Question 1 – Continuous Distribution

Given the cdf F for a continuous random variable X:

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-0.01x} & x \geq 0, \end{cases}$$

(a) Determine $P(X > 50)$.

Solution:
As the CDF is $P(X \leqslant x) = F(x) = \int_{-\infty}^{x} f(t)\ dt$ we need to take the compliment of the CDF for $x = 50$.

$$\begin{aligned} P(X > 50) &= 1 - P(X \leqslant 50) \\ &= 1 - F(50) \\ &= 1 - \left(1 - e^{-0.01\times 50}\right) \\ &= 0.6065 \end{aligned}$$

(b) Determine $P(X \leqslant 100)$.

Solution:
This time we can calculate the value directly from the CDF.

$$\begin{aligned} P(X \leqslant 100) &= F(100) \\ &= 1 - e^{-0.01*100} \\ &= 1 - e^{-1} \\ &= 0.6321 \end{aligned}$$

(c) Determine $P(-10 \leqslant X \leqslant 10)$.

Solution:
Here we need to calculate $P(x \leqslant -10)$ and subtract it from $P(x \leqslant 10)$

$$\begin{aligned} P(-10 \leqslant X \leqslant 10) &= P(X \leqslant 10) - P(X \leqslant -10) \\ &= F(10) - F(-10) \\ &= 1 - e^{-0.01\times 10} - 0 \\ &= 1 - e^{-0.1} \\ &= 0.0952 \end{aligned}$$

(d) Derive the corresponding probability density function $f(x)$ for X.

Solution:
Here the pdf $f(x) = \frac{d}{dx}F(X)$. So deriving the CDF greater than or equal to 0 we get

$$\begin{aligned} f(x) &= \frac{d}{dx}F(x) \\ &= \frac{d}{dx}(1 - e^{-0.01x}) \\ &= 0.01e^{-0.01x} \end{aligned}$$

For the CDF less than 0 $f(x) = 0$. Therefore the pdf is

$$f(x) = \begin{cases} 0 & x < 0, \\ 0.01e^{-0.01x} & x \geqslant 0. \end{cases}$$

(e) Find the mean.

Solution:
We could notice that the random variable is distributed with an exponential distribution and therefore $\mu = \frac{1}{\lambda}$. On examination of the formula we see $\lambda = 0.01$ and so $\mu = 100$. However if we did not notice this we would integrate the pdf multiplied by x across the possible values that it can take.

$$\begin{aligned}\mu = E(X) &= \int_{-\infty}^{\infty} xf(x)\ dx \\ &= \int_{-\infty}^{0} x \times 0\ dx + \int_{0}^{\infty} 0.01xe^{-0.01x}\ dx \\ &= 0.01 \int_{0}^{\infty} xe^{-0.01x}\ dx \\ &= 0.01 \left[-\frac{xe^{-0.01x}}{0.01} \right]_0^\infty + 0.01 \int_0^\infty \frac{e^{-0.01x}}{0.01}\ dx \\ &= 0.01 \times 0 - 0.01 \left[\frac{e^{-0.01x}}{0.01^2} \right]_0^\infty \\ &= \frac{0}{0.01} + \frac{1}{0.01} \\ &= 100\end{aligned}$$

(f) Find the standard deviation.

Solution:
Again if we use that the random variable is exponentially distributed then $\sigma^2 = \frac{1}{\lambda^2}$ and so $\sigma^2 = 100^2$. This means the standard deviation will be $\sigma = 100$. Alternatively we can calculate this analytically by multiplying the pdf with $(x - \mu)^2$

$$\begin{aligned}\sigma^2 = Var(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\ dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)\ dx - \mu^2 \\ &= 0.01 \int_{-\infty}^{0} x^2 \times 0 + 0.01 \int_0^\infty x^2 e^{-0.01x}\ dx - 100^2 \\ &= 0.01 \left[-\frac{x^2e^{-0.01x}}{0.01} \right]_0^\infty + 0.01 \int_0^\infty \frac{2xe^{-0.01x}}{0.01}\ dx - 100^2 \\ &= 0.002 \left[-\frac{xe^{-0.01}}{0.01^2} \right]_0^\infty + 0.002 \int_0^\infty \frac{e^{-0.01x}}{0.01^2}\ dx - 100^2 \\ &= -0.002 \left[\frac{e^{-0.01x}}{0.01^3} \right]_0^\infty - 100^2 \\ &= \frac{0}{0.01^2} + \frac{2}{0.01^2} - 100^2 \\ &= 10000\end{aligned}$$

Again this means that the standard deviation is $\sigma = 100$.

