed that the number of alpha particles emitted per unit time in a radioactive process is a random variable having a Poisson distribution. Let X denote the count over one second and suppose it has mean 5. What is the probability of observing fewer than two particles during any given second? What is the $P(X \geq 10)$? Let Y denote the count over a separate period of 1.5 seconds. What is $P(Y \geq 10)$? What is $P(X + Y \geq 10)$?

SOLUTION:
$P(X \leq 1) = e^{-5}(1 + 5) = 0.0404$. $P(X \geq 10) = 0.0398$. Y is $\text{Pn}(7.5)$ and so $P(Y \geq 10) = 1 - P(Y \leq 9) = 1 - \sum_0^9 P(Y = i) = 0.5113$. $X + Y$ is $\text{Pn}(12.5)$ and we find $P(X + Y \geq 10) = 0.7986$.

35. **QUESTION**:
A process for putting chocolate chips into cookies is random and the number of choc chips in a cookie has a Poisson distribution with mean λ. Find an expression for the probability that a cookie contains less than 3 choc chips.

SOLUTION:
The Poisson distribution gives probabilities for each possible number of choc chips but as a cookie can't contain two different numbers simultaneously we *add* the probabilities for the possible values 0, 1 and 2. Hence

$$P(X < 3) = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2}\right)$$

(if you wonder what happened to s, it equals 1 as we're only looking at a single cookie – for 10 cookies we take $s = 10$ etc).

36. **QUESTION**:
Let X have the density $f(x) = 2x$ if $0 \leq x \leq 1$ and $f(x) = 0$ otherwise. Show that X has the mean $2/3$ and the variance $1/18$. Find the mean and the variance of the random variable $Y = -2X + 3$.

SOLUTION:
To find expected values for continuous random variables we integrate e.g.

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 x \cdot 2x\, dx = 2/3$$

and similarly

$$\text{Var}(X) = \int_0^1 (x - 2/3)^2 \cdot 2x\, dx = 1/2 - 8/9 + 4/9 = 1/18$$

You could also use the formula $\text{Var}(X) = E(X^2) - E(X)^2$ where

$$E(X^2) = \int_0^1 x^2 \cdot 2x\, dx = 1/2$$

so that $\text{Var}(X) = 1/2 - (2/3)^2 = 1/18$. Use the linearity of expectation for the last bits to get

$$E(Y) = -2 \times 2/3 + 3 = 5/3 \quad \text{and} \quad \text{Var}(Y) = (-2)^2 \times 1/18 = 2/9$$

37. **QUESTION**:
Let the random variable X have the density $f(x) = kx$ if $0 \le x \le 3$. Find k. Find x_1 and x_2 such that $P(X \le x_1) = 0.1$ and $P(X \le x_2) = 0.95$. Find $P(|X - 1.8| < 0.6)$.

SOLUTION:
To make $\int_0^3 f(x)dx = 1$ we must take $k = 2/9$. For any c in $(0,3)$, $P(X \le c) = c^2/9$ so $x_1 = \sqrt{0.9} = 0.9487$ and $x_2 = \sqrt{8.55} = 2.9240$. $P(|X - 1.8| < 0.6) = P(1.2 < X < 2.4) = (2.4^2 - 1.2^2)/9 = 0.48$.

38. **QUESTION**:
A small petrol station is supplied with petrol once a week. Assume that its volume X of potential sales (in units of 10,000 litres) has the probability density function $f(x) = 6(x-2)(3-x)$ for $2 \le x \le 3$ and $f(x) = 0$ otherwise. Determine the mean and the variance of this distribution. What capacity must the tank have for the probability that the tank will be emptied in a given week to be 5%?

SOLUTION:
Proceed as usual, $E(X) = 5/2$ and $\mathrm{Var}(X) = 1/20$. Let T denote the capacity of the tank. We need to solve $0.05 = \int_T^3 6(x-2)(3-x)dx = \int_{T-2}^1 6y(1-y)\,dy = 1 - 3(T-2)^2 + 2(T-2)^3$ and doing this numerically (iterate the equation $T - 2 = \sqrt{0.95}(7-2T)^{-1/2}$) we find $T \approx 2.86465$ i.e. the tank should hold approximately 28,650 litres.

39. **QUESTION**:
Find the probability that none of the three bulbs in a set of traffic lights will have to be replaced during the first 1200 hours of operation if the lifetime X of a bulb (in thousands of hours) is a random variable with probability density function $f(x) = 6[0.25 - (x-1.5)^2]$ when $1 \le x \le 2$ and $f(x) = 0$ otherwise. You should assume that the lifetimes of different bulbs are independent.

SOLUTION:
For a single bulb, $P(X > 1.2) = 6[\frac{3}{2}x^2 - \frac{1}{3}x^3 - 2x]_{1.2}^2 = 0.8960$. Hence $P(\text{no bulbs replaced}) = 0.8960^3 = 0.7193$.

40. **QUESTION**:
Suppose X is $\mathrm{N}(10,1)$. Find (i) $P[X > 10.5]$, (ii) $P[9.5 < X < 11]$, (iii) x such that $P[X < x] = 0.95$. *You will need to use Standard Normal tables.*

SOLUTION:
Let Z denote a $N(0,1)$ random variable from now on and let Φ denote its cdf.

(i) $P[X > 10.5] = P[X - 10 > 0.5] = 1 - \Phi(0.5) = 0.3085$; (ii) $P[9.5 < X < 11] = \Phi(1) - \Phi(-0.5) = 0.5328$; (iii) $P[X < x] = P[Z < x - 10] = 0.95$ when $x - 10 = 1.645$ i.e. $x = 11.645$.

41. **QUESTION**:
Suppose X is $\mathrm{N}(-1,4)$. Find

(a) $P(X < 0)$; (b) $P(X > 1)$; (c) $P(-2 < X < 3)$; (d) $P(|X+1| < 1)$.

SOLUTION:
As $X = 2Z + 1$ we have (a) $P(X < 0) = P(2Z + 1 < 0) = \Phi(-1/2) = 1 - \Phi(1/2) = 0.3085$; (b) $P(X > 1) = \Phi(0) = 1/2$; (c) $P(-2 < X < 3) = \Phi(1) - \Phi(-3/2) = 0.7745$; (d) $P(|X+1| < 1) = P(-2 < X < 0) = \Phi(-1/2) - \Phi(-3/2) = 0.2417$.

42. **QUESTION**:
Suppose X is $\mathrm{N}(\mu, \sigma^2)$. For $a = 1$, 2, 3 find $P(|X - \mu| < a\sigma)$.

SOLUTION:
$P(|X - \mu| < a\sigma) = P(-a < Z < a) = 2\Phi(a) - 1$ so the required values are approximately 0.682, 0.954 and 0.998 respectively.

43. **QUESTION**:
The height of a randomly selected man from a population is normal with $\mu = 178$cm and $\sigma = 8$cm. What proportion of men from this population are over 185cm tall? There are 2.54cm to an inch. What is their height distribution in inches? The heights of the women in this population are normal with $\mu = 165$ cm and $\sigma = 7$cm. What proportion of the women are taller than half of the men?

SOLUTION:
Let M denote the height of a man and W the height of a woman in centimetres. We want to know $P(M > 185) = P(\frac{M-178}{8} > \frac{185-178}{8}) = P(Z > 0.875) \approx 0.19$. Let H denote the height of a man in inches. Then $H = M/2.54$ so that H is $N(70.1\,, (3.15)^2\,)$. Finally $P(M < h) = 0.5$ when $h = 178$ and so $P(W > 178) = P(Z > \frac{13}{7}) = 0.032$ so 3.2% of the women are taller than half of the men.

44. **QUESTION**:
N independent trials are to be conducted, each with "success" probability p. Let $X_i = 1$ if trial i is a success and $X_i = 0$ if it is not. What is the distribution of the random variable $X = X_1+X_2+\ldots+X_N$? Express $P[a \le X \le b]$ as a sum (where $a \le b$ and these are integers between 0 and N). Use the central limit theorem to provide an approximation to this probability. Compare your approximation with the limit theorem of De Moivre and Laplace on p1189 of Kreyszig.

SOLUTION:
As X is the total number of successes in N independent trials X is $\mathrm{Bin}(N, p)$. Thus

$$P[a \le X \le b] = \sum_{x=a}^{b} \binom{N}{x} p^x (1-p)^{N-x}.$$

As each X_i has $E(X_i) = p$ and $\mathrm{Var}(X_i) = p(1-p)$ *(you should confirm this)* the central limit theorem says that X is approximately $\mathrm{Normal}\big(Np, Np(1-p)\big)$. Let $\bar{a} = (a - Np)/\sqrt{Np(1-p)}$ and $\bar{b} = (b - Np)/\sqrt{Np(1-p)}$. The approximation is $P[a \le X \le b] \approx P[\bar{a} < Z < \bar{b}]$ where Z has the standard Normal distribution. The reason for the small correction in the version of this result in Kreyszig is that $P[a \le X \le b] = P[a - \delta < X < b + \delta]$ for any $\delta \in (0, 1)$ while the approximation varies with δ – the choice $\delta = 0.5$ is arbitrary but generally sensible.

45. **QUESTION**:
Suppose that of $1,000,000$ live births in Paris over some period, $508,000$ are boys. Suppose X is $\mathrm{Bin}(10^6, 0.5)$ and calculate approximately $P[X \ge 508,000]$. Does it seem reasonable to you that the proportion of males among Parisian babies conceived soon after the above period will be 50%. (Laplace developed his limit theorem in the late 1700's to deal with a question similar to this.)

SOLUTION:
From q17, $\bar{b} = 8,000/500 = 16$ and from the formulæ sheet $1 - \Phi(16) \approx 6.4 \times 10^{-58}$. It seems entirely unreasonable that births at that time and place should be modelled with $p = 0.5$ chance of each sex. Standard practice would be to say that the proportion is within some small distance of $0.508(= 508,000/10^6\,)$.

46. **QUESTION**:
An airfreight company has various classes of freight. In one of these classes the average weight of packages is 10kg and the variance of the weight distribution is 9kg^2. Assuming that the package weights are independent (it is not the case that a single company is sending a large number of identical packages, for instance), estimate the probability that 100 packages will have total weight more than 1020kg.

SOLUTION:
The central limit theorem says $\sum W_i$ is approximately $N(1,000\,, 30^2\,)$ so that $P(\sum W_i > 1,020) \approx P\big[Z > (1,020 - 1,000)/30 = 0.67\big] = 0.251$.

Independent Events
Video 249 on Corbettmaths

Examples

Workout

Click here

Scan here

Question 1: A fair coin is flipped twice.

(a) Find the probability that the coin lands on heads twice.
(b) Find the probability that the coin lands on tails twice.
(c) Find the probability that the coin lands on heads exactly once.

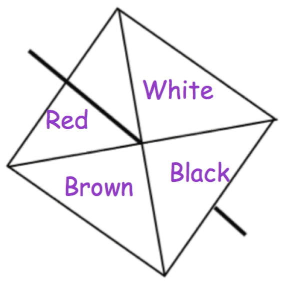
Question 2: Penelope is playing football.
When attempting to score a penalty, the probability she scores is $\frac{2}{3}$
During the game, Penelope takes two penalties.
Find the probability that Penelope scores both.

Question 3: Trevor is taking part in a quiz.
The probability that he answer a question correctly is $\frac{3}{5}$
Trevor is asked two questions.

