Part IA: Mathematics for Natural Sciences A
Examples Sheet 9: Discrete and continuous random variables, and functions of random variables

Model Solutions

Please send all comments and corrections to jmm232@cam.ac.uk.

Discrete random variables, expectation, and variance

1. A complex experiment consists of two independent stages, each of which involves the random generation of a number. In the first stage, the possible numbers are $0, 1, 2$, with probabilities $1/2, 1/4, 1/4$ respectively. In the second stage, the possible numbers are $2, 3$, with probabilities $2/3, 1/3$ respectively.

Let X_1, X_2 denote the results of the first and second stages of the experiment, respectively. Let $X = 2X_1 - X_2$, that is, let X be the random variable given by taking twice the result of the first stage, and then subtracting the result of the second stage.

(a) Find and sketch the probability mass function of X.

(b) Find and sketch the cumulative distribution function of X.

(c) Compute the probability $\mathbb{P}(0 \leq X \leq 2)$ by: (i) summing values of the probability mass function; (ii) taking the difference of two values of the cumulative distribution function.

(d) Define the *expectation* (or *mean*) $\mathbb{E}[X]$ of a discrete random variable X, taking values $x_1, ..., x_n$. Find $\mathbb{E}[X]$ for the variable X defined above.

(e) Give two expressions for the *variance* $\mathrm{Var}[X]$ of a discrete random variable X, and prove that they are equivalent. Find $\mathrm{Var}[X]$ for the random variable X defined above.

(f) Verify that:

$$\mathbb{E}[X] = 2\mathbb{E}[X_1] - \mathbb{E}[X_2], \qquad \mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2], \qquad \mathrm{Var}[X] = 4\mathrm{Var}[X_1] + \mathrm{Var}[X_2].$$

Are you expecting these results? Do they hold more generally?

Solution: The variables X_1, X_2 have the distributions:

$$\mathbb{P}(X_1 = 0) = \frac{1}{2}, \qquad \mathbb{P}(X_1 = 1) = \frac{1}{4}, \qquad \mathbb{P}(X_1 = 2) = \frac{1}{4},$$

and:

$$\mathbb{P}(X_2 = 2) = \frac{2}{3}, \qquad \mathbb{P}(X_2 = 3) = \frac{1}{3}.$$

The possible pairs of (X_1, X_2) are therefore $(0, 2)$, $(0, 3)$, $(1, 2)(1, 3)$, $(2, 2)$, $(2, 3)$ which by independence occur with the following probabilities:

$$\mathbb{P}((X_1, X_2) = (0, 2)) = \frac{1}{3}, \qquad \mathbb{P}((X_1, X_2) = (0, 3)) = \frac{1}{6}, \qquad \mathbb{P}((X_1, X_2) = (1, 2)) = \frac{1}{6},$$

$$\mathbb{P}((X_1, X_2) = (1, 3)) = \frac{1}{12}, \qquad \mathbb{P}((X_1, X_2) = (2, 2)) = \frac{1}{6}, \qquad \mathbb{P}((X_1, X_2) = (2, 3)) = \frac{1}{12}.$$

This gives the corresponding distribution of X:

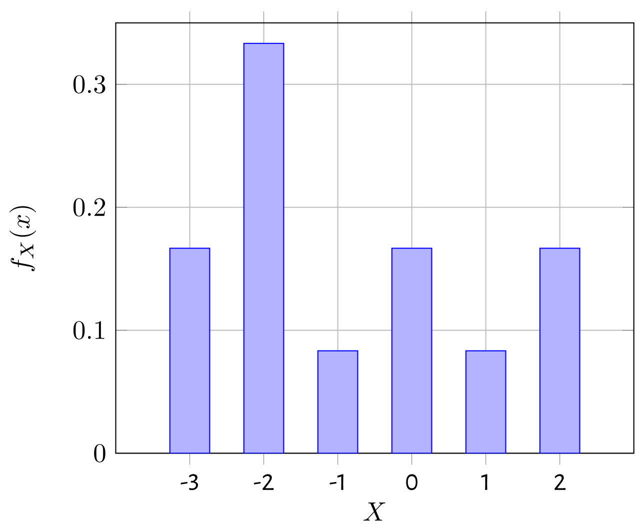
$$\mathbb{P}(X=-2)=\frac{1}{3}, \qquad \mathbb{P}(X=-3)=\frac{1}{6}, \qquad \mathbb{P}(X=0)=\frac{1}{6},$$
$$\mathbb{P}(X=-1)=\frac{1}{12}, \qquad \mathbb{P}(X=2)=\frac{1}{6}, \qquad \mathbb{P}(X=1)=\frac{1}{12}.$$

This will be very helpful throughout the rest of the question!

(a) The *probability mass function* is by definition $f_X(x)=\mathbb{P}(X=x)$ for a discrete random variable. Using the values we computed above, this gives the function $f_X:\{-3,-2,-1,0,1,2\}\to[0,1]$ given by:

$$f_X(x)=\begin{cases}1/6, & x=-3,\\ 1/3, & x=-2,\\ 1/12, & x=-1,\\ 1/6, & x=0,\\ 1/12, & x=1,\\ 1/6, & x=2.\end{cases}$$

Sketching this function, we have:



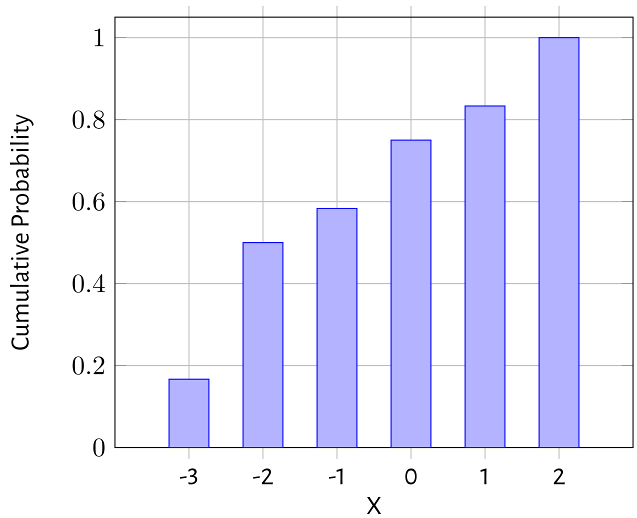
(b) The *cumulative distribution function* is by definition $F_X(x)=\mathbb{P}(X\le x)$ for a discrete random variable. This can be rewritten in the form:

$$F_X(x)=\sum_{y\le x}\mathbb{P}(X=y)=\sum_{y\le x}f_X(y),$$

where $f_X(y)$ is the probability mass function we found in part (a). Therefore, the cumulative distribution function is given by:

$$F_X(x)=\begin{cases}1/6, & x=-3,\\ 1/2, & x=-2,\\ 7/12, & x=-1,\\ 3/4, & x=0,\\ 5/6, & x=1,\\ 1, & x=2.\end{cases}$$

The required sketch is:



(c) The probability $\mathbb{P}(0 \leq X \leq 2)$ can be computed in two ways. First, we can sum values of the mass function:

$$\mathbb{P}(0 \leq X \leq 2) = f_X(0) + f_X(1) + f_X(2) = \frac{1}{6} + \frac{1}{12} + \frac{1}{6} = \frac{5}{12}.$$

Alternatively, we can take the difference of two values of the cumulative distribution function:

$$\mathbb{P}(0 \leq X \leq 2) = \mathbb{P}(X \leq 2) - \mathbb{P}(X \leq -1) = F_X(2) - F_X(-1) = 1 - \frac{7}{12} = \frac{5}{12}.$$

(d) The *expectation* of a discrete random variable taking values $x_1, x_2, \ldots, x_n$ is defined by:

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i).$$

Computing the expectation for this question, we have:

$$\mathbb{E}[X] = -\frac{3}{6} - \frac{2}{3} - \frac{1}{12} + \frac{1}{12} + \frac{2}{6} = -\frac{5}{6}.$$

(e) The *variance* of a discrete random variable taking values $x_1, x_2, \ldots, x_n$ is defined by either of the two equivalent formulae:

$$\mathrm{Var}[X] = \sum_{i=1}^{n} (x_i - \mathbb{E}[X])^2 \, \mathbb{P}(X = x_i) = \sum_{i=1}^{n} \left(x_i^2 \mathbb{P}(X = x_i)\right) - (\mathbb{E}[X])^2 .$$

To show these formulae are equivalent, we can expand the first as:

$$\sum_{i=1}^{n} x_i^2 \mathbb{P}(X = x_i) - 2\mathbb{E}[X] \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) + (\mathbb{E}[X])^2 \sum_{i=1}^{n} \mathbb{P}(X = x_i) = \sum_{i=1}^{n} x_i^2 \mathbb{P}(X = x_i) - (\mathbb{E}[X])^2 ,$$

using the definition of the expectation and the fact that the sum of all probabilities must be one.

For the random variable X, we can use either formula to calculate the variance. It is a bit easier to use the second though; we have:

$$\mathrm{Var}[X] = \frac{9}{6} + \frac{4}{3} + \frac{1}{12} + \frac{1}{12} + \frac{4}{6} - \left(-\frac{5}{6}\right)^2 = \frac{107}{36}.$$

(f) To verify the first fact, we will need both $\mathbb{E}[X_1]$ and $\mathbb{E}[X_2]$. We have:

$$\mathbb{E}[X_1] = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}, \qquad \mathbb{E}[X_2] = \frac{4}{3} + 1 = \frac{7}{3}.$$

Hence:

$$2\mathbb{E}[X_1] - \mathbb{E}[X_2] = \frac{3}{2} - \frac{7}{3} = -\frac{5}{6},$$

which agrees with what we found earlier.