Question 2 – Discrete Distribution

Consider a function p such that

x	10	11	13	16	20
$p(x)$	0.08	0.15	0.30	0.20	0.27

(a) Verify that p is a probability mass function (pmf).

Solution:
To verify that p is a probability mass function, we simply need to check that the sum of probabilities equals one.

$$\begin{aligned}\sum_{x=10}^{20} p(x) &= 0.08 + 0.15 + 0.30 + 0.20 + 0.27 \\ &= 1\end{aligned}$$

Calculate the following:

(b) $P(X \leqslant 13)$,

Solution:
To calculate this probability we simply add the probabilities of all the values of x greater than or equal to 13.

$$\begin{aligned}P(X \leqslant 13) &= \sum_{x=13}^{20} p(x) \\ &= 0.3 + 0.2 + 0.27 \\ &= 0.53\end{aligned}$$

(c) $P(X > 13)$,

Solution:
This is simply the compliment of the previous part so

$$\begin{aligned}P(X > 13) &= 1 - P(X \leqslant 13) \\ &= 1 - 0.53 \\ &= 0.47\end{aligned}$$

(d) $P(X = 11)$,

Solution:
Here we simply look up the value in the pmf

$$P(X = 11) = 0.15$$

(e) $P(X < 20 \text{ or } X \geqslant 13)$,

Solution:
Note that these intervals overlap and therefore cover the entire sample space. So

$$P(X < 20 \text{ or } X \geqslant 13) = P(X < 20 \cup X \geqslant 13) = 1$$

(f) the mean,

Solution:
Here we sum the probabilities multiplied by the respective x values to get the mean.

$$\begin{aligned}\mu = E(X) &= \sum_{x=10}^{20} xp(x) \\ &= 10 \times 0.08 + 11 \times 0.15 + 13 \times 0.3 + 16 \times 0.2 + 20 \times 0.27 \\ &= 14.95\end{aligned}$$

(g) the variance.

Solution:
Here we use that $Var(X) = E(X^2) - E(X)^2$ and calculate these expected values as above.

$$\begin{aligned}\sigma^2 = Var(X) &= E(X^2) - E(X)^2 \\ &= \sum_{x=10}^{20} x^2 p(x) - 14.95^2 \\ &= 100 \times 0.08 + 121 \times 0.15 + 169 \times 0.3 + 256 \times 0.2 + 400 \times 0.27 - 223.5025 \\ &= 12.5475\end{aligned}$$

Question 3 – Computer Fans

Suppose that the time to failure (in hours) of fans in a personal computer can be modelled by an exponential distribution with $\lambda = 0.0003$.

(a) What proportion of fans will last at least 12,000 hours?

Solution:
For an exponentially distributed variable $P(X > x) = 1 - F(x) = e^{-\lambda x}$.Using this the calculation is as follows

$$\begin{aligned}P(X \geq 12000) &= 1 - F(12000) \\ &= e^{-0.0003 \times 12000} \\ &= 0.0273237\end{aligned}$$

So the proportion of hard drives that will last at least 8000 hours is 0.0273.

(b) What proportion of fans will last at most 8,000 hours?

Solution: Again using the cdf for the exponential distribution we get the following:

$$P(X \leqslant 8000) = F(8000) = 1 - e^{-0.0003 \times 8000} = 0.909282$$

So the proportion of hard drives that will last at most 8000 hours is 0.9093

(c) What is the variance of the time until a computer fan fails?

Solution: Remember from lectures that the variance of an exponential distribution is $Var(X) = \frac{1}{\lambda^2}$, so

$$Var(X) = \frac{1}{0.0003^2} = 1.1111111 \times 10^7$$

(d) Use Monte Carlo simulation to predict the following: Assume a computer now has three independent fans and the failure of the computer occurs once all three fans are broken. What is the mean life of the computer?

Solution:
To perform this simulation in R, at each step we need to generate three random value from an exponential distributed random variable with $\lambda = 0.003$, find the maximum, then store that value. We then take the mean of the resultant vector.

```
> fanlife = mean(replicate(10^6,max(rexp(3,0.0003))))
> print(fanlife)
```

```
[1] 6104.086
```

So we can see that the mean lifetime of the fan is 6104.0865 hours.

