Practice Questions (week 11)

Semester 2, 2019

These questions are about discrete random variables, Bernoulli and binomial random variables, and continuous random variables. Difficult questions are starred. Good places to go for further questions on this topic include the exercises in:

- Ross, *A first course in Probability* (6th Ed.), Chapters 4 and 5.

1. Consider a game in which a fair six-sided die is rolled. If you roll a multiple of 2 you lose \$1, but if you roll a multiple of 3 you gain \$1. Describe a random variable that denotes your winnings and determine its expectation?

   Solution: Let X be the random variable denoting your winnings. Rolling a 1, 5 or 6 results in winning $X = 0$ dollars. Rolling a 2 or 4 results in winning $X = -1$ dollars. Rolling a 3 results in winning $X = 1$ dollar. Therefore $\Pr(X = 0) = \frac{3}{6} = \frac{1}{2}$, $\Pr(X = -1) = \frac{2}{6} = \frac{1}{3}$ and $\Pr(X = 1) = \frac{1}{6}$. Thus

   $$\mathbb{E}[X] = -1 \times \frac{1}{3} + 0 \times \frac{1}{2} + 1 \times \frac{1}{6} = -\frac{1}{6}.$$

---

2. Consider rolling two fair (six-sided) dice and let X be a random variable denoting the difference between the two numbers rolled, specifically if a and b are rolled we take $X = |a - b|$.

   (a) Describe the sample space for the experiment and determine the possible values that X can take on.

   (b) What is $\Pr(X = 2)$?

   (c) What is $\mathbb{E}[X]$?

   (d) What is $\mathrm{Var}(X)$?

   Solution:

   (a) The sample space is $S = \{(a, b) : a, b \in \{1, 2, 3, 4, 5, 6\}\}$. $X = |a - b|$ must be one of $0, 1, 2, 3, 4, 5$ (i.e. $|1 - 6| = |6 - 1| = 5$ is the largest possible outcome, and 0 is the smallest which occurs whenever the dice rolls are equal).

(b) The event $X = 2$ is equivalent to

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

Since $|E| = 8$, $|S| = 6 \times 6 = 36$ and S is an equi-probable sample space then $\Pr(X = 2) = \frac{8}{36} = \frac{2}{9}$.

(c) Since X is a discrete random variable with range $\{0,1,2,3,4,5\}$ we have

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{5} k \Pr(X = k) \\ &= \Pr(X = 1) + 2\Pr(X = 2) + 3\Pr(X = 3) + 4\Pr(X = 4) + 5\Pr(X = 5)\,.\end{aligned}$$

We have already determined $\Pr(X = 2)$ and need to determine the remainder. The event $X = 5$ occurs in only 2 of the 36 outcomes, namely $(1,6),(6,1)$, and therefore $\Pr(X = 5) = \frac{2}{36} = \frac{1}{18}$. The event $X = 4$ occurs in only 4 of the 36 outcomes, namely $(1,5),(2,6),(5,1),(6,2)$, and therefore $\Pr(X = 4) = \frac{4}{36} = \frac{1}{9}$. At this stage you may see a pattern that for $k > 0$ one has $\Pr(X = k) = \frac{2(6-k)}{36}$ (note however that $\Pr(X = 0) = \frac{1}{6}$ does not follow this rule). Thus, we have

$$\begin{aligned}\mathbb{E}[X] &= \frac{10}{36} + 2 \times \frac{8}{36} + 3 \times \frac{6}{36} + 4 \times \frac{4}{36} + 5 \times \frac{2}{36} \\ &= \frac{10 + 16 + 18 + 16 + 10}{36} \\ &= \frac{70}{36} = \frac{35}{18}\,.\end{aligned}$$

(Note: a sensible check one can perform is to make sure that $\sum_{k=0}^{5} \Pr(X = k) = 1$ which is indeed the case.)

(d) Since X is a discrete random variable with range $\{0,1,2,3,4,5\}$, and given $\mathbb{E}[X]$ and $\Pr(X = k)$ from the previous result, we have

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{k=0}^{5} (k - \mathbb{E}[X])^2 \Pr(X = k) \\ &= \frac{(-70)^2}{36^2}\frac{6}{36} + \frac{(-34)^2}{36^2}\frac{10}{36} + \frac{2^2}{36^2}\frac{8}{36} \\ &\quad + \frac{38^2}{36^2}\frac{6}{36} + \frac{74^2}{36^2}\frac{4}{36} + \frac{110^2}{36^2}\frac{2}{36} \\ &= \frac{29400 + 11560 + 32 + 8664 + 21904 + 24200}{36^3} \\ &= \frac{95760}{46656} = \frac{665}{324} \approx 2.05247\,.\end{aligned}$$

(Alternatively you could compute $\mathbb{E}[X^2]$ and use $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.)

3. Suppose a box contains 5 blue blocks and 5 red blocks. Consider an experiment where blocks are taken out of the box randomly, one at a time, without replacement.

   (a) Suppose the experiment stops when all of the blue blocks have been withdrawn and let X denote the total number of blocks withdrawn.

      i. What is $\Pr(X = 6)$?
      ii. What is $\Pr(X \leq 7)$?
      iii. Determine $\mathbb{E}[X]$.
      iv. What is $\mathrm{Var}(X)$?