(a) Calculate the probability that Trevor answers both questions correctly.
(b) Calculate the probability that Trevor answers both questions incorrectly.

Question 4: Daisy has a biased spinner.
The probability of each colour is:

Colour	Red	White	Black	Brown
Probability	0.1	0.4	0.3	0.2



Daisy spins the spinner twice.

(a) Find the probability of the spinner landing on white twice.
(b) Find the probability of the spinner landing on black and then brown.
(c) Find the probability of the spinner landing on the same colour in both spins.

Question 5: A fair six sided dice is rolled three times.

(a) Find the probability of getting a two all three times.
(b) Find the probability of getting no twos

Independent Events

Video 249 on Corbettmaths

Question 6: Mark is playing darts.
The probability he hits the bullseye is 0.4
Mark throws two darts

(a) Find the probability of Mark hitting the bullseye once.
(b) Find the probability of Mark hitting the bullseye at least once.

Question 7: A bag contains five yellow sweets, three green sweets and one purple sweet.
A sweet is taken out of the bag and replaced.
Another sweet is taken out.

(a) Find the probability that both sweets are yellow.
(b) Find the probability of neither sweet is green.
(c) Find the probability that the two sweets are different colours.

Question 8: The probability of a bus being on time is $\frac{3}{4}$

Archie catches the bus to work three times each week.

(a) Work out the probability that the bus is late every time.
(b) Work out the probability that the bus is on time every time.
(c) Work out the probability that the bus is late exactly once.

Question 9: Jackson, Frederick and Kelvin each sit a test.

The probability Jackson passes is $\frac{9}{10}$

The probability Frederick passes is $\frac{2}{3}$

The probability Kelvin passes is $\frac{1}{2}$

(a) Find the probability that Jackson and Kelvin pass, but Kelvin fails.

(b) Find the probability that Frederick passes, but Jackson and Kelvin fail.

(c) Find the probability that at least two boys pass.

Question 10: The probability that Dylan reads at night is $\frac{4}{5}$

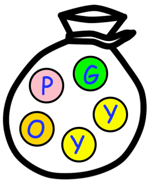
Calculate the probability that Dylan reads every night in one week.

Independent Events

Video 249 on Corbettmaths

Apply

Question 1: Amelia is organising a game for a charity fête.
She has put 1 orange, 1 pink, 1 green and 2 yellow counters into a bag.



To play, each person will pay £1 and take out a counter at random.
They will then replace the counter and then take a second counter at random.
The person will win £2.50 if both counters are the same colour.

Amelia expects 200 people to play the game.

How much money would Amelia expect to raise for charity?

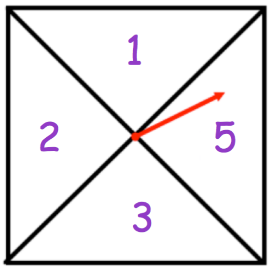
Question 2: There are12 tiles in a bag, each with a letter written on it.

C	O	R	B	E	T	T	M	A	T	H	S

A tile is selected at random and then replaced.
Another tile is then selected.

Find the probability that both tiles have different letters on them.

Question 3: A fair spinner has four sections.

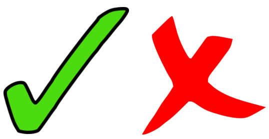


The spinner is spun three times.
The three numbers are added together to give a score.

(a) Find the probability that the score is odd.
(b) Find the probability that the score is greater than 3.

Question 4: Tom and Ben sit their driving test.
The probability Tom passes is 0.4
The probability that only one man passes is 0.56
Find the probability they both fail.

Answers



Click here

Scan here

Mean and Variance of Discrete Random Variables

Expected Value of Discrete Random Variable

Suppose you and I play a betting game: we flip a coin and if it lands heads, I give you a dollar, and if it lands tails, you give me a dollar.

On average, how much am I expected to win or lose?

$$\text{expected winnings} = \underbrace{(-1)\left(\frac{1}{2}\right)}_{\text{win -\$1 half the time}} + \underbrace{(1)\left(\frac{1}{2}\right)}_{\text{win \$1 half the time}} = 0$$

Expected Value of Discrete Random Variable

Suppose you and I play a betting game: we flip a coin and if it lands heads, I give you a dollar, and if it lands tails, you give me a dollar.

On average, how much am I expected to win or lose?

$$\text{expected winnings} = \underbrace{(-1)\left(\frac{1}{2}\right)}_{\text{win -\$1 half the time}} + \underbrace{(1)\left(\frac{1}{2}\right)}_{\text{win \$1 half the time}} = 0$$

Expected Value of Discrete Random Variable

Suppose you and I play a betting game: we flip a coin and if it lands heads, I give you a dollar, and if it lands tails, you give me a dollar.
On average, how much am I expected to win or lose?

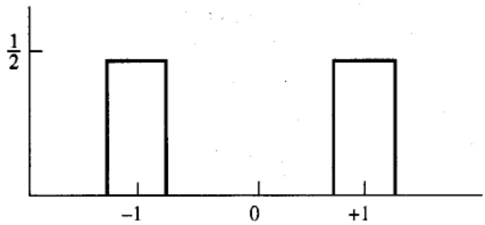
$$\text{expected winnings} = \underbrace{(-1)\left(\frac{1}{2}\right)}_{\text{win -\$1 half the time}} + \underbrace{(1)\left(\frac{1}{2}\right)}_{\text{win \$1 half the time}} = 0$$

Expected Value of Discrete Random Variable

Suppose you and I play a betting game: we flip a coin and if it lands heads, I give you a dollar, and if it lands tails, you give me a dollar.

On average, how much am I expected to win or lose?

Figure 4.4. Relative frequency of winnings

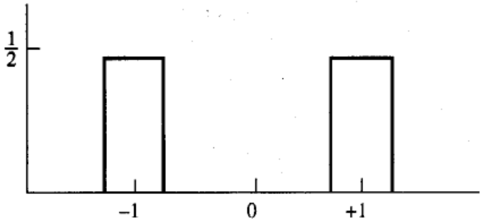


$$\text{expected winnings} = \underbrace{(-1)\left(\frac{1}{2}\right)}_{\text{win -\$1 half the time}} + \underbrace{(1)\left(\frac{1}{2}\right)}_{\text{win \$1 half the time}} = 0$$

Expected Value of Discrete Random Variable

Suppose you and I play a betting game: we flip a coin and if it lands heads, I give you a dollar, and if it lands tails, you give me a dollar.
On average, how much am I expected to win or lose?

Figure 4.4. Relative frequency of winnings



$$\text{expected winnings} = \underbrace{(-1)\left(\frac{1}{2}\right)}_{\text{win -\$1 half the time}} + \underbrace{(1)\left(\frac{1}{2}\right)}_{\text{win \$1 half the time}} = 0$$

Definition of Expected Value of a Discrete Random Variable

Definition

The **expected value of a discrete random variable** X with probability distribution $p(x)$ is given by

$$\mathrm{E}(X) \triangleq \mu = \sum_x x \, p_X(x) \qquad (\star)$$

where the sum is over all values of x for which $p_X(x) > 0$.

Note that in order for $(\star)$ to exist, the sum must converge absolutely; that is

$$\sum_x |x| \, p_X(x) < \infty \qquad (\star\star)$$

If $(\star\star)$ does not hold, we say the expected value of X does not exist.

Definition of Expected Value of a Discrete Random Variable

Definition

The **expected value of a discrete random variable** X with probability distribution $p(x)$ is given by

$$\mathrm{E}(X) \triangleq \mu = \sum_x x\, p_X(x) \qquad (\star)$$

where the sum is over all values of x for which $p_X(x) > 0$.

Note that in order for $(\star)$ to exist, the sum must converge absolutely; that is

$$\sum_x |x|\, p_X(x) < \infty \qquad (\star\star)$$

If $(\star\star)$ does not hold, we say the expected value of X does not exist.

Definition of Expected Value of a Discrete Random Variable

Definition

The **expected value of a discrete random variable** X with probability distribution $p(x)$ is given by

$$\mathrm{E}(X) \triangleq \mu = \sum_x x\, p_X(x) \qquad (\star)$$

where the sum is over all values of x for which $p_X(x) > 0$.

Note that in order for $(\star)$ to exist, the sum must converge absolutely; that is

$$\sum_x |x|\, p_X(x) < \infty \qquad (\star\star)$$

If $(\star\star)$ does not hold, we say the expected value of X does not exist.

Expected value of functions of random variables

The expected value of $g(X)$ where g is any real-valued function is naturally

$$\mathrm{E}(g(X)) = \sum_x g(x)p(x)$$

Example

Consider the Bernoulli random variable

$$X = \begin{cases} 0, & \text{w.p. } 1/2 \\ 1, & \text{w.p. } 1/2 \end{cases}$$

Compute $\mathrm{E}\left(X^2 - 1\right)$.

$$\mathrm{E}\left(X^2 - 1\right) = (0^2 - 1)\left(\frac{1}{2}\right) + ((1)^2 - 1)\left(\frac{1}{2}\right) = \frac{-1}{2}$$

Expected value of functions of random variables

The expected value of $g(X)$ where g is any real-valued function is naturally

$$\mathrm{E}(g(X)) = \sum_x g(x)p(x)$$

Example

Consider the Bernoulli random variable

$$X = \begin{cases} 0, & \text{w.p. } 1/2 \\ 1, & \text{w.p. } 1/2 \end{cases}$$

Compute $\mathrm{E}\left(X^2 - 1\right)$.

$$\mathrm{E}\left(X^2 - 1\right) = (0^2 - 1)\left(\frac{1}{2}\right) + ((1)^2 - 1)\left(\frac{1}{2}\right) = \frac{-1}{2}$$

Birthday Problem Revisited

65 people participated in the birthday game a few weeks back.

I claimed that if no two birthdays matched, then I would pay everyone 30 monopoly dollars, but otherwise each person would pay me one monopoly dollar.

Therefore either I would lose 65*30=1950 monopoly dollars, or I would win 65 monopoly dollars.

My expected earnings should be somewhere between -1950 and +65. Let's compute it.

My total earnings are represented by the following random variable

$$X = \begin{cases} 30, & \text{w.p. } p \\ -1950, & \text{w.p. } 1-p \end{cases}$$

where

$$p \approx \frac{365(364)(363)\cdots(301)}{365^{65}} \approx .9977$$

Therefore

$$\mathrm{E}(X) = 30(.9977) - 1950(.0023) = 25.446$$

Birthday Problem Revisited

65 people participated in the birthday game a few weeks back.

I claimed that if no two birthdays matched, then I would pay everyone 30 monopoly dollars, but otherwise each person would pay me one monopoly dollar.

Therefore either I would lose 65*30=1950 monopoly dollars, or I would win 65 monopoly dollars.

My expected earnings should be somewhere between -1950 and +65. Let's compute it.

My total earnings are represented by the following random variable

$$X = \begin{cases} 30, & \text{w.p. } p \\ -1950, & \text{w.p. } 1-p \end{cases}$$

where

$$p \approx \frac{365(364)(363)\cdots(301)}{365^{65}} \approx .9977$$

Therefore

$$\mathrm{E}(X) = 30(.9977) - 1950(.0023) = 25.446$$

Birthday Problem Revisited

65 people participated in the birthday game a few weeks back.

I claimed that if no two birthdays matched, then I would pay everyone 30 monopoly dollars, but otherwise each person would pay me one monopoly dollar.