To verify the second fact, we need to also compute $\mathbb{E}[X_1X_2]$. Using the probabilities of the pairs we calculated earlier, we have:

$$\mathbb{P}(X_1X_2 = 0) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}, \qquad \mathbb{P}(X_1X_2 = 2) = \frac{1}{6}, \qquad \mathbb{P}(X_1X_2 = 3) = \frac{1}{12}$$
$$\mathbb{P}(X_1X_2 = 4) = \frac{1}{6}, \qquad \mathbb{P}(X_1X_2 = 6) = \frac{1}{12}.$$

Hence the required expectation is:

$$\mathbb{E}[X_1X_2] = \frac{2}{6} + \frac{3}{12} + \frac{4}{6} + \frac{6}{12} = \frac{21}{12} = \frac{7}{4}.$$

This agrees with $\mathbb{E}[X_1]\mathbb{E}[X_2] = (3/4) \cdot (7/3)$, as required.

To verify the third fact, recall that we found earlier that $\mathrm{Var}[X] = 107/36$. Alternatively, we could also note that:

$$\mathrm{Var}[X_1] = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \frac{1}{4} + \frac{1}{4} \cdot 4 - \left(\frac{3}{4}\right)^2 = \frac{11}{16},$$

and:

$$\mathrm{Var}[X_2] = \mathbb{E}[X_2^2] - \mathbb{E}[X_2]^2 = \frac{2}{3} \cdot 4 + \frac{1}{3} \cdot 9 - \left(\frac{7}{3}\right)^2 = \frac{2}{9}.$$

Hence we have:

$$4\mathrm{Var}[X_1] + \mathrm{Var}[X_2] = \frac{11}{4} + \frac{2}{9} = \frac{107}{36},$$

as expected.

All of these facts hold more generally. Indeed:

- For any constants a, b and any random variables X, Y (not necessarily independent), we have:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

- For any two independent random variables X, Y, we have:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- For any two independent random variables X, Y and constants a, b, we have:

$$\mathrm{Var}[aX + bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y].$$

To see the proofs, do the Maths B sheet!

2. Let X be the result of a roll of a biased die, which displays one with probability $p/2$, two, three, four or five with probability p, and six with probability $2p$. Compute $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

➾ **Solution:** Since the sum of all the probabilities must be one, we require:

$$\frac{p}{2} + 4p + 2p = 1 \qquad \Leftrightarrow \qquad p = \frac{2}{13}.$$

We can use this to compute the expectation. We have:

$$\mathbb{E}[X] = \frac{1}{13} + \frac{4}{13} + \frac{6}{13} + \frac{8}{13} + \frac{10}{13} + \frac{24}{13} = \frac{53}{13}.$$

We can then use the expectation to compute the variance. We have:

$$\mathrm{Var}[X] = \frac{1}{13} + \frac{8}{13} + \frac{18}{13} + \frac{32}{13} + \frac{50}{13} + \frac{144}{13} - \left(\frac{53}{13}\right)^2 = \frac{480}{169}.$$

---

3. Three standard 6-sided dice are tossed onto a table. Calculate the mean and variance of:

(a) the sum of the values shown by the dice;

(b) the sum of the squares of the values shown by the dice.

[*Hint: use the linearity of expectation*.]

➾ **Solution:** Let X_1 be the result of the first die, X_2 be the result of the second die, and X_3 the result of the third die. Observe that each of the variables X_1, X_2, X_3 has exactly the same distribution. Further, the expectation of each is given by:

$$\mathbb{E}[X_i] = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{21}{6} = \frac{7}{2},$$

which is precisely halfway between 3 and 4, as we should expect. Similarly, the expectation of the squares is given by:

$$\mathbb{E}[X_i^2] = \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} = \frac{91}{6}.$$

The variances can be computed similarly:

$$\mathrm{Var}[X_i] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12},$$

and:

$$\mathrm{Var}[X_i^2] = \frac{1}{6} + \frac{32}{6} + \frac{81}{6} + \frac{256}{6} + \frac{625}{6} + \frac{1296}{6} - \left(\frac{91}{6}\right)^2 = \frac{5465}{36}.$$

We can now start the question properly:

(a) The expectation of the sum of the values shown by the dice is:

$$\mathbb{E}[X_1 + X_2 + X_3] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] = 3\mathbb{E}[X_1] = \frac{21}{2}.$$

The variance of the sum of the values shown by the dice is:

$$\mathrm{Var}[X_1 + X_2 + X_3] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + \mathrm{Var}[X_3] = 3\mathrm{Var}[X_1] = \frac{35}{4}$$

(b) The expectation of the sum of the squares of the values shown by the dice is:

$$\mathbb{E}[X_1^2 + X_2^2 + X_3^2] = \mathbb{E}[X_1^2] + \mathbb{E}[X_2^2] + \mathbb{E}[X_3^2] = 3\mathbb{E}[X_1^2] = \frac{91}{2}.$$

The variance of the sum of the squares of the values shown by the dice is:

$$\mathrm{Var}[X_1^2 + X_2^2 + X_3^2] = \mathrm{Var}[X_1^2] + \mathrm{Var}[X_2^2] + \mathrm{Var}[X_3^2] = 3\mathrm{Var}[X_1^2] = \frac{5465}{12}.$$

The geometric distribution

4. Explain what it means to say that a discrete random variable X has a *geometric distribution*, $X \sim \mathrm{Geo}(p)$. Prove that:

$$\mathbb{E}[X] = 1/p, \qquad \mathrm{Var}[X] = (1-p)/p^2.$$

Solution: The random variable X has a *geometric distribution* $X \sim \mathrm{Geo}(p)$ if it takes values $1, 2, 3, \dots$ up to infinity, with probabilities:

$$\mathbb{P}(X = x) = p(1-p)^{x-1}.$$

The expectation is given by:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} && \text{(definition)} \\
&= -p\frac{d}{dp}\sum_{x=1}^{\infty}(1-p)^x && \text{(trying to relate to sum of geometric series)} \\
&= -p\frac{d}{dp}\left(\frac{1-p}{1-(1-p)}\right) && \text{(sum of infinite geometric series)} \\
&= -p\frac{d}{dp}\left(\frac{1}{p} - 1\right) \\
&= \frac{1}{p},
\end{aligned}$$

as required.

The variance is given by:

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} - \frac{1}{p^2} && \text{(definition)} \\
&= -p\frac{d}{dp}\sum_{x=1}^{\infty} x(1-p)^x - \frac{1}{p^2} \\
&= p\frac{d}{dp}\left((1-p)\frac{d}{dp}\sum_{x=1}^{\infty}(1-p)^x\right) - \frac{1}{p^2} && \text{(trying to relate to sum of geometric series)} \\
&= p\frac{d}{dp}\left((1-p)\frac{d}{dp}\left(\frac{1}{p}-1\right)\right) - \frac{1}{p^2} \\
&= p\frac{d}{dp}\left(\frac{1}{p}-\frac{1}{p^2}\right) - \frac{1}{p^2} \\
&= p\left(-\frac{1}{p^2}+\frac{2}{p^3}\right) - \frac{1}{p^2} \\
&= \frac{1-p}{p^2},
\end{aligned}$$

as required.

5. After a late night out in Cambridge, you attempt to open the door to your college room with the three keys in your pocket, only one of which is the correct door key. Assuming that you are equally likely to select any of the keys in your pocket, compute the probability mass functions, the expectations, and variances of:

(a) the random variable X, which is the number of attempts required to open the door if once you try a key from your pocket, you discard it on the ground if it doesn't work;

(b) the random variable Y, which is the number of attempts required to open the door if once you try a key from your pocket, you place it back in your pocket again.

Solution:

(a) In the first case, our chance of getting the key right on the first go is $1/3$. Afterwards, on the second go, the chance is $1/2$. Finally, the chance on the third go is 1. Therefore, the variable X has the mass function:

$$f_X(1) = \frac{1}{3}, \qquad f_X(2) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}, \qquad f_X(3) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}.$$

It follows that the expectation is given by:

$$\mathbb{E}[X] = \frac{1}{3} + \frac{2}{3} + \frac{3}{3} = 2,$$

and the variance is given by:

$$\mathrm{Var}[X] = \frac{1}{3} + \frac{4}{3} + \frac{9}{3} - 4 = \frac{14}{3} - 4 = \frac{2}{3}.$$

(b) In the second case, the chance of being successful on the first attempt is $1/3$. On the second attempt, this remains constant, but we must have also failed on the first attempt. This continues indefinitely. Hence the probability mass function is:

$$f_Y(y) = \frac{1}{3}\left(\frac{2}{3}\right)^{y-1}.$$

Thus we have a geometric random variable, $Y \sim \mathrm{Geo}(1/3)$. It follows that the expectation is 3 and the variance is:

$$\frac{2/3}{(1/3)^2} = 6.$$

The binomial and Poisson distributions

6.

(a) Explain what it means to say that a random variable X has a *binomial distribution*, $X \sim B(n,p)$. Prove that:

$$\mathbb{E}[X] = np, \qquad \mathrm{Var}[X] = np(1-p).$$

(b) An opaque bag contains 10 green counters and 20 red. One counter is selected at random and then replaced: green scores one and red scores zero. Five draws are made. If X is the total score, determine its expectation and variance.

