1 Consider the continuous random variables X and Y with the following joint probability density function (pdf):

$$f_{X,Y}(x,y) = \begin{cases} c & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1-x \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of c.

Solution: The are of $\mathbb{R}^2$ where the joint pdf $f_{X,Y}(x,y)$ differs from 0 is a right-angled triangle with the right angle located at the origin $(0,0)$. The lengths of the two sides adjacent to the right angle are both 1. The area A of such a triangle is given by:

$$A = \frac{1}{2} \times \text{base} \times \text{height} = \frac{1}{2} \times 1 \times 1 = \frac{1}{2}.$$

Since, in order to be a probability distribution, the total volume under the joint pdf must be 1, we have:

$$c \times A = 1 \Rightarrow c \times \frac{1}{2} = 1,$$

which implies that:

$$c = 2.$$

Alternatively, we can find the value of c by integrating the joint pdf over the entire range of X and Y. This gives us:

$$\begin{aligned} 1 &= \int_0^1 \int_0^{1-x} c\, dy\, dx \\ &= \int_0^1 c\, [y]_0^{1-x}\, dx \\ &= \int_0^1 c(1-x) dx \\ &= c\left[x - \frac{x^2}{2}\right]_0^1 \\ &= c\left(1 - \frac{1}{2}\right) \\ &= \frac{c}{2}. \end{aligned}$$

Hence, we must have $c = 2$ to satisfy the condition that the total probability is 1.

[2 marks]

(b) Find the marginal distribution of X, $f_X(x)$.

Solution: To find the marginal pdf of X, we integrate the joint pdf $f_{X,Y}(x,y)$ over the possible values of Y for a given X. Since $0 \leq y \leq 1-x$, we have:

$$\begin{aligned} f_X(x) &= \int_0^{1-x} 2\,dy \\ &= 2\,[y]_0^{1-x} \\ &= 2(1-x) - 2(0) \\ &= 2 - 2x. \end{aligned}$$

Hence, the marginal pdf of X is $f_X(x) = 2 - 2x$ for $0 \leq x \leq 1$. For values of x outside this range, the marginal pdf of X is 0.

Hence:

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

[2 marks]

(c) Find the conditional distribution of Y given $X = x$, $f_{Y|X}(y|x)$.

Solution: To find the conditional distribution $f_{Y|X}(y|x)$, we just apply the definition of marginal pdfs, which implies that for the range where $f_X(x) > 0$ we divide the joint pdf, $f_{X,Y}(x,y)$, by the marginal pdf of X, which we have found in part (b). So when $0 \leq x \leq 1$ for $0 \leq y \leq 1-x$, the conditional pdf is obtained as follows:

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \frac{2}{2(1-x)} \\ &= \frac{1}{1-x} \quad \text{for } 0 \leq y \leq 1-x. \end{aligned}$$

The conditional distribution $f_{Y|X}(y|x)$ is $\frac{1}{1-x}$ over the interval $0 \leq y \leq 1-x$, and 0 otherwise. This makes sense as the total probability over the range of Y must be 1 when X is given, and the length of the interval for Y is $1-x$.

[3 marks]

(d) Calculate the conditional expectation $\mathbb{E}[Y|X]$.

Solution: To calculate the conditional expectation $\mathbb{E}[Y|X]$, we first need to find $\mathbb{E}[Y|X=x]$ based on the conditional distribution $f_{Y|X}(y|x)$ found in part (c), which we, as usual, do by integration;

$$\begin{aligned}\mathbb{E}[Y|X=x] &= \int_0^{1-x} y \cdot f_{Y|X}(y|x)\,dy \\ &= \int_0^{1-x} y \cdot \frac{1}{1-x}\,dy \\ &= \frac{1}{1-x}\int_0^{1-x} y\,dy \\ &= \frac{1}{1-x}\left[\frac{y^2}{2}\right]_0^{1-x} \\ &= \frac{1}{1-x}\cdot\frac{(1-x)^2}{2} \\ &= \frac{1}{2}\cdot(1-x).\end{aligned}$$

So, the conditional expectation function is $\mathbb{E}[Y|X=x] = \frac{1}{2}\cdot(1-x)$ while $0 \le x \le 1$. Hence the sought conditional expectation is found by replacing the outcome x with the random variable X, so $\mathbb{E}[Y|X] = \frac{1}{2}\cdot(1-X)$.