Therefore either I would lose 65*30=1950 monopoly dollars, or I would win 65 monopoly dollars.

My expected earnings should be somewhere between -1950 and +65. Let's compute it.

My total earnings are represented by the following random variable

$$X = \begin{cases} 30, & \text{w.p. } p \\ -1950, & \text{w.p. } 1-p \end{cases}$$

where

$$p \approx \frac{365(364)(363)\cdots(301)}{365^{65}} \approx .9977$$

Therefore

$$\mathrm{E}(X) = 30(.9977) - 1950(.0023) = 25.446$$

Birthday Problem Revisited

65 people participated in the birthday game a few weeks back.

I claimed that if no two birthdays matched, then I would pay everyone 30 monopoly dollars, but otherwise each person would pay me one monopoly dollar.

Therefore either I would lose 65*30=1950 monopoly dollars, or I would win 65 monopoly dollars.

My expected earnings should be somewhere between -1950 and +65. Let's compute it.

My total earnings are represented by the following random variable

$$X = \begin{cases} 30, & \text{w.p. } p \\ -1950, & \text{w.p. } 1-p \end{cases}$$

where

$$p \approx \frac{365(364)(363)\cdots(301)}{365^{65}} \approx .9977$$

Therefore

$$\mathrm{E}(X) = 30(.9977) - 1950(.0023) = 25.446$$

Birthday Problem Revisited

65 people participated in the birthday game a few weeks back.

I claimed that if no two birthdays matched, then I would pay everyone 30 monopoly dollars, but otherwise each person would pay me one monopoly dollar.

Therefore either I would lose 65*30=1950 monopoly dollars, or I would win 65 monopoly dollars.

My expected earnings should be somewhere between -1950 and +65. Let's compute it.

My total earnings are represented by the following random variable

$$X = \begin{cases} 30, & \text{w.p. } p \\ -1950, & \text{w.p. } 1-p \end{cases}$$

where

$$p \approx \frac{365(364)(363)\cdots(301)}{365^{65}} \approx .9977$$

Therefore

$$\mathrm{E}(X) = 30(.9977) - 1950(.0023) = 25.446$$

Birthday Problem Revisited

65 people participated in the birthday game a few weeks back.

I claimed that if no two birthdays matched, then I would pay everyone 30 monopoly dollars, but otherwise each person would pay me one monopoly dollar.

Therefore either I would lose 65*30=1950 monopoly dollars, or I would win 65 monopoly dollars.

My expected earnings should be somewhere between -1950 and +65. Let's compute it.

My total earnings are represented by the following random variable

$$X = \begin{cases} 30, & \text{w.p. } p \\ -1950, & \text{w.p. } 1-p \end{cases}$$

where

$$p \approx \frac{365(364)(363)\cdots(301)}{365^{65}} \approx .9977$$

Therefore

$$\mathrm{E}(X) = 30(.9977) - 1950(.0023) = 25.446$$

Variance

Definition (variance)

The **variance** of a random variable X with expected value μ is given by

$$\mathrm{var}(X) \triangleq \sigma^2 = \mathrm{E}\left[(X-\mu)^2\right]$$

Definition

The **standard deviation** of a random variable X is, σ, the square root of the variance, i.e.

$$\mathrm{sd}(X) \triangleq \sigma = \sqrt{\mathrm{E}\left[(X-\mu)^2\right]} = \sqrt{\mathrm{var}(X)}$$

Variance

Definition (variance)

The **variance** of a random variable X with expected value μ is given by

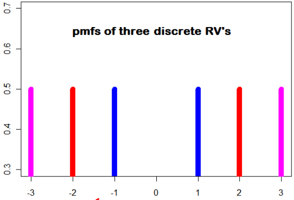
$$\mathrm{var}(X) \triangleq \sigma^2 = \mathrm{E}\left[(X-_{\mu})^2\right]$$

Definition

The **standard deviation** of a random variable X is, σ, the square root of the variance, i.e.

$$\mathrm{sd}(X) \triangleq \sigma = \sqrt{\mathrm{E}\left[(X-\mu)^2\right]} = \sqrt{\mathrm{var}(X)}$$

Some Distributions



$$X_1 = \begin{cases} -1, & \text{w.p. } 1/2 \\ 1, & \text{w.p. } 1/2 \end{cases} \quad X_2 = \begin{cases} -2, & \text{w.p. } 1/2 \\ 2, & \text{w.p. } 1/2 \end{cases} \quad X_3 = \begin{cases} -3, & \text{w.p. } 1/2 \\ 3, & \text{w.p. } 1/2 \end{cases}$$

Hence $\mathrm{E}(X_1) = \mathrm{E}(X_2) = \mathrm{E}(X_3) = 0$ and

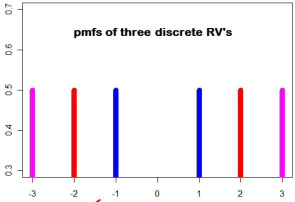
$$\mathrm{var}(X_1) = (1-0)^2\left(\frac{1}{2}\right) + (-1-0)^2\left(\frac{1}{2}\right) = 1$$

$$\mathrm{var}(X_2) = (2-0)^2\left(\frac{1}{2}\right) + (-2-0)^2\left(\frac{1}{2}\right) = 4$$

$$\mathrm{var}(X_3) = (3-0)^2\left(\frac{1}{2}\right) + (-3-0)^2\left(\frac{1}{2}\right) = 9$$

Therefore $\mathrm{sd}(X_1) = 1$, $\mathrm{sd}(X_2) = 2$, and $\mathrm{sd}(X_3) = 3$.

Some Distributions



$$X_1 = \begin{cases} -1, & \text{w.p. } 1/2 \\ 1, & \text{w.p. } 1/2 \end{cases} \quad X_2 = \begin{cases} -2, & \text{w.p. } 1/2 \\ 2, & \text{w.p. } 1/2 \end{cases} \quad X_3 = \begin{cases} -3, & \text{w.p. } 1/2 \\ 3, & \text{w.p. } 1/2 \end{cases}$$

Hence $\mathrm{E}(X_1) = \mathrm{E}(X_2) = \mathrm{E}(X_3) = 0$ and

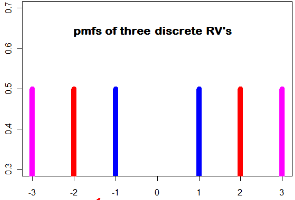
$$\mathrm{var}(X_1) = (1-0)^2\left(\frac{1}{2}\right) + (-1-0)^2\left(\frac{1}{2}\right) = 1$$

$$\mathrm{var}(X_2) = (2-0)^2\left(\frac{1}{2}\right) + (-2-0)^2\left(\frac{1}{2}\right) = 4$$

$$\mathrm{var}(X_3) = (3-0)^2\left(\frac{1}{2}\right) + (-3-0)^2\left(\frac{1}{2}\right) = 9$$

Therefore $\mathrm{sd}(X_1) = 1$, $\mathrm{sd}(X_2) = 2$, and $\mathrm{sd}(X_3) = 3$.

Some Distributions



$$X_1 = \begin{cases} -1, & \text{w.p. } 1/2 \\ 1, & \text{w.p. } 1/2 \end{cases} \quad X_2 = \begin{cases} -2, & \text{w.p. } 1/2 \\ 2, & \text{w.p. } 1/2 \end{cases} \quad X_3 = \begin{cases} -3, & \text{w.p. } 1/2 \\ 3, & \text{w.p. } 1/2 \end{cases}$$

Hence $\mathrm{E}(X_1) = \mathrm{E}(X_2) = \mathrm{E}(X_3) = 0$ and

$$\mathrm{var}(X_1) = (1-0)^2\left(\frac{1}{2}\right) + (-1-0)^2\left(\frac{1}{2}\right) = 1$$

$$\mathrm{var}(X_2) = (2-0)^2\left(\frac{1}{2}\right) + (-2-0)^2\left(\frac{1}{2}\right) = 4$$

$$\mathrm{var}(X_3) = (3-0)^2\left(\frac{1}{2}\right) + (-3-0)^2\left(\frac{1}{2}\right) = 9$$

Therefore $\mathrm{sd}(X_1) = 1$, $\mathrm{sd}(X_2) = 2$, and $\mathrm{sd}(X_3) = 3$.

Theorem 4.2

Theorem (mean and variance of $aX + b$)

For any random variable X (discrete or not) and constants a and b,

1. $\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$
2. $\mathrm{var}(aX + b) = a^2\,\mathrm{var}(X)$

It follows that if X has mean μ and standard deviation σ, then

$$Y = \frac{x - \mu}{\sigma}$$

has mean 0 and standard deviation 1. *This is the standardized form of X.*

Theorem

If X and Y are random variables, then $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$.

Theorem 4.2

Theorem (mean and variance of $aX + b$)

For any random variable X (discrete or not) and constants a and b,

1. $\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$
2. $\mathrm{var}(aX + b) = a^2\,\mathrm{var}(X)$

It follows that if X has mean μ and standard deviation σ, then

$$Y = \frac{x - \mu}{\sigma}$$

has mean 0 and standard deviation 1. *This is the standardized form of X.*

Theorem

If X and Y are random variables, then $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$.

Theorem 4.2

Theorem (mean and variance of $aX + b$)

For any random variable X (discrete or not) and constants a and b,

1. $\mathrm{E}(aX + b) = a\,\mathrm{E}(X) + b$
2. $\mathrm{var}(aX + b) = a^2\,\mathrm{var}(X)$

It follows that if X has mean μ and standard deviation σ, then

$$Y = \frac{x - \mu}{\sigma}$$

has mean 0 and standard deviation 1. *This is the standardized form of X.*

Theorem

If X and Y are random variables, then $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$.

Theorem 4.3

Theorem ($\mathrm{var}(X) = \mathrm{E}(X^2) - \mu^2$)

If X is a random variable with mean μ, then

$$\mathrm{var}(X) = \mathrm{E}(X^2) - \mu^2$$

Proof.

$$\begin{aligned}
\mathrm{var}(X) &= \mathrm{E}\left[(X-\mu)^2\right] \\
&= \mathrm{E}(X^2 - 2X\mu + \mu^2) \\
&= E(X^2) - E(2X\mu) + E(\mu^2) \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2
\end{aligned}$$

□

Theorem 4.3

Theorem ($\mathrm{var}(X) = \mathrm{E}(X^2) - \mu^2$)

If X is a random variable with mean μ, then

$$\mathrm{var}(X) = \mathrm{E}(X^2) - \mu^2$$

Proof.

$$\begin{aligned}
\mathrm{var}(X) &= \mathrm{E}\left[(X-\mu)^2\right] \\
&= \mathrm{E}(X^2 - 2X\mu + \mu^2) \\
&= E(X^2) - E(2X\mu) + E(\mu^2) \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2
\end{aligned}$$

$\square$

Chebyshev's inequality

There are many random variables X with a given mean μ and a given variance σ^2, but they all must satisfy the following inequality.

Theorem

Chebyshev's inequality Let X be a random variable with mean μ and variance σ^2. Then for any positive k,

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Letting $k = 2$ in Chebyshev's inequality gives

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{4} = \frac{3}{4}$$

That is, the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ must contain at least 3/4 of the probability mass.

Chebyshev's inequality

There are many random variables X with a given mean μ and a given variance σ^2, but they all must satisfy the following inequality.