Question 4 – Guessing on Multiple Choice Exams

A multiple-choice test contains 30 questions, each with 5 answers. Only one answer is correct. Assume that a student just guesses on each question.

(a) What is the probability that the student answers more than 15 questions correctly?

Solution: The number of correct answers is distributed with a Binomial Distribution with $n = 30$ and $p = \frac{1}{5}$. As this is a discrete distribution $P(X > 15) = P(X \geq 16)$ so the probabilities of getting 16 to 30 answers correct need to be totalled.

$$\begin{aligned} P(X \geq 16) &= P(X = 16) + P(X = 17) + \ldots + P(X = 30) \\ &= \binom{30}{16}\left(\frac{1}{5}\right)^{16}\left(\frac{4}{5}\right)^{1} 4 + \binom{30}{17}\left(\frac{1}{5}\right)^{17}\left(\frac{4}{5}\right)^{1} 3 + \ldots + \binom{30}{30}\left(\frac{1}{5}\right)^{30}\left(\frac{4}{5}\right)^{0} \\ &= 4.1915232 \times 10^{-5} + 8.6296067 \times 10^{-6} + \ldots + 1.0737418 \times 10^{-21} \\ &= 5.2387287 \times 10^{-5} \end{aligned}$$

(b) What is the probability that the student answers fewer than 10 questions correctly?

Solution: Here again as the questions is asking for strictly less than 10 questions correct and following a similar method to that of the previous question

$$\begin{aligned} P(X \leqslant 9) &= P(X = 9) + P(X = 8) + \ldots + P(X = 0) \\ &= \binom{30}{9}\left(\frac{1}{5}\right)^{9}\left(\frac{4}{5}\right)^{21} + \binom{30}{8}\left(\frac{1}{5}\right)^{8}\left(\frac{4}{5}\right)^{22} + \ldots + \binom{30}{0}\left(\frac{1}{5}\right)^{0}\left(\frac{4}{5}\right)^{30} \\ &= 0.0675636 + 0.1105586 + \ldots + 0.0012379 \\ &= 0.9389129 \end{aligned}$$

The following code generates the vector “Binvec” = pmf of a Binomial distribution with parameters $n = 10$ and $p = 0.6$. It then sums up the vector, illustrating that the sum of all of the probabilities is 1.

```
> Binvec <- dbinom(0:10,10,0.6)
> sum(Binvec) # => 1
```

(c) Modify the code above, to validate your answers in (a) and (b).

Solution:

```
> # (a)
> Binvec <- dbinom(16:30,30,1/5)
> sum(Binvec)

[1] 5.238729e-05

> #(b)
> Binvec <- dbinom(0:9,30,1/5)
> sum(Binvec)

[1] 0.9389129
```

Question 5 – Aerospace Inspections

The thickness of a flange on an aircraft component is Uniformly distributed between 0.7 and 1.2 millimetres. Determine the following:

(a) Cumulative distribution function of flange thickness.

Solution: First the probability density function of the flanges needs to be determined. This is done by dividing one by the range of the values

$$f(x) = \frac{1}{1.2 - 0.7} = \frac{1}{0.5} = 2$$

Now integrate the pdf to get the cumulative distribution function

$$F(x) = \int_{-\infty}^{x} 2 \; du = \int_{0.7}^{x} 2 \; du = 2(x - 0.7)$$

(b) Proportion of flanges that exceeds 0.93 millimetres.

Solution: The cdf above gives $P(X < x)$ so to calculate the proportion of flanges that exceeds 0.93 millimetres the following calculation is performed

$$\begin{aligned} P(X > 0.93) &= 1 - P(X < 0.93) \\ &= 1 - F(0.93) \\ &= 1 - 2(0.93 - 0.7) \\ &= 0.54 \end{aligned}$$

(c) Probability that the thickness exceeds 67% of the flanges.

Solution: This is the 0.67 quantile so the following equation is obtained

$$0.67 = 2(x - 0.7)$$

Solving this equation for x

$$\begin{aligned} 0.67 &= 2(x - 0.7) \\ 0.335 &= x - 0.7 \\ 0.335 + 0.7 &= x \\ 1.035 &= \end{aligned}$$

So the thickness needs to be at least 1.035 millimetres.