Solution:

(a) A random variable X has a *binomial distribution*, $X \sim B(n,p)$, if its probability mass function is given by:

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The expectation is given by:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} && \text{(definition of expectation)} \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{x-1} && \text{(note } x = 0 \text{ case vanishes)} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{x-1} && \text{(definition of binomial coefficients)} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!((n-1)-(x-1))!} p^{x-1} (1-p)^{(n-1)-(x-1)} && \text{(clever rearranging)} \\
&= np \sum_{r=0}^{n-1} \frac{(n-1)!}{r!((n-1)-r)!} p^r (1-p)^{(n-1)-r} && \text{(substituting } r = x-1\text{)} \\
&= np \left((p + 1 - p)^{n-1} \right) \\
&= np,
\end{aligned}$$

as required.

Similarly, the variance is given by (using some clever rewriting of $x^2 = x(x-1) + x$):

$$
\begin{aligned}
\mathrm{Var}[X] &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} - n^2p^2 \\
&= \sum_{x=1}^{n} \left(x(x-1) + x\right) \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} - n^2p^2 \\
&= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} p^x (1-p)^{n-x} + \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} - n^2p^2 \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!((n-2)-(x-2))!} p^{x-2} (1-p)^{(n-2)-(x-2)} + np - n^2p^2 \\
&= n(n-1)p^2 + np - n^2p^2 \\
&= np(1-p),
\end{aligned}
$$

as required.

(b) The total score X is just the number of green counters drawn, which follows a binomial distribution $B(5, 1/3)$, since the probability of drawing a green counter is constant. The expectation and variance are therefore $5/3$ and $10/3$ respectively.

7.

(a) Explain what it means to say that a random variable X has a *Poisson distribution*, $X \sim \mathrm{Po}(\lambda)$. Prove that:

$$\mathbb{E}[X] = \lambda, \qquad \mathrm{Var}[X] = \lambda.$$

Prove also that the limit of a binomial distribution $B(n, p)$ as $n \to \infty$, $p \to 0$ with $np = \lambda$ fixed, is a Poisson distribution $\mathrm{Po}(\lambda)$.

(b) The probability of seeing a shooting star in any given hour is 0.44. Explaining all assumptions you make, estimate the probabilty of seeing a shooting star in any given half hour.

Solution:

(a) We say that X has a *Poisson distribution*, $X \sim \mathrm{Po}(\lambda)$, if its probability mass function is given by:

$$\mathbb{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!},$$

for $x = 0, 1, 2 \ldots$. The expectation is given by:

$$\begin{aligned}
\mathbb{E}[X] &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{x\lambda^x}{x!} \\
&= e^{-\lambda} \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} && \text{(some clever rearranging)} \\
&= e^{-\lambda} \lambda \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} && \text{(substitute } r = x - 1 \text{ in the sum)} \\
&= \lambda && \text{(Taylor series of } e^{\lambda} \text{ about } \lambda = 0)
\end{aligned}$$

The variance is given by:

$$\begin{aligned}
\mathrm{Var}[X] &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{x^2 \lambda^x}{x!} - \lambda^2 \\
&= e^{-\lambda} \sum_{x=1}^{\infty} \frac{x(x-1)\lambda^x}{x!} + e^{-\lambda} \sum_{x=1}^{\infty} \frac{x\lambda^x}{x!} - \lambda^2 \\
&= e^{-\lambda} \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda.
\end{aligned}$$

Finally, we are also asked to show that the limit of a binomial distribution $B(n, p)$ as $n \to \infty, p \to 0$, with $np = \lambda$ fixed is a Poisson distribution (this result is sometimes called the *law of small numbers*). For any fixed x, we have:

$$\begin{aligned}\mathbb{P}(X = x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{1}{x!} \cdot n(n-1)...(n-(x-1))p^x(1-p)^{n-x} \\ &= \frac{1}{x!}\lambda\left(\lambda - \frac{1}{n}\right)...\left(\lambda - \frac{x-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{1}{n}\right)^{-x}.\end{aligned}$$

We have rewritten everything in terms of n, which we are taking to infinity, and $np = \lambda$, which we are keeping constant. Keeping x fixed, we now take the limit as $n \to \infty$. The result is:

$$\mathbb{P}(X = x) \to \frac{1}{x!}\lambda^x e^{-\lambda} \qquad \text{as } n \to \infty,$$

using the result from Examples Sheet 4 telling us:

$$\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

(b) Assuming that:

- there is a constant average rate of shooting stars throughout the entire evening;
- disjoint time intervals are independent;

we can model the number of shooting stars seen in an hour as a Poisson random variable X with average rate λ per hour. Then the probability of seeing at least one shooting star in an hour is:

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-\lambda}.$$

We are given that this is 0.44, hence $1 - e^{-\lambda} = 0.44$, which gives $\lambda = -\ln(0.56)$.

If the average rate of shooting stars per hour is λ, then the average rate of shooting stars per half hour is $\lambda/2$. Hence modelling the number of shooting stars seen in a half hour as another Poisson random variable Y with average rate $\lambda/2$ per half hour, the probability of seeing at least one shooting star in a half hour is:

$$\mathbb{P}(Y \geq 1) = 1 - \mathbb{P}(Y = 0) = 1 - e^{-\lambda/2} = 1 - e^{\ln(0.56)/2} = 1 - \sqrt{0.56} = 1 - \sqrt{\frac{14}{25}} = 1 - \frac{\sqrt{14}}{5}.$$

Continuous random variables

8. Let X be a continuous random variable with probability density function:

$$f_X(x) = \begin{cases} 10dx^2, & 0 \le x \le \frac{3}{5}, \\ 9d(1-x), & \frac{3}{5} \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find the value of d, and hence sketch the function f_X.

(b) Find and sketch the cumulative distribution function of X. Show also that the derivative of the cumulative distribution function is equal to the probability density function f_X - is this an accident?

(c) Compute the probability $\mathbb{P}(0 \le X \le 2/5)$ by: (i) integrating the probability density function; (ii) taking the difference of two values of the cumulative distribution function.

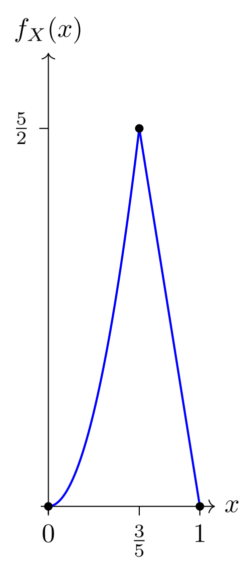
(d) Compute the mode, median, mean, and variance of the variable X.

Solution:

(a) We require that the integral of $f_X(x)$ from 0 to 1 is one for this distribution to be properly normalised. Hence:

$$\begin{aligned} 1 &= 10d \int_0^{3/5} x^2\, dx + 9d \int_{3/5}^{1} (1-x)\, dx \\ &= \frac{10d}{3}\left(\frac{3}{5}\right)^3 + \frac{9d}{2}\left[-(1-x)^2\right]_{3/5}^{1} \\ &= 10d \cdot \frac{9}{125} + \frac{9d}{2}\left(\frac{2}{5}\right)^2 \\ &= \frac{36}{25}d. \end{aligned}$$

Hence $d = 25/36$. The sketch of the function is:



(b) The cumulative distribution function is, by definition:

$$F_X(x) = \mathbb{P}(-\infty \le X) = \int_{-\infty}^{x} f_X(t)\,dt.$$

If $x < 0$, this is zero. If $0 < x < 3/5$, this is given by:

$$F_X(x) = 10d \int_0^x t^2\,dt = \frac{10dx^3}{3} = \frac{125x^3}{54}.$$

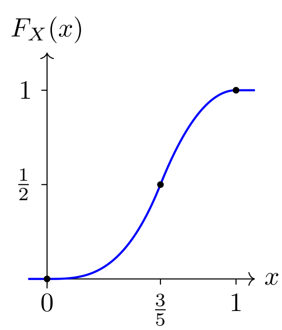
If $3/5 < x < 1$, this is given by:

$$\begin{aligned}
F_X(x) &= 10d \int_0^{3/5} t^2\,dt + 9d \int_{3/5}^{x} (1-t)\,dt \\
&= 10d \left(\frac{3}{5}\right)^3 + 9d \left[-\frac{(1-t)^2}{2}\right]_{3/5}^{x} \\
&= 10 \cdot \frac{26}{36} \cdot \left(\frac{3}{5}\right)^3 + 9 \cdot \frac{25}{36} \cdot \left[\frac{(2/5)^2}{2} - \frac{(1-x)^2}{2}\right] \\
&= \frac{181}{100} - \frac{25}{8}(1-x)^2.
\end{aligned}$$

For $x > 1$, we have $F_X(x) = 1$. Hence the cumulative distribution function is:

$$F_X(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{125x^3}{54}, & 0 \le x \le 3/5, \\ \dfrac{181}{100} - \dfrac{25}{8}(1-x)^2, & 3/5 \le x \le 1, \\ 1, & x \ge 1. \end{cases}$$

The sketch is:



Computing the derivative, we note that:

$$F_X'(x) = \begin{cases} 0, & x \leq 0, \\ \dfrac{125x^2}{18}, & 0 \leq x \leq 3/5, \\ \dfrac{25}{4}(1-x), & 3/5 \leq x \leq 1, \\ 1, & x \geq 1. \end{cases} = f_X(x).$$

It is not an accident that they agree, because:

$$F_X'(x) = \frac{d}{dx} \int\limits_{-\infty}^{x} f_X(t)\, dt = f_X(x),$$

by the fundamental theorem of calculus. So the derivative of the cumulative distribution function is always equal to the probability density function.