Theorem

Chebyshev's inequality Let X be a random variable with mean μ and variance σ^2. Then for any positive k,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Letting $k = 2$ in Chebyshev's inequality gives

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{4} = \frac{3}{4}$$

That is, the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ must contain at least 3/4 of the probability mass.

Chebyshev's inequality

There are many random variables X with a given mean μ and a given variance σ^2, but they all must satisfy the following inequality.

Theorem

Chebyshev's inequality Let X be a random variable with mean μ and variance σ^2. Then for any positive k,

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Letting $k = 2$ in Chebyshev's inequality gives

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \geq 1 - \frac{1}{4} = \frac{3}{4}$$

That is, the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ must contain at least 3/4 of the probability mass.

Exercise 4.17

4.17. You are to pay \$1.00 to play a game that consists of drawing one ticket at random from a box of numbered tickets. You win the amount (in dollars) of the number on the ticket you draw. The following two boxes of numbered tickets are available.

I. $\boxed{0, 1, 2}$ II. $\boxed{0, 0, 0, 1, 4}$

a. Find the expected value and variance of your net gain per play with box I.
b. Repeat part (a) for box II.
c. Given that you have decided to play, which box would you choose, and why?

Exercise 4.37

4.37. Four couples go to dinner together. The waiter seats the men randomly on one side of the table and the women randomly on the other side of the table. Find the expected value and variance of the number of couples who are seated across from each other.

Conditional Probability, Independence and Bayes' Theorem
Class 3, 18.05
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Know the definitions of conditional probability and independence of events.
2. Be able to compute conditional probability directly from the definition.
3. Be able to use the multiplication rule to compute the total probability of an event.
4. Be able to check if two events are independent.
5. Be able to use Bayes' formula to 'invert' conditional probabilities.
6. Be able to organize the computation of conditional probabilities using trees and tables.
7. Understand the base rate fallacy thoroughly.

2 Conditional Probability

Conditional probability answers the question 'how does the probability of an event change if we have extra information'. We'll illustrate with an example.

Example 1. Toss a fair coin 3 times.
(a) What is the probability of 3 heads?
answer: Sample space $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
All outcomes are equally likely, so $P(\text{3 heads}) = 1/8$.

(b) Suppose we are told that the first toss was heads. Given this information how should we compute the probability of 3 heads?
answer: We have a new (reduced) sample space: $\Omega' = \{HHH, HHT, HTH, HTT\}$.
All outcomes are equally likely, so

$$P(\text{3 heads given that the first toss is heads}) = 1/4.$$

This is called conditional probability, since it takes into account additional conditions. To develop the notation, we rephrase (b) in terms of *events*.

Rephrased (b) Let A be the event 'all three tosses are heads' $= \{HHH\}$.
Let B be the event 'the first toss is heads' $= \{HHH, HHT, HTH, HTT\}$.

The conditional probability of A knowing that B occurred is written

$$P(A|B)$$

This is read as

'the conditional probability of A given B'

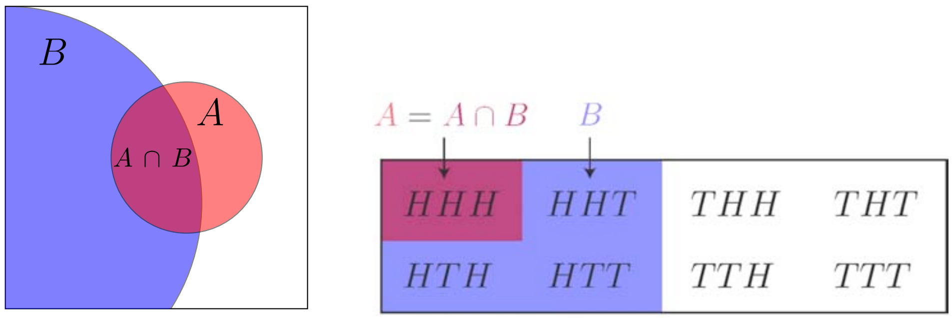
or

'the probability of A conditioned on B'

or simply

'the probability of A given B'.

We can visualize conditional probability as follows. Think of $P(A)$ as the proportion of the area of the *whole* sample space taken up by A. For $P(A|B)$ we restrict our attention to B. That is, $P(A|B)$ is the proportion of area of B taken up by A, i.e. $P(A \cap B)/P(B)$.



Conditional probability: Abstract visualization and coin example

Note, $A \subset B$ in the right-hand figure, so there are only two colors shown.

The formal definition of conditional probability catches the gist of the above example and visualization.

Formal definition of conditional probability
Let A and B be events. We define the conditional probability of A given B as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ provided } P(B) \neq 0. \tag{1}$$

Let's redo the coin tossing example using the definition in Equation (1). Recall A = '3 heads' and B = 'first toss is heads'. We have $P(A) = 1/8$ and $P(B) = 1/2$. Since $A \cap B = A$, we also have $P(A \cap B) = 1/8$. Now according to (1), $P(A|B) = \frac{1/8}{1/2} = 1/4$, which agrees with our answer in Example 1b.

3 Multiplication Rule

The following formula is called the multiplication rule.

$$P(A \cap B) = P(A|B) \cdot P(B). \tag{2}$$

This is simply a rewriting of the definition in Equation (1) of conditional probability. We will see that our use of the multiplication rule is very similar to our use of the rule of product in counting. In fact, the multiplication rule is just a souped up version of the rule of product.

We start with a simple example where we can check all the probabilities directly by counting.

Example 2. Draw two cards from a deck. Define the events: S_1 = 'first card is a spade' and S_2 = 'second card is a spade'. What is the $P(S_2|S_1)$?

answer: We can do this directly by counting: if the first card is a spade then of the 51 cards remaining, 12 are spades.

$$P(S_2|S_1) = 12/51.$$

Now, let's recompute this using formula (1). We have to compute $P(S_1)$, $P(S_2)$ and $P(S_1 \cap S_2)$: We know that $P(S_1) = 1/4$ because there are 52 equally likely ways to draw the first card and 13 of them are spades. The same logic says that there are 52 equally likely ways the second card can be drawn, so $P(S_2) = 1/4$.

Aside: The probability $P(S_2) = 1/4$ may seem surprising since the value of first card certainly affects the probabilities for the second card. However, if we look at *all* possible two card sequences we will see that every card in the deck has equal probability of being the second card. Since 13 of the 52 cards are spades we get $P(S_2) = 13/52 = 1/4$. Another way to say this is: if we are not given value of the first card then we have to consider all possibilities for the second card.

Continuing, we see that

$$P(S_1 \cap S_2) = \frac{13 \cdot 12}{52 \cdot 51} = 3/51.$$

This was found by counting the number of ways to draw a spade followed by a second spade and dividing by the number of ways to draw any card followed by any other card). Now, using (1) we get

$$P(S_2|S_1) = \frac{P(S_2 \cap S_1)}{P(S_1)} = \frac{3/51}{1/4} = 12/51.$$

Finally, we verify the multiplication rule by computing both sides of (2).

$$P(S_1 \cap S_2) = \frac{13 \cdot 12}{52 \cdot 51} = \frac{3}{51} \qquad \text{and} \qquad P(S_2|S_1) \cdot P(S_1) = \frac{12}{51} \cdot \frac{1}{4} = \frac{3}{51}. \qquad \text{QED}$$

Think: For S_1 and S_2 in the previous example, what is $P(S_2|S_1^c)$?

4 Law of Total Probability

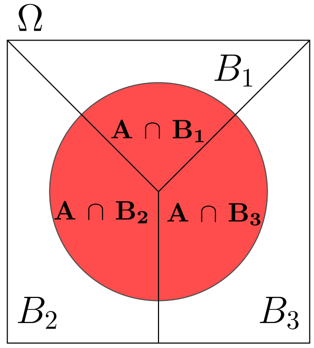
The law of total probability will allow us to use the multiplication rule to find probabilities in more interesting examples. It involves a lot of notation, but the idea is fairly simple. We state the law when the sample space is divided into 3 pieces. It is a simple matter to extend the rule when there are more than 3 pieces.

Law of Total Probability

Suppose the sample space Ω is divided into 3 disjoint events B_1, B_2, B_3 (see the figure below). Then for any event A:

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) \\ P(A) &= P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + P(A|B_3)\,P(B_3) \qquad (3) \end{aligned}$$

The top equation says 'if A is divided into 3 pieces then $P(A)$ is the sum of the probabilities of the pieces'. The bottom equation (3) is called the law of total probability. It is just a rewriting of the top equation using the multiplication rule.



The sample space Ω and the event A are each divided into 3 disjoint pieces.

The law holds if we divide Ω into any number of events, so long as they are *disjoint* and *cover* all of Ω. Such a division is often called a *partition* of Ω.

Our first example will be one where we already know the answer and can verify the law.

Example 3. An urn contains 5 red balls and 2 green balls. Two balls are drawn one after the other. What is the probability that the second ball is red?

answer: The sample space is $\Omega = \{\text{rr, rg, gr, gg}\}$.
Let R_1 be the event 'the first ball is red', G_1 = 'first ball is green', R_2 = 'second ball is red', G_2 = 'second ball is green'. We are asked to find $P(R_2)$.

The fast way to compute this is just like $P(S_2)$ in the card example above. Every ball is equally likely to be the second ball. Since 5 out of 7 balls are red, $P(R_2) = 5/7$.

Let's compute this same value using the law of total probability (3). First, we'll find the conditional probabilities. This is a simple counting exercise.

$$P(R_2|R_1) = 4/6, \qquad P(R_2|G_1) = 5/6.$$

Since R_1 and G_1 partition Ω the law of total probability says

$$\begin{aligned} P(R_2) &= P(R_2|R_1)P(R_1) + P(R_2|G_1)P(G_1) \qquad (4) \\ &= \frac{4}{6}\cdot\frac{5}{7} + \frac{5}{6}\cdot\frac{2}{7} \\ &= \frac{30}{42} = \frac{5}{7}. \end{aligned}$$

Probability urns

The example above used probability urns. Their use goes back to the beginning of the subject and we would be remiss not to introduce them. This toy model is very useful. We quote from Wikipedia: http://en.wikipedia.org/wiki/Urn_problem

> In probability and statistics, an urn problem is an idealized mental exercise in which some objects of real interest (such as atoms, people, cars, etc.) are represented as colored balls in an urn or other container. One pretends to draw (remove) one or more balls from the urn; the goal is to determine the probability of drawing one color or another, or some other properties. A key parameter is whether each ball is returned to the urn after each draw.

It doesn't take much to make an example where (3) is really the best way to compute the probability. Here is a game with slightly more complicated rules.

Example 4. An urn contains 5 red balls and 2 green balls. A ball is drawn. If it's green a red ball is added to the urn and if it's red a green ball is added to the urn. (The original ball is not returned to the urn.) Then a second ball is drawn. What is the probability the second ball is red?

answer: The law of total probability says that $P(R_2)$ can be computed using the expression in Equation (4). Only the values for the probabilities will change. We have

$$P(R_2|R_1) = 4/7, \qquad P(R_2|G_1) = 6/7.$$

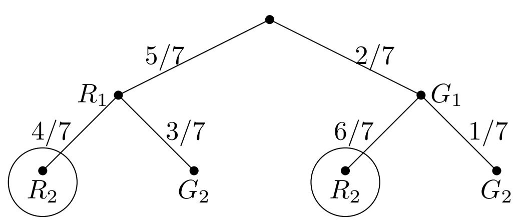
Therefore,

$$P(R_2) = P(R_2|R_1)P(R_1) + P(R_2|G_1)P(G_1) = \frac{4}{7} \cdot \frac{5}{7} + \frac{6}{7} \cdot \frac{2}{7} = \frac{32}{49}.$$

5 Using Trees to Organize the Computation

Trees are a great way to organize computations with conditional probability and the law of total probability. The figures and examples will make clear what we mean by a tree. As with the rule of product, the key is to organize the underlying process into a sequence of actions.