   (b) Now, consider the case where the experiment stops as soon as at least one red and one blue block have been withdrawn and let Y denote the total number of blocks that have been withdrawn.

      i. What is $\Pr(Y \leq 8)$?
      ii. What is $\Pr(Y = 4)$?
      iii. Determine $\mathbb{E}[Y]$.
      iv. What is $\mathrm{Var}(Y)$?

Solution:

(a) i. First observe that X must be one of $5, 6, 7, 8, 9, 10$. The sample space can be considered to be all possible permutations of the sequence $\{R, R, R, R, R, B, B, B, B, B\}$ (we can consider continuing to withdraw balls until the box is empty). There are $\binom{10}{5} = 252$ possible sequences each of which are equally likely to occur. The event $X = 6$ means that one red block and all five blues blocks have been removed (leaving four red blocks in the box), and furthermore, the sixth block to be removed must have been blue. There are only 5 such sequences (the one red block being in each of the first five places), and therefore $\Pr(X = 6) = \frac{5}{252}$.

ii. We already know $\Pr(X = 6)$. No that $X = 5$ corresponds to the first five blocks withdrawn being blue which is only one of the 252 outcomes, and thus $\Pr(X = 5) = \frac{1}{252}$. For $X = 7$ we note that the seventh block must be blue and there are two red blocks amongst the first six (leaving three red blocks in the box). Since there are $\binom{6}{2} = 15$ such outcomes we have $\Pr(X = 7) = \frac{15}{252} = \frac{5}{84}$. Thus we have

$$\Pr(X \leq 7) = \frac{1 + 5 + 15}{252} = \frac{21}{252} = \frac{1}{12}\,.$$

iii. To determine $\mathbb{E}[X]$ we need to first determine $\Pr(X = k)$ for $k = 8, 9, 10$. Following similar arguments as before we can determine there are $\binom{7}{3} = 35$, $\binom{8}{4} = 56$ and $\binom{9}{5} = 126$ possible outcomes in each case respectively (and as a sensible check we note the total outcomes for $k = 5, 6, 7, 8, 9, 10$ is 252 as expected). Therefore $\Pr(X = 8) = \frac{35}{252} = \frac{5}{36}$, $\Pr(X = 9) = \frac{56}{252} = \frac{2}{9}$ and $\Pr(X = 10) = \frac{126}{252} = \frac{1}{2}$ (the last one is somewhat obvious in that we expect a 50% chance for the last block to be removed to be red, again this is a good check). Thus we can now determine that

$$\begin{aligned}\mathbb{E}[X] &= 5 \times \frac{1}{252} + 6 \times \frac{5}{252} + 7 \times \frac{15}{252} + 8 \times \frac{35}{252} + 9 \times \frac{56}{252} + 10 \times \frac{126}{252} \\ &= \frac{5 + 30 + 105 + 280 + 504 + 1260}{252} \\ &= \frac{2184}{252} = \frac{26}{3}\,.\end{aligned}$$

iv. We have

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{k=5}^{10} (k - \mathbb{E}[X])^2 \Pr(X = k) \\ &= \frac{121}{9}\frac{1}{252} + \frac{64}{9}\frac{5}{252} + \frac{25}{9}\frac{15}{252} \\ &\quad + \frac{4}{9}\frac{35}{252} + \frac{1}{9}\frac{56}{252} + \frac{16}{9}\frac{126}{252} \\ &= \frac{121 + 320 + 375 + 140 + 56 + 2016}{9 \times 252} \\ &= \frac{3028}{2268} = \frac{757}{567} \approx 1.3351\,.\end{aligned}$$

(Alternatively you could compute $\mathbb{E}[X^2]$ and use $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.)

(b) i. The sample space can be taken as the same as that considered in the first part and observe that Y must be one of $2, 3, 4, 5, 6$ (since drawing 6 blocks is guaranteed to have at least one of each colour block). Therefore it is immediate that $\Pr(Y \leq 8) = 1$.

ii. The event $Y = 4$ means that the first three blocks are blue and the fourth is red, or vice versa. Call these events $B_1B_2B_3R_4$ and $R_1R_2R_3B_4$. There are a couple of ways we can the probability of each of these. The first is to note that

$$\begin{aligned}\Pr(B_1B_2B_3R_4) &= \Pr(B_1)\Pr(B_2|B_1)\Pr(B_3|B_1B_2)\Pr(R_4|B_1B_2B_3) \\ &= \frac{5}{10}\frac{4}{9}\frac{3}{8}\frac{5}{7} \\ &= \frac{300}{5040} = \frac{5}{84}\,.\end{aligned}$$

Alternatively observe that the number of combinations for the remaining two blue and four red blocks in the box is $\binom{6}{2} = 15$ and thus $\Pr(B_1B_2B_3R_4) = \frac{15}{252} = \frac{5}{84}$. By symmetry $\Pr(R_1R_2R_3B_4)$ is identical and therefore $\Pr(Y = 4) = \frac{5}{42}$.