(c) The given probability can be computed in two ways:

- Directly integrating the density function, we have:

$$10d \int\limits_0^{2/5} x^2\, dx = \frac{10d}{3}\left(\frac{2}{5}\right)^3 = \frac{250}{108}\left(\frac{2}{5}\right)^3 = \frac{4}{27}.$$

- Alternatively, using the cumulative density function, we have:

$$F_X(2/5) - F_X(0) = \frac{125}{54}\left(\frac{2}{5}\right)^3 = \frac{4}{27},$$

  which is evidently the same.

(d) Finally, we are asked to calculate various statistical estimators for the distribution. The mode is clearly $3/5$ from the sketch. The median is the value for which $F_X(x) = 1/2$, which also occurs at $x = 3/5$. The mean can be calculated via:

$$\begin{aligned} \mathbb{E}[X] &= \int\limits_0^1 x f_X(x)\, dx \\ &= 10d \int\limits_0^{3/5} x^3\, dx + 9d \int\limits_{3/5}^{1} (x - x^2)\, dx \\ &= \frac{10d}{4}\left(\frac{3}{5}\right)^4 + 9d\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_{3/5}^{1} \\ &= \frac{10d}{4}\left(\frac{3}{5}\right)^4 + 9d\left(\frac{1}{2} - \frac{1}{3} - \frac{1}{2}\left(\frac{3}{5}\right)^2 + \frac{1}{3}\left(\frac{3}{5}\right)^3\right) \\ &= \frac{213d}{250} \\ &= \frac{71}{120}. \end{aligned}$$

To calculate the variance, we first calculate:

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_0^1 x^2 f_X(x)\,dx \\
&= 10d \int_0^{3/5} x^4\,dx + 9d \int_{3/5}^1 (x^2 - x^3)\,dx \\
&= \frac{10d}{5}\left(\frac{3}{5}\right)^5 + 9d\left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_{3/5}^1 \\
&= 2d\left(\frac{3}{5}\right)^5 + 9d\left(\frac{1}{3} - \frac{1}{4} - \frac{1}{3}\left(\frac{3}{5}\right)^3 + \frac{1}{4}\left(\frac{3}{5}\right)^4\right) \\
&= \frac{1716d}{3135} \\
&= \frac{143}{375}.
\end{aligned}$$

Hence, the variance is given by:

$$\mathrm{Var}[X] = \frac{143}{375} - \left(\frac{71}{120}\right)^2 = \frac{2251}{72000}.$$

9. A continuous random variable X, taking values in the interval $[0, 1]$, has cumulative distribution function:

$$F_X(x) = A\left(\frac{x^3}{3} - \frac{x^4}{4}\right),$$

for values in the range $0 \leq x \leq 1$.

(a) Find the value of the constant A, and hence sketch the function F_X for all values of x.

(b) Find and sketch the probability density function of X.

(c) Calculate the mode, median, mean, and variance of the random variable X.

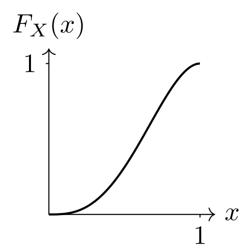
(d) Calculate the probability that X lies in the interval $[\mathbb{E}[X] - \sigma, \mathbb{E}[X] + \sigma]$, where σ is the standard deviation of the random variable X.

Solution:

(a) Since X can only take values in the interval $[0, 1]$, we must have $F_X(1) = 1$. Hence we require:

$$A\left(\frac{1}{3} - \frac{1}{4}\right) = \frac{A}{12} = 1 \qquad \Rightarrow \qquad A = 12.$$

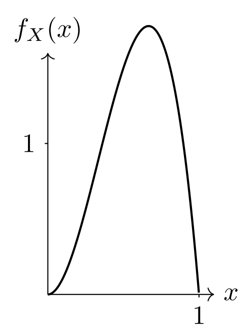
The sketch is:



(b) The probability density function is:

$$f_X(x) = F_X'(x) = A(x^2 - x^3),$$

in the region $[0, 1]$ and 0 elsewhere. Sketching this function, we have:



(c) The mode is the maximum value of the density function; to obtain this, we could use differentiation. We have $f_X'(x) = A(2x - 3x^2) = 0$ when $x = 0$ or $x = 2/3$. Evidently $x = 0$ is a minimum and $x = 2/3$ is a maximum, from the sketch, which gives $2/3$ as the mode.

The median is the value for which $F_X(x) = 1/2$. This occurs when:

$$A\left(\frac{x^3}{3} - \frac{x^4}{4}\right) = \frac{1}{2} \qquad \Rightarrow \qquad 8x^3 - 6x^4 = 1,$$

so the median is the solution of this quartic equation in $[0, 1]$ (which we cannot solve via elementary methods sadly!). The mean is given by:

$$\mathbb{E}[X] = \int_0^1 A(x^3 - x^4)\,dx = A\left(\frac{1}{4} - \frac{1}{5}\right) = A\left(\frac{1}{20}\right) = \frac{3}{5}.$$

To compute the variance, we calculate:

$$\mathbb{E}[X^2] = \int_0^1 A(x^4 - x^5)\,dx = A\left(\frac{1}{5} - \frac{1}{6}\right) = A\left(\frac{1}{30}\right) = \frac{2}{5}.$$

This gives:

$$\mathrm{Var}[X] = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{10}{25} - \frac{9}{25} = \frac{1}{25}.$$

(d) The required probability is the probability of being in the interval $[3/5 - 1/5, 3/5 + 1/5] = [2/5, 4/5]$, given by:

$$F_X(4/5) - F_X(2/5) = 12\left(\frac{1}{3}\left(\frac{4}{5}\right)^3 - \frac{1}{4}\left(\frac{4}{5}\right)^4\right) - 12\left(\frac{1}{3}\left(\frac{2}{5}\right)^3 - \frac{1}{4}\left(\frac{2}{5}\right)^4\right) = \frac{16}{25}.$$

10. A continuous random variable X, taking values in the interval $[-2, 2]$, has probability density function:

$$f_X(x) = \frac{1 + e^{-|x|}}{N},$$

for values in the range $-2 \leq x \leq 2$.

(a) Find the value of the constant N, and sketch the function f_X for all values of x.

(b) Find and sketch the cumulative distribution function of X.

(c) Calculate the mode, median, mean, and variance of the random variable X.

Another continuous random variable Y, taking values in the interval $[-2, 2]$, has probability density function:

$$f_Y(y) = \begin{cases} 0, & y < -2, \\ \mathbb{P}(X \leq y)/M, & -2 \leq y \leq 2, \\ 0, & y > 2. \end{cases}$$

(d) Calculate the constant M.

➻ Solution:

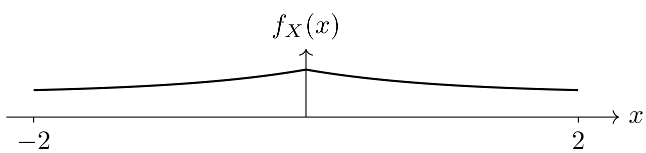
(a) Notice that the density function is even. Hence we require:

$$1 = \frac{2}{N} \int_0^2 (1 + e^{-x})\, dx = \frac{2}{N} \left[x - e^{-x}\right]_0^2 = \frac{2}{N} \left(3 - \frac{1}{e^2}\right),$$

giving:

$$N = 2\left(3 - \frac{1}{e^2}\right).$$

The sketch is:



(b) The cumulative distribution function F_X satisfies $F_X(x) = 0$ for $x < -2$. For $-2 < x < 0$, we have:

$$F_X(x) = \frac{1}{N} \int_{-2}^{x} (1 + e^{x})\, dx = \frac{1}{N} \left[x + e^{x}\right]_{-2}^{x} = \frac{1}{N} \left(2 - \frac{1}{e^2} + x + e^{x}\right).$$

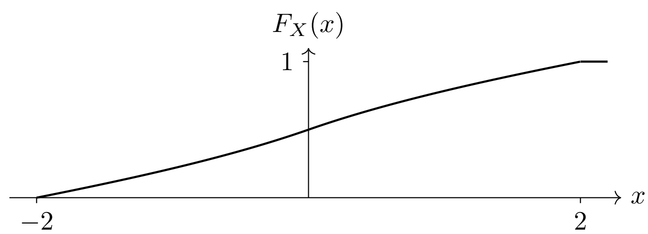
For $0 < x < 2$, we have:

$$\begin{aligned} F_X(x) &= \frac{1}{N}\left(3 - \frac{1}{e^2}\right) + \frac{1}{N}\int_0^x (1 + e^{-x})\,dx \\ &= \frac{1}{N}\left(3 - \frac{1}{e^2}\right) + \frac{1}{N}\left[x - e^{-x}\right]_0^x \\ &= \frac{1}{N}\left(3 - \frac{1}{e^2}\right) + \frac{1}{N}\left(x - e^{-x} + 1\right) \\ &= \frac{1}{N}\left(4 - \frac{1}{e^2} + x - e^{-x}\right). \end{aligned}$$