Note: This makes sense as the conditional distribution of Y given $X=x$ is a uniform distribution between 0 and $1-x$, which has exactly this expectation.

[3 marks]

(e) From the answer to part (b) we can calculate the marginal expectation of X to be $\mathbb{E}[X] = \dfrac{1}{3}$. Use this and the law of total expectation to find the marginal expectation of Y, $\mathbb{E}[Y]$.

Solution: By the law of total expectation, the expected value of Y can be found using the conditional expectation of Y given X:

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

We start by plugging in the answer from part (d), that $\mathbb{E}[Y|X] = \frac{1}{2}\cdot(1-X)$ to find that

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{1}{2} \cdot (1 - X)\right] = \frac{1}{2} \cdot (1 - \mathbb{E}[X]).$$

We then use the information given in the question text, that $\mathbb{E}[X] = \frac{1}{3}$, to get:

$$\mathbb{E}[Y] = \frac{1}{2} \cdot \left(1 - \frac{1}{3}\right) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Finally we conclude that, the marginal expectation of Y is also $\mathbb{E}[Y] = \frac{1}{3}$, which make sense due to the symmetry of the joint distribution.

[2 marks]

2 Consider a Poisson process with rate $\lambda = 3$.

(a) Find the distribution of the waiting time, T, for this process, including parameter(s).

Solution: The waiting time in a Poisson process with rate λ is exponentially distributed. Specifically, for a Poisson process with rate $\lambda = 3$, the waiting time T until the first event follows an exponential distribution with the same rate. So, $T \sim \text{Exp}(3)$.

Or we could give the distribution by simply writing down the pdf, $f_T(t)$ which in the case of an exponential distribution with parameter λ is $\lambda e^{-\lambda t}$ when $t \geq 0$, so

$$f_T(t) = \begin{cases} 3e^{-3t} & \text{for } 0 \leq t \\ 0 & \text{otherwise} \end{cases}$$

[2 marks]

(b) Given that an amount of time $t_w = 0.5$ has already passed since the last event in the process, calculate the probability that the total waiting time until the next event will be between 0.7 and 1.

Solution: The memoryless property of the exponential distribution tells us that the probability of waiting for an additional time given that some time has already passed is the same as waiting the additional time in the situation when no time had previously passed. Specifically, for the probability that the total waiting time is at least 0.7 we have:

$$P(T > 0.7 \,|\, t_w = 0.5) = P(T > 0.5 + 0.2 \,|\, T > 0.5) = P(T > 0.2).$$

It is obvious that an essentially identical argument holds for the other end of the interval, so that

$$P(0.7 < T < 1 \,|\, t_w = 0.5) = P(0.2 < T < 0.5).$$

and so we can either calculate this probability by applying the Cumulative distribution function or simply by direct integration of the probability density function from part (a);

$$P(0.2 < T < 0.5),$$

so

$$
\begin{aligned}
P(0.2 < T < 0.5) &= \int_{0.2}^{0.5} 3e^{-3t}\,dt \qquad = \left[-e^{-3t}\right]_{0.2}^{0.5} \\
&= -e^{-1.5} + e^{-0.6} \\
&= e^{-0.6} - e^{-1.5}. \qquad = 0.325681.
\end{aligned}
$$

So, in conclusion,

$$P(0.7 < T < 1 \,|\, t_w = 0.5) = 0.3257.$$

[3 marks]

(d) Now consider a stretch of lightly trafficked country road where cars pass a certain point according to a Poisson process with an expectation of 3 cars passing each hour.

Use some number or numbers from Table 1 to find the probability that it takes between 1 and 2 hours to observe 5 cars pass by this point.

Hint: Remember that the waiting time between two consecutive events in a Poisson process is independent of the waiting time between any other pair of consecutive events.