iii. Using arguments similar to the previous questions we can determine that $\Pr(Y = 2) = \frac{2\times 25}{90} = \frac{10}{18}$, $\Pr(Y = 3) = \frac{2\times 100}{720} = \frac{5}{18}$, $\Pr(Y = 5) = \frac{2\times 5}{252} = \frac{5}{126}$ and $\Pr(Y = 6) = \frac{2}{252} = \frac{1}{126}$. (It is good to check at this stage that $\sum_{k=2}^{6} \Pr(Y = k) = 1$!) Therefore we have

$$\begin{aligned}
\mathbb{E}[Y] &= 2 \times \frac{140}{252} + 3 \times \frac{70}{252} + 4 \times \frac{30}{252} + 5 \times \frac{10}{252} + 6 \times \frac{2}{252} \\
&= \frac{280 + 210 + 120 + 50 + 12}{252} \\
&= \frac{672}{252} = \frac{8}{3}\,.
\end{aligned}$$

iv. We have

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{k=2}^{6} (k - \mathbb{E}[X])^2 \Pr(X = k) \\
&= \frac{4}{9}\frac{140}{252} + \frac{1}{9}\frac{70}{252} + \frac{16}{9}\frac{30}{252} + \frac{49}{9}\frac{10}{252} + \frac{100}{9}\frac{2}{252} \\
&= \frac{560 + 70 + 480 + 490 + 200}{9 \times 252} \\
&= \frac{1800}{2268} = \frac{50}{63} \approx 0.7937\,.
\end{aligned}$$

(Alternatively you could compute $\mathbb{E}[Y^2]$ and use $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$.)

---

4. Let X be a non-negative integer valued random variable.

(a) Show that

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \Pr(X \geq k)\,.$$

(Hint: express $\Pr(X \geq k)$ as a sum and interchange the order of summation.)

(b) Come up with a similar expression in the case that X is an integer valued random variable (i.e. it may take on negative values).

Solution:

(a) Since we can write $\Pr(X \geq k) = \sum_{j=k}^{\infty} \Pr(X = j)$ we have

$$\begin{aligned}
\sum_{k=1}^{\infty} \Pr(X \geq k) &= \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} (X = j) \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{j} \Pr(X = j) \\
&= \sum_{j=1}^{\infty} j \Pr(X = j) \\
&= \mathbb{E}[X] \,,
\end{aligned}$$

as required.

(b) If X can be any integer then we would have

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{j=-\infty}^{\infty} j \Pr(X = j) \\
&= \sum_{j=1}^{\infty} j(\Pr(X = j) - \Pr(X = -j)) \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{j} (\Pr(X = j) - \Pr(X = -j)) \\
&= \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} (\Pr(X = j) - \Pr(X = -j)) \\
&= \sum_{k=1}^{\infty} \Pr(X \geq k) - \Pr(X \leq -k) \,.
\end{aligned}$$

Note the negative sign for $\Pr(X \leq -k)$ terms.

---

5. Let X_1, X_2, X_3 be independent Bernoulli random variables each with chance p_1, p_2, p_3 of success. Let $Y = X_1 + X_2 + X_3$ and $Z = X_1 \times X_2 \times X_3$.

   (a) What is $\mathbb{E}[Y]$ and $\mathbb{E}[Z]$?

   (b) What is $\mathrm{Var}(Y)$ and $\mathrm{Var}(Z)$?

   Solution:

(a) From the linearity of expectation we have

$$\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] = p_1 + p_2 + p_3 \,.$$

(Note this is much easier than first working out $\Pr(Y = k)$ for $k = 0, 1, 2, 3$.)
For Z, note that $Z = 1$ if and only if $X_1 = X_2 = X_3 = 1$ which occurs with probability $p_1p_2p_3$. Since $Z = 0$ the remainder of the time it follows that $\mathbb{E}[Z] = p_1p_2p_3$.

(b) For Y it is easiest to use the formula $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$ to obtain

$$\begin{aligned}
\mathrm{Var}(Y) &= \Pr(Y = 1) + 4\Pr(Y = 2) + 9\Pr(Y = 3) - (p_1 + p_2 + p_3)^2 \\
&= (p_1(1 - p_2)(1 - p_3) + (1 - p_1)p_2(1 - p_3) + (1 - p_1)(1 - p_2)p_3) \\
&\quad + 4(p_1p_2(1 - p_3) + p_1(1 - p_2)p_3 + (1 - p_1)p_2p_3) + 9p_1p_2p_3 - (p_1 + p_2 + p_3)^2 \\
&= p_1 + p_2 + p_3 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3 - (p_1 + p_2 + p_3)^2 \\
&= p_1(1 - p_1) + p_2(1 - p_2) + p_3(1 - p_3) \,.
\end{aligned}$$

For Z we have

$$\begin{aligned}
\mathrm{Var}(Z) &= (p_1p_2p_3)^2(1 - p_1p_2p_3) + (1 - p_1p_2p_3)^2 p_1p_2p_3 \\
&= p_1p_2p_3(1 - p_1p_2p_3) \,.
\end{aligned}$$

(Note that Z is effectively a Bernoulli random variable with success rate $p_1p_2p_3$.)

---

6. Consider an egg farm where the chickens have a 1 in 10,000 chance of laying a 'double yolker', that is an egg with two yolks. Assume the occurrence of these is independent and that eggs are randomly packaged into cartons.

   (a) What is the probability of at least one double yolker in a carton of 6 eggs?

   (b) What is the probability of at least one double yolker in a carton of 12 eggs?

   (c) How many eggs would you have to buy so that the chance of having at least one double yolker is 90%?

   (d) Given the number of eggs determined from the previous question, what is the probability of obtaining exactly 2 double yolkers, and what is the expected number of double yolkers?