For $x > 2$, we have $F_X(x) = 1$. Thus the cumulative distribution function is:

$$F_X(x) = \frac{1}{N}\begin{cases} 0, & x \le 0, \\ 2 - \dfrac{1}{e^2} + x + e^x, & -2 \le x \le 0, \\ 4 - \dfrac{1}{e^2} + x - e^{-x}, & 0 \le x \le 2, \\ 1, & x \ge 2. \end{cases}$$

The required sketch is:



(c) The distribution is completely symmetric about $x = 0$. Thus the mode, median and mean are all 0. Since $\mathbb{E}[X] = 0$, the variance is given by:

$$
\begin{aligned}
\mathbb{E}[X^2] &= \frac{2}{N}\int_0^2 x^2(1+e^{-x})\,dx \\
&= \frac{2}{N}\cdot\frac{8}{3} + \frac{2}{N}\int_0^2 x^2 e^{-x}\,dx \\
&= \frac{16N}{3} + \frac{2}{N}\left(\left[-x^2e^{-x}\right]_0^2 + 2\int_0^2 xe^{-x}\,dx\right) \\
&= \frac{16N}{3} + \frac{2}{N}\left(-4e^{-2} + 2\left[-xe^{-x}\right]_0^2 + 2\int_0^2 e^{-x}\,dx\right) \\
&= \frac{16N}{3} + \frac{2}{N}\left(-\frac{8}{e^2} + 2\left[-e^{-x}\right]_0^2\right) \\
&= \frac{16N}{3} - \frac{16N}{e^2} + \frac{4}{N}\left(-\frac{1}{e^2} + 1\right) \\
&= \frac{4}{N}\left(\frac{7}{3} - \frac{5}{e^2}\right).
\end{aligned}
$$

(d) We require:

$$1 = \frac{1}{M}\int_{-2}^2 \mathbb{P}(X \le y)\,dy = \frac{1}{M}\int_{-2}^2 F_X(y)\,dy.$$

We can make this slightly faster to evaluate by integrating by parts. We have:

$$\int_{-2}^2 F_X(y)\,dy = \left[yF_X(y)\right]_{-2}^2 - \int_{-2}^2 yf_X(y)\,dy = 2F_X(2) + 2F_X(-2) - N\mathbb{E}[X].$$

This gives:

$$M = 2F_X(2) + 2F_X(-2) - N\mathbb{E}[X] = 2.$$

The normal distribution and the central limit theorem

11. Explain what it means to say that a random variable X is *normally distributed*, $X \sim N(\mu, \sigma^2)$. Prove that:

$$\mathbb{E}[X] = \mu, \qquad \mathrm{Var}[X] = \sigma^2.$$

➾ **Solution:** A random variable X is *normally distributed* with mean μ and variance σ^2 if it has the distribution function:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

To prove the expectation is indeed μ, we compute:

$$\begin{aligned}\int_{-\infty}^{\infty} x f_X(x)\,dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,dx + \frac{\mu}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,dx.\end{aligned}$$

Since the distribution is normalised, the second term must equal μ. The first term can be shifted by the substitution $x' = x - \mu$, resulting in an odd integrand, so vanishes. Hence the mean is indeed μ.

The variance can be obtained by considering the integral:

$$\begin{aligned}\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \left(\left[-\sigma^2(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right]_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\,dx \right) \quad \text{(by parts)} \\ &= \sigma^2,\end{aligned}$$

assuming the distribution is properly normalised. The variance is indeed σ^2, as required.

12.

(a) State the *central limit theorem*.

(b) Using the central limit theorem, show that the limit of a binomial distribution $B(n,p)$ as $n \to \infty$, with $\mu = np, \sigma = np(1-p)$ fixed, is a normal distribution $N(\mu, \sigma^2)$. This is called the *normal approximation* to a binomial distribution.

(c) Using the normal approximation, estimate the probability that a binomial random variable $X \sim B(100, 1/5)$ lies within one standard deviation of its mean. [*You will need to look up some values of the cumulative distribution function of the normal distribution online.*]

Solution:

(a) The central limit theorem states that if $X_1, X_2, ..., X_n$ are a collection of independent, identically distributed random variables drawn from a distribution with mean μ and variance σ^2, then the distribution of the 'sample mean':

$$\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n}$$

approaches $N(\mu, \sigma^2/n)$ as $n \to \infty$.

(b) Let $X_1, X_2, ..., X_n \sim B(1,p)$ be independent binomial random variables, each for single random trials. Then:

$$X_1 + X_2 + ... + X_n \sim B(n,p),$$

since the trials are independent. By the central limit theorem though, we have:

$$\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n} \sim N(p, p(1-p)/n),$$

as $n \to \infty$, using the mean and variance of a binomial random variable. Scaling a normal random variable results in another normal random variable, except with shifted mean and variance, giving:

$$n\bar{X} = X_1 + X_2 + ... + X_n \sim N(np, np(1-p)).$$

Hence, the normal distribution to the binomial for large n is $B(n,p) \approx N(np, np(1-p))$.

(c) The mean of the given distribution is 20 and the variance is 16. Hence the normal approximation to the distribution is $X \sim N(20, 16)$. We want to compute the probability of being one standard deviation away from the mean:

$$\mathbb{P}(16 \leq X \leq 24) = \mathbb{P}\left(-1 \leq \frac{X-20}{4} \leq 1\right) = \Phi(1) - \Phi(-1),$$

where Φ is the cumulative distribution function of a standard normal $N(0,1)$, since $(X-20)/4 \sim N(0,1)$. By symmetry, $\Phi(1) = 1 - \Phi(-1)$, hence we can rewrite the whole probability as $1 - 2\Phi(-1)$. Using a standard normal table (e.g. looking at `https://en.wikipedia.org/wiki/Standard_normal_table`), we see that $\Phi(-1) = 0.15866$, hence the probability is $1 - 2(0.15866) = 0.68268$.

Functions of random variables

13. Let X be a continuous random variable with probability density function f_X. Let $U = X + c$ be a new random variable, where c is a constant. By considering $\mathbb{P}(u \leq U \leq u + du)$, for small du, find the probability density function for U in terms of f_X.

➻ **Solution:** By definition, we have $\mathbb{P}(u \leq U \leq u + du) = f_U(u)du$, for infinitesimally small du. But also, we have:

$$\mathbb{P}(u \leq U \leq u + du) = \mathbb{P}(u \leq X + c \leq u + du) = \mathbb{P}(u - c \leq X \leq u - c + du) = f_X(u - c)du,$$

for infinitesimal du, where f_X is the probability density function of X. Hence $f_U(u) = f_X(u - c)$.

---

14. Let X be a continuous random variable, let f_X be its probability density, and let F_X be its cumulative distribution function. Suppose that $Y = g(X)$ is another random variable, where g is an arbitrary function. Show that if g is invertible, differentiable, and increasing, then the probability density function of Y is given by:

$$f_Y(y) = \frac{dg^{-1}}{dy}(y) \cdot f_X(g^{-1}(y)).$$

[*Hint: consider the equality of cumulative distribution functions* $\mathbb{P}(y \leq Y) = \mathbb{P}(y \leq g(X)) = \mathbb{P}(g^{-1}(y) \leq X)$.] How does this formula change if g is invertible, differentiable, and *decreasing*?

➻ **Solution:** As suggested in the hint, we consider:

$$\mathbb{P}(y \leq Y) = \mathbb{P}(y \leq g(X)) = \mathbb{P}(g^{-1}(y) \leq X),$$

which holds since g is invertible and g is increasing, hence its inverse is also increasing (remember it is a reflection in the line $y = x$!). Writing out both sides explicitly, we have:

$$\int\limits_{y}^{\infty} f_Y(y')\, dy' = \int\limits_{g^{-1}(y)}^{\infty} f_X(x)\, dx.$$

Now differentiate both sides with respect to y, and use the fundamental theorem of calculus (coupled with the chain rule in the case of the right hand side!). We have:

$$f_Y(y) = \frac{dg^{-1}(y)}{dy} \cdot f_X(g^{-1}(y)),$$

as required.

In the case that g is invertible, differentiable and *decreasing*, then the inequality flips in the first line:

$$\mathbb{P}(y \leq Y) = \mathbb{P}(y \leq g(X)) = \mathbb{P}(g^{-1}(y) \geq X).$$

Hence, there is a relative sign difference throughout the rest of the question. As a result, we have:

$$f_Y(y) = -\frac{dg^{-1}(y)}{dy} \cdot f_X(g^{-1}(y)).$$

Hence if g is either increasing or decreasing, and both invertible and differentiable, then:

$$f_Y(y) = \left|\frac{dg^{-1}(y)}{dy}\right| f_X(g^{-1}(y)).$$

15. A cannon fixed at the origin in the xy plane fires cannonballs at a screen placed $x = x_0$. The cannonballs are inclined at a random angle Θ to the x-axis, where Θ is uniformly distributed on the interval $[-\pi/2, \pi/2]$. Let Y be the random y-coordinate of the point of collision with the screen.

(a) Using the result of the previous question, show that the probability density function of Y is given by:

$$f_Y(y) = \frac{x_0}{\pi(x_0^2 + y^2)}.$$

This distribution is called a *Cauchy distribution*.

(b) Hence, show that the mean and standard deviation of Y are not well-defined.