We start by redoing Example 4. The sequence of actions are: first draw ball 1 (and add the appropriate ball to the urn) and then draw ball 2.



You interpret this tree as follows. Each dot is called a node. The tree is organized by levels. The top node (root node) is at level 0. The next layer down is level 1 and so on. Each level shows the outcomes at one stage of the game. Level 1 shows the possible outcomes of the first draw. Level 2 shows the possible outcomes of the second draw starting from each node in level 1.

Probabilities are written along the branches. The probability of R_1 (red on the first draw) is 5/7. It is written along the branch from the root node to the one labeled R_1. At the next level we put in conditional probabilities. The probability along the branch from R_1 to R_2 is $P(R_2|R_1) = 4/7$. It represents the probability of going to node R_2 given that you are already at R_1.

The muliplication rule says that the probability of getting to any node is just the product of the probabilities along the path to get there. For example, the node labeled R_2 at the far left really represents the event $R_1 \cap R_2$ because it comes from the R_1 node. The multiplication rule now says

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2|R_1) = \frac{5}{7} \cdot \frac{4}{7},$$

which is exactly multiplying along the path to the node.

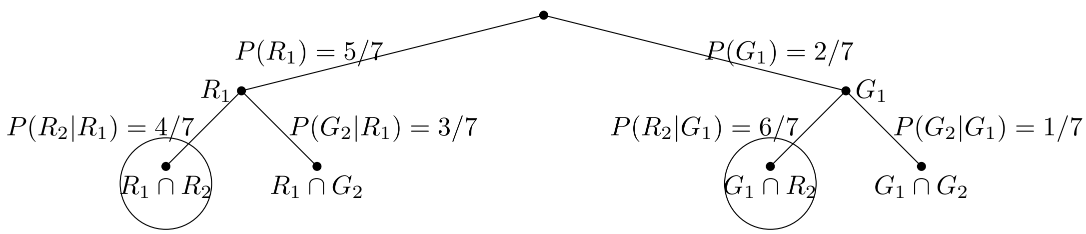
The law of total probability is just the statement that $P(R_2)$ is the sum of the probabilities of all paths leading to R_2 (the two circled nodes in the figure). In this case,

$$P(R_2) = \frac{5}{7}\cdot\frac{4}{7} + \frac{2}{7}\cdot\frac{6}{7} = \frac{32}{49},$$

exactly as in the previous example.

5.1 Shorthand vs. precise trees

The tree given above involves some shorthand. For example, the node marked R_2 at the far left really represents the event $R_1 \cap R_2$, since it ends the path from the root through R_1 to R_2. Here is the same tree with everything labeled precisely. As you can see this tree is more cumbersome to make and use. We usually use the shorthand version of trees. You should make sure you know how to interpret them precisely.



6 Independence

Two events are independent if knowledge that one occurred does not change the probability that the other occurred. Informally, events are independent if they do not influence one another.

Example 5. Toss a coin twice. We expect the outcomes of the two tosses to be independent of one another. In real experiments this always has to be checked. If my coin lands in honey and I don't bother to clean it, then the second toss might be affected by the outcome of the first toss.

More seriously, the independence of experiments can by undermined by the failure to clean or recalibrate equipment between experiments or to isolate supposedly independent observers from each other or a common influence. We've all experienced hearing the same 'fact' from different people. Hearing it from different sources tends to lend it credence until we learn that they all heard it from a common source. That is, our sources were not independent.

Translating the verbal description of independence into symbols gives

$$A \text{ is independent of } B \quad \text{if} \quad P(A|B) = P(A). \tag{5}$$

That is, knowing that B occurred does not change the probability that A occurred. In terms of events as subsets, knowing that the realized outcome is in B does not change the probability that it is in A.

If A and B are independent in the above sense, then the multiplication rule gives $P(A \cap B) = P(A|B) \cdot P(B) = P(A) \cdot P(B)$. This justifies the following technical definition of independence.

Formal definition of independence: Two events A and B are independent if

$$P(A \cap B) = P(A) \cdot P(B) \tag{6}$$

This is a nice symmetric definition which makes clear that A is independent of B if and only if B is independent of A. Unlike the equation with conditional probabilities, this definition makes sense even when $P(B) = 0$. In terms of conditional probabilities, we have:
1. If $P(B) \neq 0$ then A and B are independent if and only if $P(A|B) = P(A)$.
2. If $P(A) \neq 0$ then A and B are independent if and only if $P(B|A) = P(B)$.

Independent events commonly arise as different trials in an experiment, as in the following example.

Example 6. Toss a fair coin twice. Let H_1 = 'heads on first toss' and let H_2 = 'heads on second toss'. Are H_1 and H_2 independent?

answer: Since $H_1 \cap H_2$ is the event 'both tosses are heads' we have

$$P(H_1 \cap H_2) = 1/4 = P(H_1)P(H_2).$$

Therefore the events are independent.

We can ask about the independence of any two events, as in the following two examples.

Example 7. Toss a fair coin 3 times. Let H_1 = 'heads on first toss' and A = 'two heads total'. Are H_1 and A independent?

answer: We know that $P(A) = 3/8$. Since this is not 0 we can check if the formula in Equation 5 holds. Now, H_1 = {HHH, HHT, HTH, HTT} contains exactly two outcomes (HHT, HTH) from A, so we have $P(A|H_1) = 2/4$. Since $P(A|H_1) \neq P(A)$ these events are not independent.

Example 8. Draw one card from a standard deck of playing cards. Let's examine the independence of 3 events 'the card is an ace', 'the card is a heart' and 'the card is red'.

Define the events as A = 'ace', H = 'hearts', R = 'red'.

(a) We know that $P(A) = 4/52 = 1/13$, $P(A|H) = 1/13$. Since $P(A) = P(A|H)$ we have that A is independent of H.

(b) $P(A|R) = 2/26 = 1/13$. So A is independent of R. That is, whether the card is an ace is independent of whether it's red.

(c) Finally, what about H and R? Since $P(H) = 1/4$ and $P(H|R) = 1/2$, H and R are not independent. We could also see this the other way around: $P(R) = 1/2$ and $P(R|H) = 1$, so H and R are not independent.

6.1 Paradoxes of Independence

An event A with probability 0 is independent of itself, since in this case both sides of equation (6) are 0. This appears paradoxical because knowledge that A occurred certainly gives information about whether A occurred. We resolve the paradox by noting that since $P(A) = 0$ the statement 'A occurred' is vacuous.

Think: For what other value(s) of $P(A)$ is A independent of itself?

7 Bayes' Theorem

Bayes' theorem is a pillar of both probability and statistics and it is central to the rest of this course. For two events A and B Bayes' theorem (also called Bayes' rule and Bayes' formula) says

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}. \tag{7}$$

Comments: 1. Bayes' rule tells us how to 'invert' conditional probabilities, i.e. to find $P(B|A)$ from $P(A|B)$.
2. In practice, $P(A)$ is often computed using the law of total probability.

Proof of Bayes' rule
The key point is that $A \cap B$ is symmetric in A and B. So the multiplication rule says

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

Now divide through by $P(A)$ to get Bayes' rule.

A common mistake is to confuse $P(A|B)$ and $P(B|A)$. They can be very different. This is illustrated in the next example.

Example 9. Toss a coin 5 times. Let H_1 = 'first toss is heads' and let H_A = 'all 5 tosses are heads'. Then $P(H_1|H_A) = 1$ but $P(H_A|H_1) = 1/16$.

For practice, let's use Bayes' theorem to compute $P(H_1|H_A)$ using $P(H_A|A_1)$.The terms are $P(H_A|H_1) = 1/16$, $P(H_1) = 1/2$, $P(H_A) = 1/32$. So,

$$P(H_1|H_A) = \frac{P(H_A|H_1)P(H_1)}{P(H_A)} = \frac{(1/16) \cdot (1/2)}{1/32} = 1,$$

which agrees with our previous calculation.

7.1 The Base Rate Fallacy

The base rate fallacy is one of many examples showing that it's easy to confuse the meaning of $P(B|A)$ and $P(A|B)$ when a situation is described in words. This is one of the key examples from probability and it will inform much of our practice and interpretation of statistics. You should strive to understand it thoroughly.

Example 10. The Base Rate Fallacy
Consider a routine screening test for a disease. Suppose the frequency of the disease in the population (base rate) is 0.5%. The test is highly accurate with a 5% false positive rate and a 10% false negative rate.

You take the test and it comes back positive. What is the probability that you have the disease?

answer: We will do the computation three times: using trees, tables and symbols. We'll use the following notation for the relevant events:
$D+$ = 'you have the disease'
$D-$ = 'you do not have the disease
$T+$ = 'you tested positive'
$T-$ = 'you tested negative'.

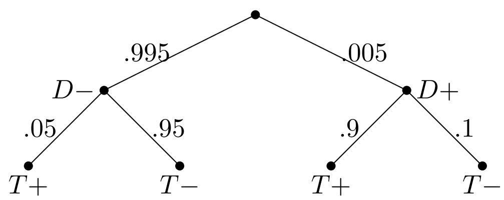
We are given $P(D+) = .005$ and therefore $P(D-) = .995$. The false positive and false negative rates are (by definition) conditional probabilities.

$$P(\text{false positive}) = P(T + |D-) = .05 \quad \text{and} \quad P(\text{false negative}) = P(T - |D+) = .1.$$

The complementary probabilities are known as the true negative and true positive rates:

$$P(T - |D-) = 1 - P(T + |D-) = .95 \quad \text{and} \quad P(T + |D+) = 1 - P(T - |D+) = .9.$$

Trees: All of these probabilities can be displayed quite nicely in a tree.



The question asks for the probability that you have the disease given that you tested positive, i.e. what is the value of $P(D+|T+)$. We aren't given this value, but we do know $P(T+|D+)$, so we can use Bayes' theorem.

$$P(D + |T+) = \frac{P(T + |D+) \cdot P(D+)}{P(T+)}.$$

The two probabilities in the numerator are given. We compute the denominator $P(T+)$ using the law of total probability. Using the tree we just have to sum the probabilities for each of the nodes marked $T+$

$$P(T+) = .995 \times .05 + .005 \times .9 = .05425$$

Thus,

$$P(D + |T+) = \frac{.9 \times .005}{.05425} = 0.082949 \approx 8.3\%.$$

Remarks: This is called the base rate fallacy because the base rate of the disease in the population is so low that the vast majority of the people taking the test are healthy, and even with an accurate test most of the positives will be healthy people. Ask your doctor for his/her guess at the odds.

To summarize the base rate fallacy with specific numbers

95% of all tests are accurate does not imply 95% of positive tests are accurate

We will refer back to this example frequently. It and similar examples are at the heart of many statistical misunderstandings.

Other ways to work Example 10

Tables: Another trick that is useful for computing probabilities is to make a table. Let's redo the previous example using a table built with 10000 total people divided according to the probabilites in this example.