   Solution:

(a) The number of double yolkers in a carton of 6 eggs is a binomial random variable with $n = 6$ and $p = 0.0001$. Letting X be this random variable we want to know $\Pr(X \geq 1)$, or equivalently $1 - \Pr(X = 0)$. Note that

$$\Pr(X = 0) = \binom{6}{0} p^0 (1-p)^6 = 0.9999^6 \approx 0.9994\,,$$

and therefore $\Pr(X \geq 1) \approx 0.0006$.

(b) In this case $n = 12$ and so we have

$$\Pr(X \geq 1) = 1 - \Pr(X = 0) = 1 - \binom{12}{0} p^0 (1-p)^{12} \approx 1 - 0.9988 \approx 0.0012\,.$$

(c) We want to determine the smallest n such that $\Pr(X \geq 1) > 0.9$. In general we have

$$\Pr(X \geq 1) = 1 - \Pr(X = 0) = 1 - \binom{n}{0} p^0 (1-p)^n = 1 - 0.9999^n\,.$$

Thus we require $0.9 < 1 - 0.9999^n$ which can be re-arranged to $n > \frac{\log(0.1)}{\log(0.9999)} \approx 23024.7$ (note that since $\log(0.9999) < 0$ there is a change in the direction of the inequality sign when dividing by $\log(0.9999)$). Therefore you will need to buy 23025 eggs (almost 1919 full carton of 12).

(d) With $n = 23025$ and $p = 0.0001$ we have an expectation of $\mathbb{E}[X] = np = 2.3025$ double yolkers. The probability of exactly two double yolkers is

$$\begin{aligned}\Pr(X = 2) &= \binom{23025}{2} 0.0001^2 0.9999^{23023} \\ &\approx 0.2651\,.\end{aligned}$$

---

7. Consider an experiment where a coin, which is possibly biased, is flipped some number of times. Let p be the probability of heads coming up on each flip of the coin.

   (a) Suppose the coin is flipped 10 times and comes up heads for 6 of those. What value of p is most likely to produce this outcome?

   (b) Generalise this for k heads out of n flips.

   Solution:

(a) Let X be the number of heads, then X is binomial distributed with $n = 10$ and probability p of 'success' (i.e. heads). Flipping 6 heads is the event $X = 6$ for which we know

$$\Pr(X = 6) = \binom{10}{6} p^6(1-p)^4 = 210p^6(1-p)^4 .$$

We want to determine the value of p which maximises this. Differentiating we have

$$210\big(6p^5(1-p)^4 - 4p^6(1-p)^3\big) = 210p^5(1-p)^3(6(1-p)-4p) = 210p^5(1-p)^3(6-10p) ,$$

which is zero at $0, 1, 0.6$. From these it is straightforward to determine that $p = 0.6$ maximises $\Pr(X = 6)$.
(Note: this is an example of maximum likelihood estimation).

(b) It would be reasonable to guess $p = k/n$, but let's make sure! We have in general

$$\Pr(X = k) = \binom{n}{k} p^k(1-p)^{n-k}$$

and differentiating with respect to p gives

$$\binom{n}{k} p^{k-1}(1-p)^{n-k-1}(k(1-p)-(n-k)p) = \binom{n}{k} p^{k-1}(1-p)^{n-k-1}(k-np) ,$$

which has zeros at $0, 1, k/n$ for which $p = k/n$ is the maximum (note the special cases $k = 0, n$ for which $p = 0, 1$ respectively).

---

8. Consider a continuous random variable X with probability density function

$$f(x) = \begin{cases} ax(1-x) & \text{for } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

for some $a \in \mathbb{R}$.

(a) Determine the value of a.

(b) Determine $\mathbb{E}[X]$.

(c) Determine $\mathrm{Var}(X)$.

Solution:

(a) Since the total probability must be one we have

$$\begin{aligned}
1 &= \int_0^1 f(x)\,dx \\
&= \int_0^1 ax(1-x)\,dx \\
&= a\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 \\
&= a\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{a}{6},
\end{aligned}$$

and therefore $a = 6$.

(b) The expectation of X is given by

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^1 xf(x)\,dx \\
&= \int_0^1 6x^2(1-x)\,dx \\
&= 6\left[\frac{x^3}{3} - \frac{x^4}{4}\right] \\
&= 6 \times \frac{1}{12} = \frac{1}{2}.
\end{aligned}$$

(c) The variance of X is given by

$$\begin{aligned}
\mathrm{Var}(X) &= \int_0^1 (x - \mathbb{E}[X])^2 f(x)\,dx \\
&= \int_0^1 \left(x - \frac{1}{2}\right)^2 6x(1-x)\,dx \\
&= 6\int_0^1 \frac{x}{4} - \frac{5x^2}{4} + 2x^3 - x^4\,dx \\
&= 6\left[\frac{x^2}{8} - \frac{5x^3}{12} + \frac{x^4}{2} - \frac{x^5}{5}\right] \\
&= 6 \times \frac{15 - 50 + 60 - 24}{120} = \frac{1}{20}.
\end{aligned}$$

---

9. Consider the function $f(x)$ defined by

$$f(x) = \begin{cases} c & \text{for } 0 \le x \le 1 \\ ax^{-3} & \text{for } x > 1 \\ 0 & \text{otherwise} \end{cases}$$

Suppose X is a random variable with f as its probability distribution function.