➾ Solution:

(a) Since Θ is uniformly distributed on $[-\pi/2, \pi/2]$, its density is $f_\Theta(\theta) = 1/\pi$ on that interval, and zero elsewhere. The y-coordinate is $Y = x_0 \tan(\Theta)$, which is an increasing, invertible, differentiable function on $[-\pi/2, \pi/2]$. As a result, we have:

$$f_Y(y) = \left(\frac{d}{dy}\arctan\left(\frac{y}{x_0}\right)\right) \cdot f_\Theta\left(\arctan\left(\frac{y}{x_0}\right)\right) = \frac{1/x_0}{\pi(1 + (y/x_0)^2)} = \frac{x_0}{\pi(y^2 + x_0^2)},$$

as required.

(b) The mean of this distribution is:

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} \frac{x_0 y}{\pi(y^2 + x_0^2)}\, dy.$$

This is an improper integral in both infinite directions, so must be regularised as:

$$\lim_{s,t\to\infty} \int_{-s}^{t} \frac{x_0 y}{\pi(y^2 + x_0^2)}\, dy,$$

where the limits are taken independently. But this results in a divergent logarithmic result in both cases, which implies the mean does not exist.

The standard deviation is defined to be the square root of the variance. But the variance is not defined in this case either for the same reason, that:

$$\int_{-\infty}^{\infty} \frac{x_0 y^2}{\pi(y^2 + x_0^2)}\, dy$$

is a divergent integral.

Functions of multiple random variables

16.

(a) If X, Y are independent discrete random variables taking integer values, explain why the distribution of their sum $Z = X + Y$ is given by the convolutional sum:

$$\mathbb{P}(Z = z) = \sum_{k=-\infty}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y = z - k).$$

(b) Hence, show that if $X \sim \mathrm{Po}(\lambda), Y \sim \mathrm{Po}(\mu)$ are independent Poisson variables, their sum is also a Poisson variable, $X + Y \sim \mathrm{Po}(\lambda + \mu)$.

(c) The number of patients arriving at Addenbrooke's hospital in a morning is on average 42, whilst the number of patients arriving at Addenbrooke's hospital in the afternoon is on average 64. Estimate the probability that the number of patients arriving on a single day will exceed 100, explaining any modelling assumptions you require. What is the variance of the number of patients arriving on a single day, under your model?

Solution:

(a) Observe that:

$$\mathbb{P}(Z = z) = \sum_{k=-\infty}^{\infty} \mathbb{P}(Y = z - k | X = k)\mathbb{P}(X = k) = \sum_{k=-\infty}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y = z - k),$$

assuming that X, Y are independent.

(b) If $X \sim \mathrm{Po}(\lambda), Y \sim \mathrm{Po}(\mu)$, we have:

$$\begin{aligned}
\mathbb{P}(X + Y = z) &= \sum_{k=-\infty}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y = z - k) \\
&= \sum_{k=0}^{z} \frac{\lambda^k e^{-\lambda}}{k!} \frac{\mu^{z-k} e^{-\mu}}{(z-k)!} \\
&= \frac{e^{-(\lambda+\mu)}}{z!} \sum_{k=0}^{z} \frac{z!}{k!(z-k)!} \lambda^k \mu^{z-k} \\
&= \frac{(\lambda+\mu)^z e^{-(\lambda+\mu)}}{z!}.
\end{aligned}$$

Hence $X + Y \sim \mathrm{Po}(\lambda + \mu)$, as required.

(c) Let X be the number of patients arriving in the morning, and let Y be the number of patients arriving in the afternoon. Then $X \sim \mathrm{Po}(42)$ and $Y \sim \mathrm{Po}(64)$. The total number of patients arriving during the day is $X + Y \sim \mathrm{Po}(106)$. Hence the variance in the number of daily patients is 106. We needed to assume that:

- The average rate is constant in the morning, and the average rate is constant in the afternoon.
- The number of patients arriving in the morning does not affect the number of patients arriving in the afternoon, and vice-versa, i.e. they are independent random variables.

The probability that the total number of patients will exceed 100 is:

$$\mathbb{P}(X+Y>100)=1-\mathbb{P}(X+Y\leq 100)=1-e^{-106}\sum_{k=0}^{100}\frac{106^k}{k!},$$

which cannot be further simplified (it can be calculated numerically, though).

---

17. By considering the probability $\mathbb{P}(X_1+X_2=r)$, where $X_1\sim B(m,p)$, $X_2\sim B(n,p)$ are independent binomial variables, prove *Vandermonde's identity* for the binomial coefficients:

$$\binom{m+n}{r}=\sum_{k=0}^{r}\binom{m}{k}\binom{n}{r-k}.$$

Solution: We have:

$$\begin{aligned}\mathbb{P}(X_1+X_2=r)&=\sum_{k=-\infty}^{\infty}\mathbb{P}(X_1=k)\mathbb{P}(X_2=r-k)\\&=\sum_{k=0}^{r}\binom{m}{k}p^k(1-p)^{m-k}\binom{n}{r-k}p^{r-k}(1-p)^{n-(r-k)}\\&=p^r(1-p)^{m+n-r}\sum_{k=0}^{r}\binom{m}{k}\binom{n}{r-k}.\end{aligned}$$

However, since X_1 is the number of successful trials in m, with success rate p, and X_2 is the number of successful trials in n, with the *same* success rate p, it must be that X_1+X_2 is the number of successful trials in $m+n$ with success rate p. Hence X_1+X_2 is also binomially distributed, as $B(m+n,p)$. Comparing the expected formula with the one above, we deduce Vandermonde's identity:

$$\binom{m+n}{r}=\sum_{k=0}^{r}\binom{m}{k}\binom{n}{r-k}.$$

18.

(a) If X, Y are independent continuous random variables with densities $f_X(x)$, $f_Y(y)$ respectively, explain why the distribution of their sum $Z = X + Y$ has density $f_Z(z)$ given by the convolutional integral:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx.$$

(b) Hence, show that if $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent normal variables, their sum is also a normal variable, $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Solution:

(a) We can either argue using cumulative distribution functions, or using infinitesimals, mirroring the approaches of Questions 16 and 17. Using cumulative distribution functions, we have:

$$\mathbb{P}(Z \leq z) = \mathbb{P}(X + Y \leq z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{(X,Y)}(X = x \cap Y = y)\, dydx,$$

where $f_{(X,Y)}$ is the joint density function. For independent random variables, this will be equal to the product of the density functions, hence we can simplify this to:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y)\, dydx.$$

Taking the derivative with respect to z on both sides, and using the fundamental theorem of calculus, we have:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx,$$

as required.

Alternatively, using infinitesimals, observe that:

$$\begin{aligned}
f_Z(z)dz = d\mathbb{P}(z \leq Z \leq z + dz) &= d\mathbb{P}(z \leq X + Y \leq z + dz) \\
&= \int_{x=-\infty}^{x=\infty} d\mathbb{P}(z \leq x + Y \leq z + dz | X = x) d\mathbb{P}(x \leq X \leq x + dx) \\
&= \int_{x=-\infty}^{x=\infty} d\mathbb{P}(z - x \leq Y \leq z - x + dz) d\mathbb{P}(x \leq X \leq x + dx) \qquad \text{(independence)} \\
&= \left(\int_{x=-\infty}^{x=\infty} f_Y(z-x) f_X(x)\, dx \right) dz,
\end{aligned}$$

which on cancelling the dz from both sides gives the required result.

(b) If X, Y are independent normal variables, we have:

$$f_Z(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \exp\left(-\frac{(z-x-\mu_2)^2}{2\sigma_2^2}\right)\, dx.$$

Instead of doing the algebra directly, we can be a bit sneaky and avoid some. Let $u = x - \mu_1$, and let $z' = z - \mu_1 - \mu_2$. Then the integral becomes:

$$f_Z(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2\sigma_1^2}\right) \exp\left(-\frac{(u-z')^2}{2\sigma_2^2}\right)\, du.$$

Looking at the exponent, we note that:

$$\begin{aligned}
\frac{u^2}{\sigma_1^2} + \frac{(u-z')^2}{\sigma_2^2} &= \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) u^2 - \frac{2u}{\sigma_2^2} z' + \frac{(z')^2}{\sigma_2^2} \\
&= \left(\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\sigma_2^2}\right) \left(u - \left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right)\frac{z'}{\sigma_2^2}\right)^2 - \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\frac{(z')^2}{\sigma_2^4} + \frac{(z')^2}{\sigma_2^2} \\
&= \left(\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\sigma_2^2}\right) \left(u - \frac{\sigma_1^2 z'}{\sigma_1^2+\sigma_2^2}\right)^2 + \frac{(z')^2}{\sigma_1^2+\sigma_2^2}.
\end{aligned}$$

Inserting this back into the integral, we have:

$$\begin{aligned}
f_Z(z) &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\right)\left(u - \frac{\sigma_1^2 z'}{\sigma_1^2+\sigma_2^2}\right)^2 - \frac{(z')^2}{2(\sigma_1^2+\sigma_2^2)}\right] du \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\right)\left(u - \frac{\sigma_1^2 z'}{\sigma_1^2+\sigma_2^2}\right)^2\right] du.
\end{aligned}$$

We can shift the integral again by setting $u' = u - \sigma_1^2 z'/(\sigma_1^2+\sigma_2^2)$, which shows that the integral is independent of z'. This means that:

$$f_Z(z) \propto \exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right),$$

so indeed Z is Gaussian with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$.