We construct the table as follows. Pick a number, say 10000 people, and place it as the grand total in the lower right. Using $P(D+) = .005$ we compute that 50 out of the 10000 people are sick ($D+$). Likewise 9950 people are healthy ($D-$). At this point the table looks like:

	$D+$	$D-$	total
$T+$			
$T-$			
total	50	9950	10000

Using $P(T+|D+) = .9$ we can compute that the number of sick people who tested positive as 90% of 50 or 45. The other entries are similar. At this point the table looks like the table below on the left. Finally we sum the $T+$ and $T-$ rows to get the completed table on the right.

	$D+$	$D-$	total
$T+$	45	498	
$T-$	5	9452	
total	50	9950	10000

	$D+$	$D-$	total
$T+$	45	498	543
$T-$	5	9452	9457
total	50	9950	10000

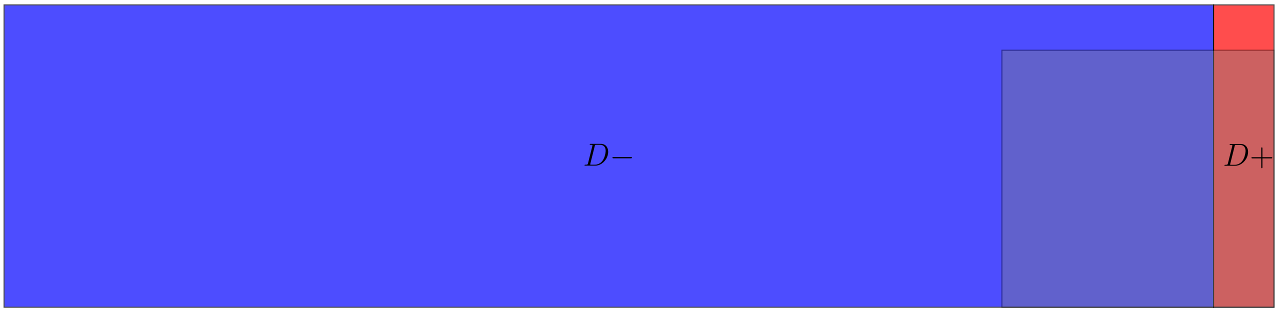
Using the complete table we can compute

$$P(D+|T+) = \frac{|D+ \cap T+|}{|T+|} = \frac{45}{543} = 8.3\%.$$

Symbols: For completeness, we show how the solution looks when written out directly in symbols.

$$\begin{aligned} P(D+|T+) &= \frac{P(T+|D+) \cdot P(D+)}{P(T+)} \\ &= \frac{P(T+|D+) \cdot P(D+)}{P(T+|D+) \cdot P(D+) + P(T+|D-) \cdot P(D-)} \\ &= \frac{.9 \times .005}{.9 \times .005 + .05 \times .995} \\ &= 8.3\% \end{aligned}$$

Visualization: The figure below illustrates the base rate fallacy. The large blue area represents all the healthy people. The much smaller red area represents the sick people. The shaded rectangle represents the the people who test positive. The shaded area covers most of the red area and only a small part of the blue area. Even so, the most of the shaded area is over the blue. That is, most of the positive tests are of healthy people.



7.2 Bayes' rule in 18.05

As we said at the start of this section, Bayes' rule is a pillar of probability and statistics. We have seen that Bayes' rule allows us to 'invert' conditional probabilities. When we learn statistics we will see that the art of statistical inference involves deciding how to proceed when one (or more) of the terms on the right side of Bayes' rule is unknown.

Chapter 2: Probability

The aim of this chapter is to revise the basic rules of probability. By the end of this chapter, you should be comfortable with:

- conditional probability, and what you can and can't do with conditional expressions;
- the Partition Theorem and Bayes' Theorem;
- First-Step Analysis for finding the probability that a process reaches some state, by conditioning on the outcome of the first step;
- calculating probabilities for continuous and discrete random variables.

2.1 Sample spaces and events

Definition: A **sample space**, Ω, is a *set of possible outcomes of a random experiment.*

Definition: An **event**, A, is a *subset of the sample space.*

This means that event A is simply *a collection of outcomes.*

Example:

Random experiment: Pick a person in this class at random.
Sample space: $\Omega = \{$*all people in class*$\}$
Event A: $A = \{$*all males in class*$\}$.

Definition: Event A **occurs** if *the outcome of the random experiment is a member of the set A.*

In the example above, event A occurs if *the person we pick is male.*

2.2 Probability Reference List

The following properties hold for all events A, B.

- $\mathbb{P}(\emptyset) = 0$.
- $0 \le \mathbb{P}(A) \le 1$.
- **Complement:** $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
- **Probability of a union:** $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
  For three events A, B, C:
  $$\mathbb{P}(A\cup B\cup C) = \mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)-\mathbb{P}(A\cap B)-\mathbb{P}(A\cap C)-\mathbb{P}(B\cap C)+\mathbb{P}(A\cap B\cap C)\,.$$
  If A and B are **mutually exclusive**, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- **Conditional probability:** $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- **Multiplication rule:** $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$.
- **The Partition Theorem:** if $B_1, B_2, \ldots, B_m$ form a partition of Ω, then
  $$\mathbb{P}(A) = \sum_{i=1}^{m} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i) \quad \text{for any event } A.$$
  As a special case, B and $\overline{B}$ partition Ω, so:
  $$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \\ &= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid \overline{B})\mathbb{P}(\overline{B}) \quad \text{for any } A, B. \end{aligned}$$
- **Bayes' Theorem:** $\mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}$.
  More generally, if $B_1, B_2, \ldots, B_m$ form a partition of Ω, then
  $$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)} \quad \text{for any } j.$$
- **Chains of events:** for any events $A_1, A_2, \ldots, A_n$,
  $$\mathbb{P}(A_1\cap A_2\cap\ldots\cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_2\cap A_1)\ldots\mathbb{P}(A_n \mid A_{n-1}\cap\ldots\cap A_1).$$

2.3 Conditional Probability

Suppose we are working with sample space $\Omega = \{\text{people in class}\}$. I want to find the proportion of people in the class who ski. What do I do?

Count up the number of people in the class who ski, and divide by the total number of people in the class.

Now suppose I want to find the proportion of *females* in the class who ski. What do I do?

By changing from asking about everyone to asking about females only, we have:

•

or:

or:

We could write the above as:

Conditioning is like changing the sample space: we are now working in a new sample space of females in class.

In the above example, we could replace 'skiing' with *any* attribute B. We have:

$$\mathbb{P}(\text{skis}) = \frac{\#\text{ skiers in class}}{\#\text{ class}}; \qquad \mathbb{P}(\text{skis}\,|\,\text{female}) = \frac{\#\text{ female skiers in class}}{\#\text{ females in class}};$$

so:

$$\mathbb{P}(B) =$$

and:

$$\mathbb{P}(B\,|\,\text{female}) =$$

Likewise, we could replace 'female' with any attribute A:

$$\mathbb{P}(B\,|\,A) =$$

This is how we get the definition of conditional probability:

$$\mathbb{P}(B\,|\,A) =$$

By conditioning on event A, we have

Definition: Let A and B be events on the same sample space: so
The **conditional probability of event B, given event A**, is

Multiplication Rule: (Immediate from above). For any events A and B,

Conditioning as 'changing the sample space'

The idea that can be very helpful in understanding how to manipulate conditional probabilities.

Any 'unconditional' probability can be written as a conditional probability:

Writing $\mathbb{P}(B) = \mathbb{P}(B \mid \Omega)$ just means that we are looking for the probability of event B, out of all possible outcomes in the set Ω.

In fact, the symbol $\mathbb{P}$ ***belongs*** to the set Ω: it has ***no meaning without*** Ω. To remind ourselves of this, we can write

Then

Similarly, $\mathbb{P}(B \mid A)$ means that we are looking for the probability of event B, out of all possible outcomes in the set

So A is just another sample space. Thus

The trick: Because we can think of A as just another sample space, let's write

Then ***we can use*** $\mathbb{P}_A$ ***just like*** $\mathbb{P}$***, as long as we remember to keep the*** A ***subscript on EVERY*** $\mathbb{P}$ ***that we write.***

This helps us to make quite complex manipulations of conditional probabilities without thinking too hard or making mistakes. There is only one rule you need to learn to use this tool effectively:

(Proof: Exercise).

The rules:

$$\boxed{\begin{array}{c}\mathbb{P}(\,\cdot \mid A) = \mathbb{P}_A(\,\cdot\,) \\ \mathbb{P}_A(B \mid C) = \mathbb{P}(B \mid C \cap A) \text{ for any } A,\, B,\, C.\end{array}}$$

Examples:

1. Probability of a union. In general,

   $\mathbb{P}(B \cup C) =$

   So,

   Thus,

2. Which of the following is equal to $\mathbb{P}(B \cap C \mid A)$?

   (a) $\mathbb{P}(B \mid C \cap A)$.

   (b) $\dfrac{\mathbb{P}(B \mid C)}{\mathbb{P}(A)}$.

   (c) $\mathbb{P}(B \mid C \cap A)\mathbb{P}(C \mid A)$.

   (d) $\mathbb{P}(B \mid C)\mathbb{P}(C \mid A)$.

   Solution:

3. Which of the following is true?

   (a) $\mathbb{P}(\overline{B} \mid A) = 1 - \mathbb{P}(B \mid A)$. (b) $\mathbb{P}(\overline{B} \mid A) = \mathbb{P}(B) - \mathbb{P}(B \mid A)$.

   Solution:

4. Which of the following is true?

   (a) $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(A) - \mathbb{P}(B \cap A)$. (b) $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(B) - \mathbb{P}(B \cap A)$.

   Solution:

5. True or false: $\mathbb{P}(B \mid A) = 1 - \mathbb{P}(B \mid \overline{A})$?

   Answer:

Exercise: if we wish to express $\mathbb{P}(B \mid A)$ in terms of only B and $\overline{A}$, show that
$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B) - \mathbb{P}(B \mid \overline{A})\mathbb{P}(\overline{A})}{1 - \mathbb{P}(\overline{A})}.$$
Note that this does not simplify nicely!

2.4 The Partition Theorem (Law of Total Probability)

Definition: Events A and B are **mutually exclusive**, or **disjoint**, if

This means events A and B cannot happen together. If A happens, it excludes B from happening, and vice-versa.

If A and B are mutually exclusive,
For all other A and B,

Definition: Any number of events $B_1, B_2, \ldots, B_k$ are **mutually exclusive** if every pair of the events is mutually exclusive: ie. $B_i \cap B_j = \emptyset$ *for all* i, j *with* $i \neq j$.