(a) What must the value of a be?

(b) What is $\Pr(X \le 2)$?

(c) Is $\frac{1}{10} < X < 10$ more likely than $X \le 2$?

(d) What is $\mathbb{E}[X]$?

(e) What is $\mathrm{Var}(X)$?

Solution:

(a) We require $1 = \int_{-\infty}^{\infty} f(x)\,dx$. Splitting this integral into the three regions we have

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)\,dx &= \int_{-\infty}^{0} 0\,dx + \int_{0}^{1} a\,dx + \int_{1}^{\infty} ax^{-3}\,dx \\ &= 0 + a + \left[-\frac{a}{2}x^{-2}\right]_{x=1}^{\infty} \\ &= a + \frac{a}{2} = \frac{3}{2}a\,.\end{aligned}$$

Therefore we require $a = \frac{2}{3}$.

(b) We have

$$\begin{aligned}\Pr(X \le 2) &= \int_{-\infty}^{2} f(x)\,dx \\ &= \int_{0}^{1} a\,dx + \int_{1}^{2} ax^{-3}\,dx \\ &= a + \left[-\frac{a}{2}x^{-2}\right]_{x=1}^{2} \\ &= a - \frac{a}{8} + \frac{a}{2} \\ &= \frac{11}{8}a = \frac{11}{12}\,.\end{aligned}$$

(Alternatively we could work this out via $\Pr(X \le 2) = 1 - \Pr(X > 2) = 1 - a\frac{1}{8} = \frac{11}{12}$.)

(c) We have

$$\begin{aligned}\Pr(\frac{1}{10} < X < 10) &= \int_{\frac{1}{10}}^{10} f(x)\,dx \\ &= \int_{\frac{1}{10}}^{1} a\,dx + \int_{1}^{10} ax^{-3}\,dx \\ &= \frac{9a}{10} + \left[-\frac{a}{2}x^{-2}\right]_{x=1}^{10} \\ &= \frac{9a}{10} - \frac{a}{200} + \frac{a}{2} \\ &= \frac{279}{200}a = \frac{93}{100} = 0.93\,.\end{aligned}$$

Since $0.93 > \frac{11}{12}$ it follows that $\frac{1}{2} < X < 10$ is more likely than $X \leq 2$.

(d) We have

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x)\, dx \\
&= \int_{-\infty}^{0} x \times 0\, dx + \int_{0}^{1} ax\, dx + \int_{1}^{\infty} ax^{-2}\, dx \\
&= 0 + \frac{a}{2} + \left[-ax^{-1}\right]_{x=1}^{\infty} \\
&= \frac{a}{2} - 0 + a = \frac{3a}{2} = 1\,.
\end{aligned}$$

(e) We have

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x)\, dx \\
&= \int_{-\infty}^{0} (x-1)^2 \times 0\, dx + \int_{0}^{1} a(x-1)^2\, dx + \int_{1}^{\infty} a\frac{(x-1)^2}{x^3}\, dx \\
&= 0 + \left[\frac{a}{3}(x-1)^3\right]_{x=0}^{1} + a\left[\log(x) + \frac{2}{x} - \frac{1}{2x^2}\right]_{x=1}^{\infty} \\
&= -\frac{a}{3} + \infty - 0 = \infty\,.
\end{aligned}$$

(Alternatively, we could try to calculate this via $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ in which case we would find that $\mathbb{E}[X^2] = \infty$.)

---

10. Consider the continuous random variable X with probability distribution function

$$f(x) = \begin{cases} 0 & x \leq -\pi \\ \dfrac{1}{2\pi}\cos(x)^2 & \text{for } x \in (-\pi, \pi) \\ 0 & x \geq \pi \end{cases}$$

Let $F(x) = \int_{-\infty}^{x} f(x)\, dx$ (which is the cumulative distribution function of X), specifically

$$F(x) = \begin{cases} 0 & x \leq -\pi \\ \dfrac{1}{2} + \dfrac{1}{2\pi}(x + \cos(x)\sin(x)) & \text{for } x \in (-\pi, \pi) \\ 1 & x \geq \pi \end{cases}$$

(a) Confirm that $\dfrac{d}{dx}F(x) = f(x)$.

(b) What is $\Pr(-1 \le X \le 1)$?

(c) What is $\mathbb{E}[X]$?

Solution:

(a) For $x \in (-\pi, \pi)$ we have

$$\begin{aligned}\frac{\partial F}{\partial x} &= \frac{1}{2\pi}\left(1 - \sin(x)^2 + \cos(x)^2\right) \\ &= \left(\cos(x)^2 + \sin(x)^2 - \sin(x)^2 + \cos(x)^2\right) \quad = \frac{1}{\pi}\cos(x)^2\,,\end{aligned}$$

noting the second equality is obtained by substituting the identity $1 = \cos(x)^2 + \sin(x)^2$. Additionally, is it clear that $\frac{d}{dx}F(x) = f(x) = 0$ for $x \le -\pi$ and $x \ge \pi$.