$x =$ \ $X \sim$	Gamma($k = 5$, $\lambda = 3$)	$\chi^2_{\nu=10}$	Normal($\mu = 1.67$, $\sigma^2 = 0.556$)
0.5	0.0186	0.0011	0.0588
1.0	0.1847	0.0186	0.1855
1.5	0.4679	0.0780	0.4115
2.0	0.7149	0.1847	0.6726
2.5	0.8679	0.3225	0.8682

Table 1: The Cumulative distribution function $F_X(x)$ for selected values of x and three different distributions of X.

Solution: The hint reminds us that we are considering the sum of 5 independent exponentially distributed random variables each with $\lambda = 3$, which is the definition of a Gamma Distribution, with $\lambda = 3$ and $k = 5$. We are now given units to work with, but that does not change the situation in any way. So, if we call the time until the 5th car passes T_5 the probability we are looking for is

$$\mathbb{P}(1 < T_5 < 2)$$

where the kind of inequalities used make no difference to the final answer. We note that this can be rewritten in terms of CDFs as

$$\mathbb{P}(1 < T_5 < 2) = F_X(2) - F_X(1)$$

when Gamma($k = 5, \lambda = 3$). So, plugging in the relevant values from the table we find that

$$\mathbb{P}(1 < T_5 < 2) = 0.7149 - 0.1847 = 0.5302.$$

[3 marks]

3 In a statistics class of 70 students, the teacher runs a surprise Quiz to check whether the students are keeping up with the course content. The quiz is multiple choice with a total of 25 questions, with each question having four answer choices. Since the students are very attentive the teacher estimates that the probability of each student getting each individual question right is 0.6, independent of all other questions.

Let X denote the total score of an individual student on this quiz.

(a) Explain why it could be appropriate to use a Normal Approximation for the random variable X in this case.

Solution: Since the total score is the sum of $n = 25$ Bernoulli random variables, each with probability of success $p = 0.6$, we conclude that the exact distribution of X is Binomial($n = 25, p = 0.6$). This is a relatively large set of trials, and the value of p is close to 0.5, making the distribution close to symmetric.

Thus this fulfills the criteria of being the sum of a large number iid random variables with a close to symmetric distribution, making it a candidate for Normal Approximation.

[1 marks]

(b) Write down the normal approximation for the random variable X, including the mean and variance.

Solution: The mean and variance of the normal approximation to a random variable are the same as the mean and variance of the exact distribution. As noted in part (a), the exact distribution of X is in this case Binomial($n = 25, p = 0.6$), so

$$\mathbb{E}[X] = n \times p = 25 \times 0.6 = 15$$

and

$$\mathrm{Var}(X) = n \times p \times (1-p) = 25 \times 0.6 \times 0.4 = 6.$$

, so

$$X \sim \text{approxNormal}(\mu = 15, \sigma^2 = 6).$$

[2 marks]

(c) Find the probability that a student scores more than 20 marks on the quiz based on the normal approximation. For your calculation, utilize the **most appropriate** value or values from the list of `R`-commands with associated outputs shown in Table 2 .

Command	**Output**
`pnorm(20, mean = 15, sd = 6)`	0.7976716
`pnorm(20, mean = 12.5, sd = 2.4495)`	0.9989002
`pnorm(20, mean = 15, sd = 2.4495)`	0.9793866
`pnorm(21, mean = 15, sd = 6)`	0.8413447

Table 2: `R`-commands and their corresponding outputs

Solution: We are after the probability

$$\mathbb{P}(X > 20) = 1 - \mathbb{P}(X \leq 20) = 1 - F_X(20).$$

When utilizing the normal approximation of X, and particularly with the options we have to chose from in Table 2, we use the value for the normal approximation of $F_X(20)$ remembering that we need to take the square root out of the variance (to find the standard deviation) when using it in `R` Therefore the third entry in the table,
`pnorm(20, mean = 15, sd = 2.4495)`
is the correct one to use. Hence

$$\mathbb{P}(X > 20) = 1 - 0.9793866 = 0.0206134.$$

So, the sought probability is

$$\mathbb{P}(X > 20) \approx 0.021.$$

[2 marks]