Definition: A **partition** of Ω is a

That is, sets $B_1, B_2, \ldots, B_k$ form a partition of Ω if

$$B_i \cap B_j = \emptyset \text{ for all } i, j \text{ with } i \neq j,$$

$$\textbf{and} \quad \bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots \cup B_k = \Omega.$$

$B_1, \ldots, B_k$ form a partition of Ω if they

Examples:

Partitioning an event A

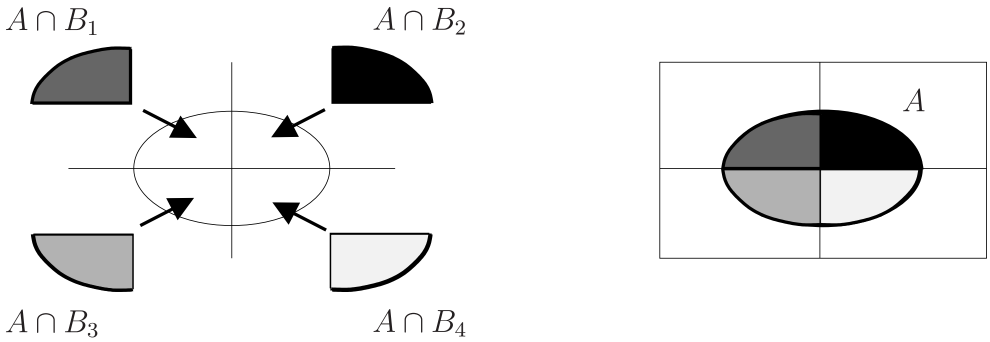
Any set A can be partitioned: it doesn't have to be Ω.
In particular, if $B_1, \ldots, B_k$ form a partition of Ω, then $(A \cap B_1), \ldots, (A \cap B_k)$ form a partition of A.

Theorem 2.4: The Partition Theorem (Law of Total Probability)

Both formulations of the Partition Theorem are very widely used, but especially the conditional formulation $\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$.

Intuition behind the Partition Theorem:

The Partition Theorem is easy to understand because it simply states that "the whole is the sum of its parts."



$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3) + \mathbb{P}(A \cap B_4).$$

2.5 Bayes' Theorem: inverting conditional probabilities

Bayes' Theorem allows us to "invert" a conditional statement, ie. *to express* $\mathbb{P}(B \mid A)$ *in terms of* $\mathbb{P}(A \mid B)$.

Theorem 2.5: Bayes' Theorem

For any events A and B:

$$\boxed{\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}.}$$

Proof:

$$\begin{aligned} \mathbb{P}(B \cap A) &= \mathbb{P}(A \cap B) \\ \mathbb{P}(B \mid A)\mathbb{P}(A) &= \mathbb{P}(A \mid B)\mathbb{P}(B) \qquad \text{(multiplication rule)} \\ \therefore \qquad \mathbb{P}(B \mid A) &= \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}. \qquad \square \end{aligned}$$

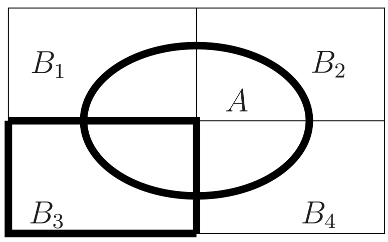
Extension of Bayes' Theorem

Suppose that $B_1, B_2, \ldots, B_m$ form a partition of Ω. By the Partition Theorem,

$$\mathbb{P}(A) = \sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Thus, for ***any single partition member*** B_j, put $B = B_j$ in Bayes' Theorem to obtain:

$$\boxed{\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_{i=1}^{m} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.}$$



Special case: $m = 2$

Given any event B, the events B and $\overline{B}$ form a partition of Ω. Thus:

$$\boxed{\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid \overline{B})\mathbb{P}(\overline{B})}.}$$

Example: In screening for a certain disease, the probability that a healthy person wrongly gets a positive result is 0.05. The probability that a diseased person wrongly gets a negative result is 0.002. The overall rate of the disease in the population being screened is 1%. If my test gives a positive result, what is the probability I actually have the disease?

2.6 First-Step Analysis for calculating probabilities in a process

In a stochastic process, what happens at the next step depends upon the current state of the process. We often wish to know the probability of *eventually* reaching some particular state, given our current position.

Throughout this course, we will tackle this sort of problem using a technique called

The idea is to consider all possible first steps away from the current state. We derive a system of equations that specify the probability of the eventual outcome given each of the possible first steps. We then try to solve these equations for the probability of interest.

First-Step Analysis depends upon ***conditional probability*** and the ***Partition Theorem.*** Let $S_1, \ldots, S_k$ be the k possible first steps we can take away from our current state. We wish to find the probability that event E happens eventually. First-Step Analysis calculates $\mathbb{P}(E)$ as follows:

Here, $\mathbb{P}(S_1), \ldots, \mathbb{P}(S_k)$ give the probabilities of taking the different first steps $1, 2, \ldots, k$.

Example: Tennis game at Deuce.

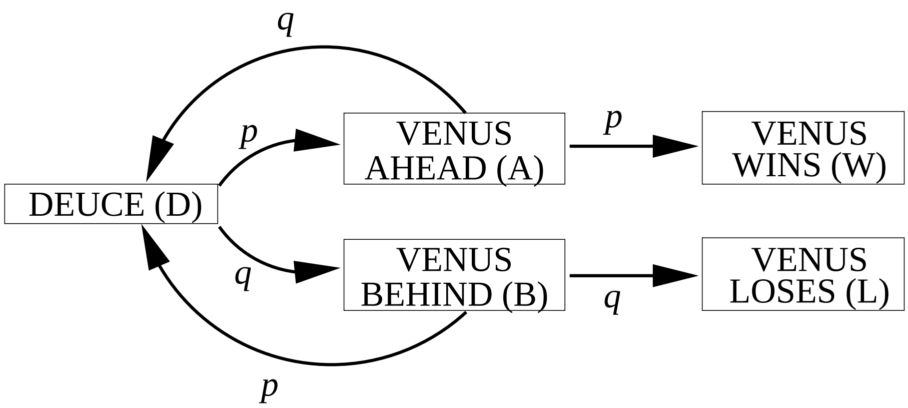
Venus and Serena are playing tennis, and have reached the score Deuce (40-40). (*Deuce* comes from the French word *Deux* for 'two', meaning that each player needs to win two consecutive points to win the game.)

For each point, let:

$$p = \mathbb{P}(\text{Venus wins point}), \qquad q = 1 - p = \mathbb{P}(\text{Serena wins point}).$$

Assume that all points are independent.

Let v be the probability that Venus wins the game eventually, starting from Deuce. Find v.


DEUCE (D)
VENUS AHEAD (A)
VENUS BEHIND (B)
VENUS WINS (W)
VENUS LOSES (L)
p
q
p
q
p
q

Note: Because $p + q = 1$, we have:

So the final probability that Venus wins the game is:

Note how this result makes intuitive sense. For the game to finish from Deuce, either Venus has to win two points in a row (probability p^2), or Serena does (probability q^2). The ratio $p^2/(p^2 + q^2)$ describes Venus's 'share' of the winning probability.

First-step analysis as the Partition Theorem:

Our approach to finding $v = \mathbb{P}(\text{Venus wins})$ can be summarized as:

First-step analysis is just the **Partition Theorem:**

An example of a sample point is:
Another example is:

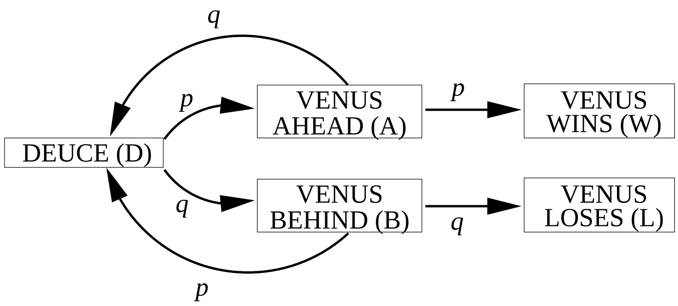
The **partition** of the sample space that we use in first-step analysis is:

Then first-step analysis simply states:

Notation for quick solutions of first-step analysis problems

Defining a helpful notation is central to modelling with stochastic processes. Setting up well-defined notation helps you to solve problems quickly and easily. Defining your notation is one of the most important steps in modelling, because it provides the conversion from words (which is how your problem starts) to mathematics (which is how your problem is solved).

Several marks are allotted on first-step analysis questions for setting up a well-defined and helpful notation.



Here is the correct way to formulate and solve this first-step analysis problem.

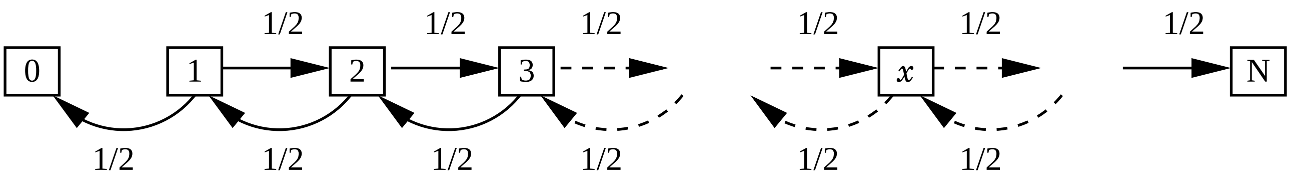
2.7 Special Process: the Gambler's Ruin


With permission, copyright Martin Ouellet

This is a famous problem in probability. A gambler starts with $\$x$. She tosses a fair coin repeatedly.

If she gets a Head, she wins \$1. If she gets a Tail, she loses \$1.

The coin tossing is repeated until the gambler has either \$0 or $\$N$, when she stops. What is the probability of the Gambler's Ruin, i.e. that the gambler ends up with \$0?



Solution of difference equation $(\star)$:

$$\begin{aligned} p_x &= \tfrac{1}{2}p_{x+1} + \tfrac{1}{2}p_{x-1} \quad \text{for } x = 1, 2, \ldots, N-1\,; \\ p_0 &= 1 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\star) \\ p_N &= 0. \end{aligned}$$

We usually solve equations like this using the theory of 2nd-order difference equations. For this special case we will also verify the answer by two other methods.

1. Theory of linear 2nd order difference equations

Theory tells us that the general solution of $(\star)$ is $p_x = A + Bx$ for some constants A, B and for $x = 0, 1, \ldots, N$. Our job is to find A and B using the boundary conditions:

So our solution is:

For Stats 325, you will be told the general solution of the 2nd-order difference equation and expected to solve it using the boundary conditions.

For Stats 721, we will study the theory of 2nd-order difference equations. You will be able to derive the general solution for yourself before solving it.

Question: What is the probability that the gambler wins (ends with \$N), starting with \$x?

2. Solution by inspection

The problem shown in this section is the ***symmetric*** Gambler's Ruin, where the probability is $\frac{1}{2}$ of moving up or down on any step. For this special case, we can solve the difference equation by inspection.

We have:

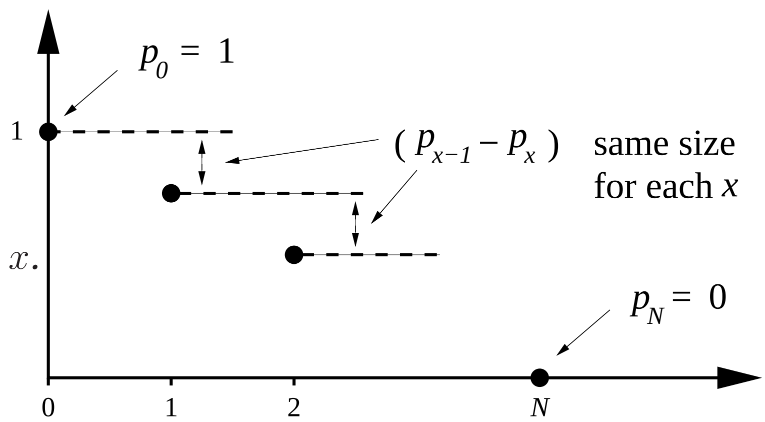
$$\begin{aligned} p_x &= \tfrac{1}{2}p_{x+1} + \tfrac{1}{2}p_{x-1} \\ \tfrac{1}{2}p_x + \tfrac{1}{2}p_x &= \tfrac{1}{2}p_{x+1} + \tfrac{1}{2}p_{x-1} \\ \text{Rearranging:} \quad p_{x-1} - p_x &= p_x - p_{x+1}. \end{aligned}$$

Boundaries: $p_0 = 1, p_N = 0.$

There are N *steps to go down from* $p_0 = 1$ *to* $p_N = 0$.
Each step is the same size, because
$(p_{x-1} - p_x) = (p_x - p_{x+1})$ *for all* x.
So each step has size $1/N$*,*
$\Rightarrow \quad p_0 = 1, p_1 = 1 - 1/N,$
$p_2 = 1 - 2/N$*, etc.*

So

$$p_x = 1 - \frac{x}{N} \quad \textit{as before.}$$



3. Solution by repeated substitution.

In principle, all systems could be solved by this method, but it is usually too tedious to apply in practice.