(b) Using the fundamental theorem of calculus we have

$$\begin{aligned}\Pr(-1 \le X \le 1) &= \int_{-1}^{1} f(x)\,dx \\ &= F(1) - F(-1) \\ &= \frac{1 + \cos(1)\sin(1)}{\pi} \approx 0.46303\,.\end{aligned}$$

(c) Observe that we can use the even symmetry of f, that is $f(x) = f(-x)$, to determine

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\pi}^{\pi} x f(x)\,dx \\ &= \int_{0}^{\pi} x f(x) + (-x) f(-x)\,dx \\ &= \int_{0}^{\pi} x f(x) - x f(x)\,dx = 0\,.\end{aligned}$$

Alternatively you could evaluate $\mathbb{E}[X]$ directly, e.g.

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\pi}^{\pi} x f(x)\,dx \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} x\cos(x)^2\,dx \\ &= \frac{1}{\pi}\left[\frac{x^2}{4} + \frac{x\sin(x)\cos(x)}{2} - \frac{\sin(x)^2}{4}\right]_{x=-\pi}^{\pi} \\ &= \frac{1}{\pi}\left(\frac{\pi^2}{4} - \frac{\pi^2}{4}\right) = 0\,.\end{aligned}$$

(noting you may need to use integration by parts and trigonometric identities to work out the third inequality). Yet another approach is to determine $\mathbb{E}[X]$ using $F(x)$, e.g.

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\pi}^{\pi} x f(x)\, dx \\
&= [xF(x)]_{x=-\pi}^{\pi} - \int_{-\pi}^{\pi} F(x)\, dx \\
&= \pi - \int_{-\pi}^{\pi} \frac{1}{2} + \frac{1}{2\pi}(x + \cos(x)\sin(x))\, dx \\
&= \pi - \left[\frac{x}{2} + \frac{x}{4\pi} + \frac{\sin(x)^2}{4\pi}\right]_{x=-\pi}^{\pi} \\
&= \pi - \pi = 0\,,
\end{aligned}$$

noting the second equality uses integration by parts (and again you may need to use integration by parts and trigonometric identities to work out the fourth inequality).

---

11. Suppose a company produces light bulbs with a lifetime L (in hours) which is exponentially distributed with parameter $\lambda = 1/5000$.

    (a) What is the probability density function f of L?

    (b) What is the probability a light bulb will work continuously for 365 days?

    (c) What is the probability a light bulb will work continuously for at least 730 days if it has already operated for 365 days?

    (d) What is the mean and variance of bulb lifetime?

Solution:

(a) Since L is exponentially distributed with $\lambda = 1/5000$ we have

$$f(x) = \lambda e^{-\lambda x} = \frac{1}{5000} e^{-x/5000}\,.$$

(b) 365 days equates to $24 \times 365 = 8760$ hours. Working continuously for 365 days is therefore equivalent to the event $L \geq 8760$. We then have

$$\begin{aligned}
\Pr(L \geq 8760) &= \int_{8760}^{\infty} f(x)\, dx \\
&= \int_{8760}^{\infty} \frac{1}{5000} e^{-x/5000}\, dx \\
&= \left[-e^{-x/5000}\right]_{x=8760}^{\infty} \\
&= e^{-8760/5000} = e^{-1.752} \approx 0.1734\,.
\end{aligned}$$

(c) Noting that 730 days is equivalent to 17520 hours, this question is asking for the conditional probability $\Pr(X > 17520|X > 8760)$. Using the formula for conditional probability we have

$$\begin{aligned}\Pr(X > 17520|X > 8760) &= \frac{\Pr((X > 17520) \cap (X > 8760))}{\Pr(X > 8760)} \\ &= \frac{\Pr(X > 17520)}{\Pr(X > 8760)}\,.\end{aligned}$$

Observe that, similar to above, one has

$$\Pr(L \geq 17520) = \int_{17520}^{\infty} f(x)\,dx = e^{-17520/5000}\,,$$

and therefore

$$\begin{aligned}\Pr(X > 17520|X > 8760) &= \frac{e^{-17520/5000}}{e^{-8760/5000}} \\ &= e^{(-17520+8760)/5000} \\ &= e^{-8760/5000} = e^{-1.752} \approx 0.1734\,.\end{aligned}$$

(Alternatively, this also follows from the *memory-less* property of the exponential distribution, that is for $s, t > 0$ and X exponentially distributed one has $\Pr(X > s + t|X > s) = \Pr(X > t)$. In this particular example we have $s = t = 8760$.)

(d) Since L is exponentially distributed with $\lambda = 1/5000$ we have $\mathbb{E}[L] = \lambda^{-1} = 5000$ hours and $\mathrm{Var}(L) = \lambda^{-2} = 25000000$.

---

12. Consider the following game. You pick a positive real number a. A random number X is then generated which happens to be exponentially distributed with parameter $\lambda = 1/10$. You are then assigned a score $S = 20 - |X - a|$.

    (a) For some fixed $a > 0$ is the probability of a score of at least 15?

    (b) Which choice of a will maximise the probability of obtaining a score of at least 15?

    (c) Do you expect this answer to change for different λ?

    (d) Suppose you pick $a = \mathbb{E}[X]$ and play this game several times with the same a. Describe the random variable which describes how many times you get a score of at least 15 after playing 8 times, then determine the expectation.