19. Let X be a continuous random variable with probability density function $f(x) = \pi^{-1}(1+x^2)^{-1}$, for $-\infty < x < \infty$, and let Y be a continuous random variable with uniform density on the interval $[-1/2, 1/2]$. Let $Z = X + Y$.

(a) Find the probability density function of Z, and sketch it.

(b) Find the mode and mean values of Z.

Solution: Observe that the density function of Y is:

$$f_Y(y) = 1,$$

if $y \in [-1/2, 1/2]$ and 0 otherwise.

(a) Using the density function of Y, we have the density function of Z:

$$f_Z(z) = \frac{1}{\pi}\int_{-1/2}^{1/2} \frac{1}{1+(z-y)^2}\,dy = \frac{1}{\pi}\left[\arctan(y-z)\right]_{-1/2}^{1/2} = \frac{1}{\pi}\left[\arctan\left(\frac{1}{2}-z\right) - \arctan\left(-\frac{1}{2}-z\right)\right].$$

We can simplify this using the addition of arctangents formula:

$$\arctan(A) + \arctan(B) = \arctan\left(\frac{A+B}{1-AB}\right),$$

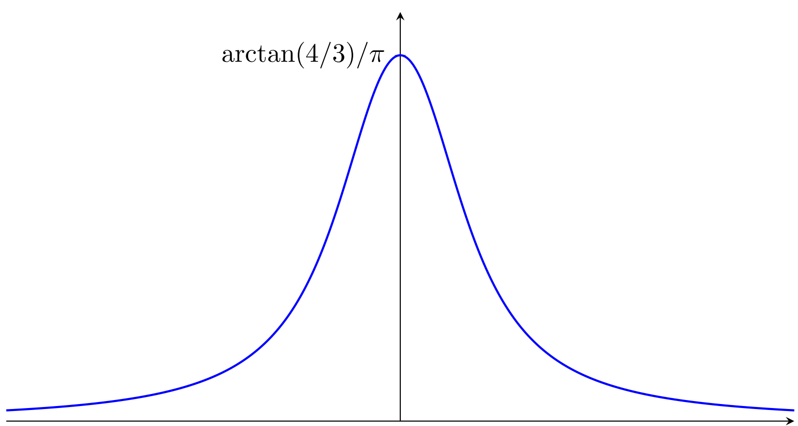
from earlier in the course. The result is:

$$\frac{1}{\pi}\left(\arctan\left(\frac{1}{2}-z\right) + \arctan\left(\frac{1}{2}+z\right)\right) = \frac{1}{\pi}\arctan\left(\frac{1}{z^2+3/4}\right).$$

To plot this function, first observe that as $z \to \pm\infty$, the function tends to zero. Also observe that the argument of the arctangent is at most $4/3$, occurring when $z = 0$, which must be a global maximum for the function. Finally, it is useful to consider its derivative. We have:

$$f_Z'(z) = \frac{1}{\pi}\left(-\frac{2z}{(z^2+3/4)^2}\right)\cdot\frac{1}{1+(z^2+3/4)^{-2}} = -\frac{2z}{\pi((z^2+3/4)^2+1)}.$$

We see that there is a single stationary points when $z = 0$, and that the gradient is positive for $z < 0$, and negative for $z > 0$. This allows us to produce the plot:



(b) The mode is obviously $z = 0$. The mean, on the other hand, is not well-defined. The reason is that for large values of z, we have $\arctan(1/(z^2+4/3)) \approx \arctan(1/z^2) \approx 1/z^2$ (using the Taylor series for $\arctan(x)$, which one can easily develop using the techniques we saw earlier in the course). As a result, when we additionally multiple by z and integrate we have the integral of $1/z$, which is logarithmic, and hence divergent in the infinite limit. This is consistent with what we found for the Cauchy distribution in Question 18 earlier on in the sheet.

20. A pedestrian arrives at a crossing, where they press the button and wait for the lights to change. The time taken before the lights change colour is a uniform random variable T_1, taking values between 0 and 1 minutes.

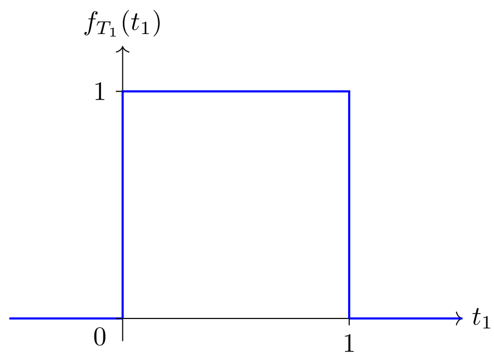
(a) Sketch the probability density function of T_1, and find its mean and variance.

After the lights change, the pedestrian crosses and walks to another independent set of lights, which operate under the same conditions, changing at a time T_2 after the button is pressed. The pedestrian takes 5 minutes in total to walk to the second set of lights, after the first set of lights has changed. Let U denote the difference between the total time taken for the journey, from the moment that the first button is pressed until the second set of lights changes, and 5 minutes.

(b) Find an expression for U in terms of T_1, T_2, and hence obtain its probability density.

(c) By direct integration of the probability density, find the mean and variance of U, and show that this is consistent with the general properties of expectation and variance from Questions 3 and 4.

Solution: (a) The density of T_1 is just $f_{T_1}(t_1) = 1$ if $t_1 \in [0, 1]$, and 0 otherwise. The plot is given below.



The mean is:

$$\mathbb{E}[T_1] = \int_0^1 t_1 \, dt_1 = \frac{1}{2},$$

and the variance is:

$$\mathrm{Var}[T_1] = \int_0^1 t_1^2 \, dt_1 - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

(b) The journey is as follows. The pedestrian presses the button on the first set of lights, which starts the timer running. A time T_1 later, the lights change, and the pedestrian crosses. Then, five minutes elapse, before the pedestrian reaches the second set of lights and presses the button. A time T_2 later, the lights change, and the journey finishes. Thus, the total time taken is $T_1 + 5 + T_2$. Therefore, the variable U is given by $U = T_1 + T_2$.

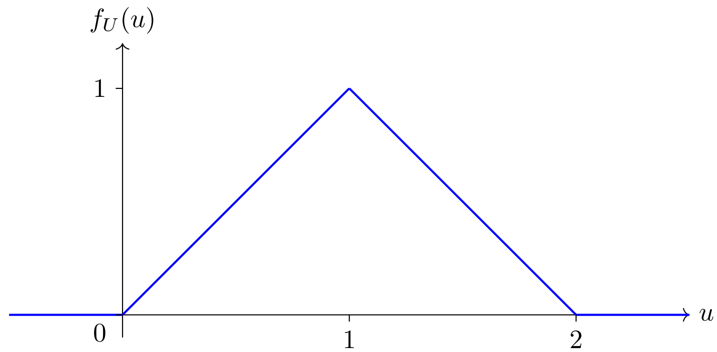
The probability density of U is given by the convolutional integral:

$$f_U(u) = \int_{-\infty}^{\infty} f_{T_1}(t) f_{T_2}(u - t) \, dt = \int_0^1 f_{T_2}(u - t) \, dt.$$

Make the substitution $v = u - t$, so that $dv = -dt$ and the limits transform as $[0,1] \mapsto [u, u-1]$. Then we have:

$$f_U(u) = \int_{u-1}^{u} f_{T_2}(v)\, dv = \begin{cases} 0, & \text{if } u < 0, \\ u, & \text{if } 0 < u < 1, \\ 2-u, & \text{if } 1 < u < 2, \\ 0, & \text{if } u > 2. \end{cases}$$

Sketching this density, we get a little triangular hat:



(c) The mean of U is:

$$\mathbb{E}[U] = \int_0^1 u^2\, du + \int_1^2 (2u - u^2)\, du = \frac{1}{3} + \left[u^2 - \frac{1}{3}u^3\right]_1^2 = \frac{1}{3} + 4 - \frac{8}{3} - 1 + \frac{1}{3} = 1.$$

This is consistent with $\mathbb{E}[T_1] + \mathbb{E}[T_2] = \frac{1}{2} + \frac{1}{2} = 1$.

Meanwhile, the variance of U is:

$$\begin{aligned} \mathrm{Var}[U] &= \int_0^1 u^3\, du + \int_1^2 (2u^2 - u^3)\, du - 1 \\ &= \frac{1}{4} + \left[\frac{2}{3}u^3 - \frac{1}{4}u^4\right]_1^2 - 1 \\ &= \frac{1}{4} + \frac{16}{3} - 4 - \frac{2}{3} + \frac{1}{4} - 1 \\ &= \frac{1}{6}. \end{aligned}$$

This is consistent with $\mathrm{Var}[T_1] + \mathrm{Var}[T_2] = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}$.

21. A coffee machine produces random amounts of liquid for each cup, always in the range 270 to 330ml, with the distribution function:

$$f(v) = \frac{\pi}{120}\cos\left(\frac{\pi(v - 300\,\text{ml})}{60\,\text{ml}}\right)(\text{ml})^{-1}$$

Each coffee is made independently of all other coffees.

(a) Sketch the distribution function, and find both the mean amount of liquid per cup and the variance in the amount of liquid per cup.

(b) Suppose that the machine is used to make two coffees. Calculate the probability that the total amount of liquid produced by the machine for these two cups exceeds 630ml.