Rearrange $(\star)$ *to give:*

$$
\begin{array}{rrcl}
 & p_{x+1} & = & 2p_x - p_{x-1} \\
\Rightarrow \quad (x=1) & p_2 & = & 2p_1 - 1 \qquad \textit{(recall } p_0 = 1\textit{)} \\
(x=2) & p_3 & = & 2p_2 - p_1 = 2(2p_1 - 1) - p_1 = 3p_1 - 2 \\
(x=3) & p_4 & = & 2p_3 - p_2 = 2(3p_1 - 2) - (2p_1 - 1) = 4p_1 - 3 \qquad \textit{etc} \\
 & & \vdots & \\
\textit{giving} & p_x & = & xp_1 - (x-1) \qquad \textit{in general,} \qquad (\star\star) \\
\textit{likewise} & p_N & = & Np_1 - (N-1) \qquad \textit{at endpoint.}
\end{array}
$$

Boundary condition: $p_N = 0 \quad \Rightarrow \quad Np_1 - (N-1) = 0 \quad \Rightarrow \quad p_1 = 1 - 1/N.$

Substitute in $(\star\star)$*:*

$$
\begin{array}{rcl}
p_x & = & xp_1 - (x-1) \\
 & = & x\left(1 - \frac{1}{N}\right) - (x-1) \\
 & = & x - \frac{x}{N} - x + 1 \\
p_x & = & 1 - \frac{x}{N} \qquad \textit{as before.} \qquad \square
\end{array}
$$

2.8 Independence

Definition: Events A and B are **statistically independent** if and only if

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

This implies that A and B are statistically independent if and only if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.

Note: If events are *physically* independent, they will also be statistically indept.

For interest: more than two events

Definition: For more than two events, $A_1, A_2, \ldots, A_n$, we say that $A_1, A_2, \ldots, A_n$ are **mutually independent** if

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \quad \textit{for ALL finite subsets } J \subseteq \{1, 2, \ldots, n\}.$$

Example: events A_1, A_2, A_3, A_4 are mutually independent if

i) $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all i, j with $i \neq j$; AND

ii) $\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i)\mathbb{P}(A_j)\mathbb{P}(A_k)$ for all i, j, k that are all different; AND

iii) $\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)\mathbb{P}(A_4)$.

Note: For mutual independence, it is ***not*** enough to check that $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all $i \neq j$. Pairwise independence does not imply mutual independence.

2.9 The Continuity Theorem

The Continuity Theorem states that probability is a *continuous set function*:

Theorem 2.9: The Continuity Theorem

a) Let $A_1, A_2, \ldots$ be an *increasing sequence of events*: i.e.

$$A_1 \subseteq A_2 \subseteq \ldots \subseteq A_n \subseteq A_{n+1} \subseteq \ldots .$$

Then

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Note: because $A_1 \subseteq A_2 \subseteq \ldots$, we have: $\displaystyle \lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

b) Let $B_1, B_2, \ldots$ be a *decreasing sequence of events*: i.e.

$$B_1 \supseteq B_2 \supseteq \ldots \supseteq B_n \supseteq B_{n+1} \supseteq \ldots .$$

Then

$$\mathbb{P}\left(\lim_{n\to\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Note: because $B_1 \supseteq B_2 \supseteq \ldots$, we have: $\displaystyle\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n.$

Proof (a) only: for (b), take complements and use (a).

Define $C_1 = A_1$, and $C_i = A_i \backslash A_{i-1}$ for $i = 2, 3, \ldots$. Then $C_1, C_2, \ldots$ are mutually exclusive, and $\bigcup_{i=1}^{n} C_i = \bigcup_{i=1}^{n} A_i$, and likewise, $\bigcup_{i=1}^{\infty} C_i = \bigcup_{i=1}^{\infty} A_i$.

Thus

$$\begin{aligned}
\mathbb{P}(\lim_{n\to\infty} A_n) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} C_i\right) &= \sum_{i=1}^{\infty} \mathbb{P}(C_i) \quad (C_i \text{ mutually exclusive}) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(C_i) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} C_i\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n). \quad \square
\end{aligned}$$

2.10 Random Variables

Definition: A **random variable**, X, is defined as *a function from the sample space to the real numbers:* $X : \Omega \to \mathbb{R}$.

A random variable therefore *assigns a real number to every possible outcome of a random experiment.*

A random variable is essentially *a rule or mechanism for generating random real numbers.*

The Distribution Function

Definition: The **cumulative distribution function** of a random variable X is given by

$$F_X(x) = \mathbb{P}(X \leq x)$$

$F_X(x)$ is often referred to as simply the **distribution function**.

Properties of the distribution function

1) $F_X(-\infty) = \mathbb{P}(X \leq -\infty) = 0$.
   $F_X(+\infty) = \mathbb{P}(X \leq \infty) = 1$.

2) $F_X(x)$ is a non-decreasing function of x:
   if $x_1 < x_2$*, then* $F_X(x_1) \leq F_X(x_2)$.

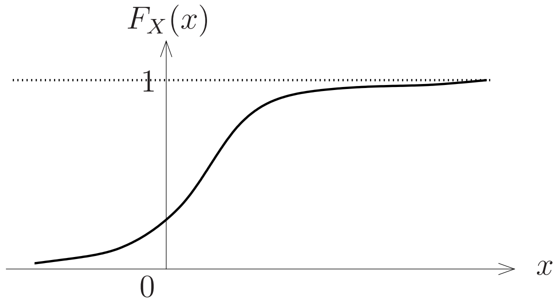
3) If $b > a$, then $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.

4) F_X is right-continuous: i.e. $\lim_{h \downarrow 0} F_X(x + h) = F_X(x)$.

2.11 Continuous Random Variables

Definition: The random variable X is **continuous** if *the distribution function $F_X(x)$ is a continuous function.*

In practice, this means that a continuous random variable *takes values in a continuous subset of $\mathbb{R}$: e.g. $X : \Omega \to [0,1]$ or $X : \Omega \to [0,\infty)$.*

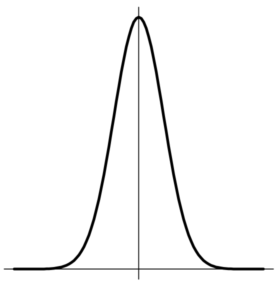


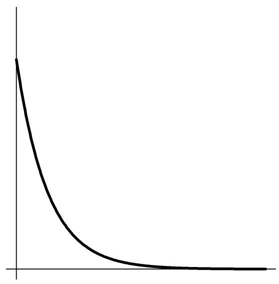
Probability Density Function for continuous random variables

Definition: Let X be a continuous random variable with continuous distribution function $F_X(x)$. The **probability density function (p.d.f.)** of X is defined as

$$f_X(x) = F'_X(x) = \frac{d}{dx}(F_X(x))$$

The pdf, $f_X(x)$, gives the *shape* of the distribution of X.


Normal distribution


Exponential distribution


Gamma distribution

By the Fundamental Theorem of Calculus, the distribution function $F_X(x)$ can be written in terms of the probability density function, $f_X(x)$, as follows:

$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(u)\, du}$$

Endpoints of intervals

For continuous random variables, every point x has $\mathbb{P}(X = x) = 0$. This means that the endpoints of intervals are not important for continuous random variables.

Thus, $\mathbb{P}(a \le X \le b) = \mathbb{P}(a < X \le b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a < X < b)$.

This is ***only*** true for ***continuous*** random variables.

Calculating probabilities for continuous random variables

To calculate $\mathbb{P}(a \le X \le b)$, use ***either***

$$\mathbb{P}(a \le X \le b) = F_X(b) - F_X(a)$$

or

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\, dx$$

Example: Let X be a continuous random variable with p.d.f.

$$f_X(x) = \begin{cases} 2x^{-2} & \text{for } 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

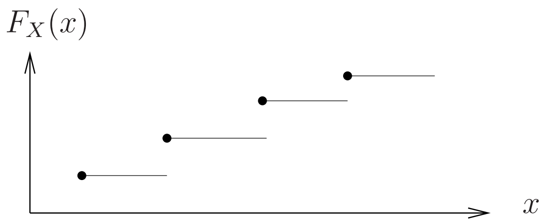
(a) Find the cumulative distribution function, $F_X(x)$.

(b) Find $\mathbb{P}(X \le 1.5)$.

2.12 Discrete Random Variables

Definition: The random variable X is **discrete** if X *takes values in a finite or countable subset of* $\mathbb{R}$*: thus,* $X : \Omega \to \{x_1, x_2, \ldots\}$.

When X is a discrete random variable, the distribution function $F_X(x)$ is a *step function.*



Probability function

Definition: Let X be a discrete random variable with distribution function $F_X(x)$. The **probability function** of X is defined as

$$f_X(x) = \mathbb{P}(X = x).$$

Endpoints of intervals

For discrete random variables, *individual points can have* $\mathbb{P}(X = x) > 0$.

This means that *the endpoints of intervals ARE important for discrete random variables.*

For example, if X takes values $0, 1, 2, \ldots$, and a, b are integers with $b > a$, then

$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a-1 < X \leq b) = \mathbb{P}(a \leq X < b+1) = \mathbb{P}(a-1 < X < b+1)$.

Calculating probabilities for discrete random variables

To calculate $\mathbb{P}(X \in A)$ for any countable set A, use

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

Partition Theorem for probabilities of discrete random variables

Recall the Partition Theorem: for any event A, and for events $B_1, B_2, \ldots$ that form a ***partition*** of Ω,

We can use the Partition Theorem to find probabilities for random variables. Let X and Y be **discrete** random variables.

-
-
-

2.13 Independent Random Variables

Random variables X and Y are independent if they have no effect on each other. This means that the probability that they both take specified values simultaneously is the product of the individual probabilities.

Definition: Let X and Y be random variables. The **joint distribution function** of X and Y is given by

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x \text{ and } Y \leq y) = \mathbb{P}(X \leq x, Y \leq y).$$

Definition: Let X and Y be any random variables (continuous or discrete). X and Y are **independent** if

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \textit{ for ALL } x, y \in \mathbb{R}.$$

If X and Y are ***discrete***, they are independent if and only if their joint probability function is the product of their individual probability functions:

$$\begin{aligned} \textit{Discrete } X, Y \textit{ are indept} \quad &\iff \quad \mathbb{P}(X = x \textit{ AND } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &\qquad\qquad \textit{for ALL } x, y \\ &\iff \quad f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \textit{for ALL } x, y. \end{aligned}$$