Solution:

(a) Observe that S has the range $(-\infty, 20]$. Obtaining a score of at least 15 is then the event $15 \leq S \leq 20$, or equivalently $a - 5 \leq X \leq a + 5$. Therefore, if $a \geq 5$ one has

$$\begin{aligned}\Pr(S \geq 15) &= \Pr(a-5 \leq X \leq a+5) \\ &= \int_{a-5}^{a+5} \frac{1}{10} e^{-x/10}\, dx \\ &= -e^{-(a+5)/10} + e^{-(a-5)/10}\,.\end{aligned}$$

Note that if $a < 5$ then

$$\begin{aligned}\Pr(S \geq 15) &= \Pr(0 \leq X \leq a+5) \\ &= \int_{0}^{a+5} \frac{1}{10} e^{-x/10}\, dx \\ &= 1 - e^{-(a+5)/10}\,.\end{aligned}$$

(b) If $0 < a < 5$ then we have from above $\Pr(S \geq 15) = 1 - e^{-(a+5)/10}$ which is strictly monotone increasing. One the other hand observe that for $a \geq 5$ we have

$$\Pr(S \geq 15) = -e^{-(a+5)/10} + e^{-(a-5)/10} = e^{-a/10}\left(e^{1/2} - e^{-1/2}\right)\,,$$

which is strictly monotone decreasing. Therefore the maximum is clearly attained by choosing $a = 5$.

(c) No, because X is always more likely to be in the range $[0, 10]$ than $[s, 10 + s]$ for any $s > 0$ given that the probability density function f is monotonically decreasing.
(Alternatively, it is straightforward to repeat the argument in the previous answer for arbitrary $\lambda > 0$.)

(d) Note that $\mathbb{E}[X] = \lambda^{-1} = 10$. Therefore we note that with $a = \mathbb{E}[X] = 10$ we have from (a)

$$\Pr(S \geq 15) = -e^{-15/10} + e^{-5/10} = e^{-1/2} - e^{-3/2} \approx 0.3834\,.$$

Therefore each time the game is played with this a there is a probability ≈ 0.3834 of 'success' (defined as a score of 15 or greater). Each individual game can be considered an independent Bernoulli experiment with $p = e^{-1/2} - e^{-3/2}$. Consequently, the number of times a score of at least 15 is obtained over 8 games is a binomial random variable, denoted Y say, with $n = 8$ and $p = e^{-1/2} - e^{-3/2}$. It then follows that $\mathbb{E}[Y] = np = 8(e^{-1/2} - e^{-3/2}) \approx 3.0672$.

---

13. Suppose the total annual solar irradiation in the Adelaide region, denoted X is normally distributed with mean $1750\,\mathrm{kWh/m^2}$ and standard deviation $100\,\mathrm{kWh/m^2}$. Let Z be the standard normal distribution.

    (a) What is X in terms of Z, and further, given some $x \in \mathbb{R}$, then what is the corresponding z such that $\Pr(X \le x) = \Pr(Z \le z)$?

    (b) What is the probability that the total solar irradiation exceeds $1900\,\mathrm{kWh/m^2}$?

    (c) What is the probability the total solar irradiation is between $1700\,\mathrm{kWh/m^2}$ and $1900\,\mathrm{kWh/m^2}$?

    (d) What range of solar irradiance is in the top 42.07%?

    (e) A household that uses $4000\,\mathrm{kWh}$ of energy annually decides to install $12\mathrm{m}^2$ of solar panels that are rated to be 20% efficient at converting solar radiation to electricity. Determine the distribution which describes the total energy produced by the panels annually. Using this, find the probability that solar irradiation is not enough to provide all of the energy used by the household?

    Solution:

    (a) Given X is normally distributed with mean $\mu = 1750$ and standard deviation $\sigma = 100$ then $X = \mu + \sigma Z = 1750 + 100Z$. The event $X \le x$ is then equivalent to $1750 + 100Z \le x$ which we can re-arrange to obtain $Z \le \frac{x-1750}{100}$, and thus $z = \frac{x-1750}{100}$.

    (b) We wish to know $\Pr(X > 1900)$. Expressing this in terms of the standard normal distribution we have

    $$\Pr(X > 1900) = \Pr(Z > \frac{150}{100}) = \Pr(Z > 1.5) = 1 - \Pr(Z \le 1.5)\,.$$

    Using a lookup table, or calculator, we have $\Pr(Z \le 1.5) \approx 0.9332$ and therefore $\Pr(X > 1900) \approx 1 - 0.9332 = 0.0668$.

    (c) We want to know $\Pr(1700 < X < 1900)$. Expressing this in terms of the standard normal distribution we have

    $$\Pr(1700 < X < 1900) = \Pr(-\frac{50}{100} < Z < \frac{150}{100}) = \Pr(Z < 1.5) - \Pr(Z \le -0.5)\,.$$

    Using a lookup table, or calculator, we have $\Pr(Z \le 1.5) \approx 0.9332$ and $\Pr(Z \le -0.5) \approx 0.3085$ (for the latter one might need to express it as $\Pr(Z \le -0.5) = 1 - \Pr(Z \le 0.5)$ if using a lookup table that only has non-negative z values.) It now follows that $\Pr(1700 < X < 1900) \approx 0.9332 - 0.3085 = 0.6247$.

    (d) We want to determine x such that $\Pr(X > x) = 0.4207$, or equivalently $\Pr(X \le x) = 1 - 0.4207 = 0.5793$. Using a lookup table, or calculator, we can determine that for the standard normal variable Z we have $\Pr(Z < 0.2) = 0.5793$. Therefore the corresponding x is $x = \sigma z + \mu = 100 \times 0.2 + 1750 = 1770$.