(c) Find the mean and standard deviation in the amount of liquid produced by the machine when making two coffees by: (i) using standard properties of expectations and variances of independent random variables; (ii) direct calculation of the expectation and variance from the distribution function of the amount of liquid produced by the machine when making two coffees.

Solution: First, observe that we can make the distribution look a bit nicer by writing:

$$\cos\left(\frac{\pi(v-300)}{60}\right) = \cos\left(\frac{\pi(v-270)}{60} - \frac{30\pi}{60}\right) = \cos\left(\frac{\pi(v-270)}{60} - \frac{\pi}{2}\right) = \sin\left(\frac{\pi(v-270)}{60}\right).$$

(a) From the second form of the distribution, we can easily construct a sketch, which is simply a sine graph starting at $v = 270$ ml, giving an argument 0 for the sine, and ending at $v = 330$ ml, giving an argument π for the sine.

The mean amount of liquid produced by the machine is obviously 300 ml by symmetry of the distribution about this point. If we wanted to check explicitly with an integral, we have:

$$\begin{aligned}
\frac{\pi}{120}\int_{270}^{330} v\sin\left(\frac{\pi(v-270)}{60}\right)dv &= \frac{1}{2}\int_0^{\pi}\left(\frac{60u}{\pi} + 270\right)\sin(u)\,du && \text{(substituting } u = \pi(v-270)/60\text{)} \\
&= \left[-\left(\frac{30u}{\pi} + 135\right)\cos(u)\right]_0^{\pi} + \frac{30}{\pi}\int_0^{\pi}\cos(u)\,du && \text{(by parts)} \\
&= (30 + 135) + 135 \\
&= 300\text{ ml}
\end{aligned}$$

The variance in the amount of liquid produced by the machine can be computed by subtracting the square of the mean from the quantity:

$$\begin{aligned}
\frac{\pi}{120}\int_{270}^{330} v^2\sin\left(\frac{\pi(v-270)}{60}\right)dv &= \frac{1}{2}\int_0^{\pi}\left(\frac{60u}{\pi} + 270\right)^2\sin(u)\,du \\
&= \frac{1}{2}\left(\left[-\left(\frac{60u}{\pi} + 270\right)^2\cos(u)\right]_0^{\pi} + \frac{120}{\pi}\int_0^{\pi}\left(\frac{60u}{\pi} + 270\right)\cos(u)\,du\right) \\
&= \frac{1}{2}\left(330^2 + 270^2 + \frac{120}{\pi}\left[\left(\frac{60u}{\pi} + 270\right)\sin(u)\right]_0^{\pi} - \frac{120\cdot 60}{\pi^2}\int_0^{\pi}\sin(u)\,du\right) \\
&= \frac{1}{2}\left(330^2 + 270^2 - \frac{240\cdot 60}{\pi^2}\right).
\end{aligned}$$

Cleaning up the constants, we end up with the variance:

$$\frac{1}{2}\left(330^2 + 270^2 - \frac{240 \cdot 60}{\pi^2}\right) - 300^2 = 900\left(1 - \frac{8}{\pi^2}\right) \text{ ml}^2.$$

(b) Let C_1 be the amount of coffee produced for the first cup, and C_2 be the amount of coffee produced by the second cup. We want the probability $C_1 + C_2 > 630$ ml. Both C_1, C_2 have density functions:

$$f_{C_i}(v) = \frac{\pi}{120}\sin\left(\frac{\pi(v-270)}{60}\right) \text{ (ml)}^{-1}.$$

We can make our lives a lot easier by shifting the variables, though, by changing some of the constants. Using the result of Question 17, the density of $C_i - 270$ ml is given by:

$$f_{C_i - 270\,\text{ml}}(v) = \frac{\pi}{120}\sin\left(\frac{\pi v}{60}\right) \text{ (ml)}^{-1},$$

where the support is now shifted to $[0, 60]$. We can also scale the distribution using the result of Question 18. The distribution of $\tilde{C}_i = \pi(C_i - 270)/60$ (which involves multiplication by $\pi/60$, whose inverse operation is multiplication by $60/\pi$) is given by:

$$f_{\tilde{C}_i}(v) = \frac{1}{2}\sin(v),$$

where the support of the distribution is now shifted to $[0, \pi]$. Much better!

The event we are interested in, $C_1 + C_2 > 630$ ml, is now transformed to:

$$\frac{\pi}{60}(C_1 - 270) + \frac{\pi}{60}(C_2 - 270) > \frac{3\pi}{2} \qquad \Leftrightarrow \qquad \tilde{C}_1 + \tilde{C}_2 > \frac{3\pi}{2}.$$

We know that the density of $\tilde{C}_1 + \tilde{C}_2$ is given by the convolutional integral:

$$\begin{aligned} f_{\tilde{C}_1+\tilde{C}_2}(v') &= \int_{-\infty}^{\infty} f_{\tilde{C}_1}(v) f_{\tilde{C}_2}(v'-v)\,dv \\ &= \frac{1}{2}\int_0^{\pi} \sin(v) f_{\tilde{C}_2}(v'-v)\,dv \end{aligned}$$

Observe that $0 < v < \pi$ in the integral, hence $v' - \pi < v' - v < v'$. This gives rise to the various cases:

- If $v' < 0$, then the integral vanishes, and the density is zero.
- If $0 < v' < \pi$, then the density becomes:

$$\begin{aligned} f_{\tilde{C}_1+\tilde{C}_2}(v') &= \frac{1}{4}\int_0^{v'} \sin(v)\sin(v'-v)\,dv \\ &= \frac{1}{8}\int_0^{v'} \left(\cos\left(2v - v'\right) - \cos(v')\right)\,dv \\ &= \frac{1}{8}\left[\frac{1}{2}\sin(2v - v') - v\cos(v')\right]_0^{v'} \\ &= \frac{1}{8}\left(\sin(v') - v'\cos(v')\right). \end{aligned}$$

- If $\pi < v' < 2\pi$, then the density becomes:

$$\begin{aligned}
f_{\tilde{C}_1+\tilde{C}_2}(v') &= \frac{1}{4}\int_{v'-\pi}^{\pi} \sin(v)\sin(v'-v)\,dv \\
&= \frac{1}{8}\left[\frac{1}{2}\sin(2v-v') - v\cos(v')\right]_{v'-\pi}^{\pi} \\
&= \frac{1}{8}\left(-\frac{1}{2}\sin(v') - \pi\cos(v') - \frac{1}{2}\sin(v') + (v'-\pi)\cos(v')\right) \\
&= \frac{1}{8}\left((v'-2\pi)\cos(v') - \sin(v')\right).
\end{aligned}$$

- If $v' > 2\pi$, the density vanishes entirely.

With all that in mind, the required probability is:

$$\frac{1}{8}\int_{3\pi/2}^{2\pi} \left((v'-2\pi)\cos(v') - \sin(v')\right)\,dv'.$$

Observe that:

$$\int v'\cos(v')\,dv' = v'\sin(v') + \cos(v') + c.$$

Hence, the integral is given by:

$$\begin{aligned}
\frac{1}{8}\left[v'\sin(v') + 2\cos(v') - 2\pi\sin(v')\right]_{3\pi/2}^{2\pi} &= \frac{1}{8}\left[2 + \frac{3\pi}{2} - 2\pi\right] \\
&= \frac{1}{8}\left[2 - \frac{\pi}{2}\right] \\
&= \frac{1}{16}\left(4-\pi\right).
\end{aligned}$$

(c) By standard properties of the mean, we expect the mean for two cups to be 600 ml. Evaluating the integrals instead, we have:

$$\begin{aligned}
&\frac{1}{8}\int_{0}^{\pi} \left(v'\sin(v') - (v')^2\cos(v')\right)\,dv' + \frac{1}{8}\int_{\pi}^{2\pi} \left(v'(v'-2\pi)\cos(v') - v'\sin(v')\right)\,dv' \\
&= \ldots = \pi,
\end{aligned}$$

from doing lots of integration by parts. This is the expectation for $\tilde{C}_1 + \tilde{C}_2$, hence we have:

$$\mathbb{E}[C_1 + C_2] = \frac{60}{\pi}\mathbb{E}[\tilde{C}_1 + \tilde{C}_2] + 2\cdot 270 = 600,$$

as anticipated.

Similarly, by standard properties of the variance, we expect the variance for two cups to be $1800(1-8/\pi^2)$. Evaluating the integrals instead, we have the expectation of $(\tilde{C}_1 + \tilde{C}_2)^2$ given by:

$$\frac{1}{8}\int_0^\pi \left((v')^2 \sin(v') - (v')^3 \cos(v')\right)\, dv' + \frac{1}{8}\int_\pi^{2\pi} \left((v')^2 (v' - 2\pi)\cos(v') - (v')^2 \sin(v')\right)\, dv'$$

$$= ... = \frac{3\pi^2}{2} - 4,$$

again doing lots of laborious integration by parts. Subtracting the previous result squared, we obtain the variance of $\tilde{C}_1 + \tilde{C}_2$, given by $\frac{1}{2}\pi^2 - 4$. This is the variance for $\tilde{C}_1 + \tilde{C}_2$, hence:

$$\mathrm{Var}[C_1 + C_2] = \frac{60^2}{\pi^2}\left(\frac{1}{2}\pi^2 - 4\right) = 3600\left(\frac{1}{2} - \frac{4}{\pi^2}\right) = 1800\left(1 - \frac{8}{\pi^2}\right),$$

as anticipated.