(e) The total amount of annual irradiation over 12m^2 of solar panels is $12X$, and further, at 20% efficiency, the total energy produced is $Y = 0.2 \times 12X = 2.4X$. The total solar irradiation in a given year is not enough to power the household if $Y < 4000$, or equivalently $X < \frac{4000}{2.4} = \frac{5000}{3}$. In terms of Z this is

$$Z < \frac{\frac{5000}{3} - 1750}{100} = \frac{5000 - 5250}{300} = \frac{-5}{6},$$

and using a lookup table, or calculator, we can determine $\Pr(Z < -\frac{5}{6}) \approx 0.2023$ (again, for a lookup table with non-negative z one might need to express as $\Pr(Z < -\frac{5}{6}) = 1 - \Pr(Z \leq \frac{5}{6})$).

---

14. $*$ Suppose X is a random variable with $\mathbb{E}[X] = 5$ and $\mathrm{Var}(X) = 2$.

(a) Let $Y = 3X - 8$, what is $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$?

(b) Let $Z = 6X^2 - 4X + 7$, what is $\mathbb{E}[Z]$?

Solution:

(a) Recall in general that $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ we have

$$\mathbb{E}[Y] = \mathbb{E}[3X - 8] = 3\mathbb{E}[X] - 8 = 15 - 8 = 7\,.$$

For the variance recall that in general $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$, therefore

$$\mathrm{Var}(Y) = \mathrm{Var}(3X - 8) = 9\mathrm{Var}(X) = 18\,.$$

(b) Observe that

$$\mathbb{E}[Z] = \mathbb{E}[6X^2 - 4X + 7] = 6\mathbb{E}[X^2] - 4\mathbb{E}[X] + 7\,.$$

Recall that we can express $\mathrm{Var}(X)$ as $\mathbb{E}[X^2] - \mathbb{E}[X]^2$, therefore

$$\mathbb{E}[X^2] = \mathrm{Var}(X) + \mathbb{E}[X]^2 = 2 + 5^2 = 27\,.$$

Thus

$$\mathbb{E}[Z] = 6 \times 27 - 4 \times 5 + 7 = 149\,.$$

---

15. Let $a < b$ be real numbers and consider a random variable X with uniform distribution over the interval $[a, b]$.

(a) What is the probability density function $f(x)$ of X?

(b) What is $\mathbb{E}[X]$ and $\mathrm{Var}(X)$?

(c) $*$ Let $Y = e^X$. What is the probability density function $g(x)$ of Y?

(d) $*$ Still with $Y = e^X$, what is $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$?

Solution:

(a) Since X is uniform over $[a, b]$ then f must have the form

$$f(x) = \begin{cases} c & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

for come constant c. Since $\int_a^b f(x)\,dx = c(b-a)$ is required to be 1 then $c = \frac{1}{b-a}$.

(b) The expectation of a uniform distribution is always the midpoint of the interval, that is $\frac{a+b}{2}$. Alternatively this can be determined explicitly via

$$\mathbb{E}[X] = \int_a^b \frac{x}{b-a}\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}\,.$$

For the variance we can first calculate

$$\mathbb{E}[X^2] = \int_a^b \frac{x^2}{b-a}\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}\,,$$

and then

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} \\ &= \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}\,.\end{aligned}$$

(c) Observe that since e^x is a continuous and strictly monotone increasing function then Y must take on values in the interval $[e^a, e^b]$. Further, it must be the case that for $x \in [a, b]$ one has $\Pr(X \leq x) = \Pr(Y \leq e^x)$. Since $\Pr(X \leq x) = \frac{x-a}{b-a}$, and letting $y = e^x$, then we have a cumulative distribution G for Y given by

$$G(y) = \Pr(Y \leq y) = \frac{\log(y) - a}{b-a}\,.$$

Therefore

$$g(y) = \frac{dG}{dy} = \frac{1}{b-a}\frac{1}{y}\,.$$

(Alternatively, one could use the general result that given a continuous random variable X with probability distribution function f and a strictly monotone and differentiable function h, then the probability density function of the random variable $h(X)$ is given by $f(h^{-1}(y))\frac{d}{dy}h^{-1}(y)$ over the appropriate domain.)

(d) For $\mathbb{E}[Y]$ we have

$$\begin{aligned}\mathbb{E}[Y] = \mathbb{E}[e^X] &= \int_{-\infty}^{\infty} e^x f(x)\,dx \\ &= \int_a^b e^x \frac{1}{b-a}\,dx \\ &= \frac{e^b - e^a}{b-a}\,.\end{aligned}$$

For $\mathrm{Var}(Y)$ we first observe that

$$\begin{aligned}\mathbb{E}[Y^2] = \mathbb{E}[e^{2X}] &= \int_{-\infty}^{\infty} e^{2x} f(x)\,dx \\ &= \int_a^b e^{2x} \frac{1}{b-a}\,dx \\ &= \frac{e^{2b} - e^{2a}}{2(b-a)}\,,\end{aligned}$$

and therefore

$$\begin{aligned}\mathrm{Var}(Y) &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\ &= \frac{e^{2b} - e^{2a}}{2(b-a)} - \frac{e^{2b} - 2e^{a+b} + e^{2a}}{(b-a)^2}\end{aligned}$$