(d) Mean and standard deviation of flange thickness.

Solution: To calculate the mean the equation given in the course lecture notes or $\int_{-\infty}^{\infty} xf(x) \; dx$ can be used. However an alternative method is to set $U \sim \text{Uniform}(0, 1)$. Then $X = 0.5U + 0.7$. For U it is know that $E(X) = \frac{1}{2}$ and $Var(U) = \frac{1}{12}$, so $E(X) = 0.5E(U) + 0.7 = 0.95$. Further $Var(X) = 0.5^2 Var(U) = \frac{0.25}{12} = 0.0208333$.

So the mean is $\mu = 0.95$ mm and the standard deviation, which is the square root of the variance, is $\sigma = \sqrt{0.0208333} = 0.1443376$ mm.

(e) Assume now that you are sampling 12 independent flanges. What is the variance of the number of flanges with a thickness less than 0.76 millimetres?

Solution:
Here we first have to work out what the probability of having a flange with a thickness less than 0.76 mm. This is calculated using the cdf as

$$P(X < 0.76) = F(0.76) = 2(0.76 - 0.7) = 0.12$$

Now define the random variable Y which is the number of flanges with a thickness less than 0.76. Y will be distributed with a Binomial distribution (as the flange is either less than or not), with $p = 0.12$. The variance of a binomial variable is calculated as

$$\begin{aligned} Var(Y) &= np(1-p) \\ &= 12 \times 0.12 \times 0.88 \\ &= 1.2672 \end{aligned}$$

So the variance of the number of flanges with a thickness less than 0.76 millimetres is 1.2672 millimetres squared.

Question 6 – The Prototype Shoe

The weight of a sophisticated running shoe is normally distributed with a mean of 12 ounces and a standard deviation of 0.7 ounce.

(a) What is the probability that a shoe weighs more than 14 ounces?

Solution: This probability is the compliment of $P(X < 14)$ and so is calculated as follows.

$$\begin{aligned} P(X > 14) &= 1 - P(X < 14) \\ &= 1 - P\left(Z < \frac{14-12}{0.7}\right) \\ &= 1 - P(Z < 2.86) \\ &= 1 - 0.9979 \\ &= 0.0021 \end{aligned}$$

(b) What must the standard deviation of weights be in order for the company to state that 99% of its shoes weighs less than 14 ounces?

Solution: This is the 99% quantile so $P(Z < z) = 0.99$ when $z = 2.33$. Now take the standardisation formula and enter known values for x μ and z. Rearrange to solve for σ

$$\begin{aligned} 2.33 &= \frac{14-12}{\sigma} \\ \sigma &= \frac{2}{2.33} \\ &= 0.8584 \quad \text{actual } 0.8597166 \end{aligned}$$

(c) If the standard deviation remains at 0.7 ounce, what must the mean weight be for the company to state that 99% of its shoes weigh less than 13 ounces?

Solution: Again as $z = 2.33$ for the 99% quantile, substitute in known values for σ, z and x and rearrange for μ

$$\begin{aligned} 2.33 &= \frac{14-\mu}{0.7} \\ 1.631 &= 14 - \mu \\ \mu &= 12.369 \quad \text{actual } 12.3715565 \end{aligned}$$

Question 7 – Water Pollution

Each sample of water has a 10% chance of containing a particular organic pollutant. Assume you take a sample every hour and that the samples are independent with regard to the presence of the pollutant.

(a) Find the probability that in the next 15 samples, exactly 2 contain the pollutant.

Solution:
Define a binomial random variable X, that is the number of samples that contain the pollutant. Here $p = 0.1$ and so we can calculate this probability by

$$\begin{aligned} P(X = x) &= \binom{n}{r} p^r (1-p)^{n-r} \\ P(X = 2) &= \binom{15}{2} 0.1^2 \times 0.9^{13} \\ &= 105 \times 0.01 \times 0.2541866 \\ &= 0.2668959 \end{aligned}$$

(b) Find the probability that in the next 18 samples, more than 4 contain the pollutant.

Solution:
For this question note $P(X > 4) = 1 - P(X \leqslant 4)$ as X is a discrete random variable.

$$\begin{aligned} P(X > 4) &= 1 - P(X \leqslant 4) \\ &= 1 - \left(\binom{18}{4} 0.1^4 \times 0.9^{14} + \binom{18}{3} 0.1^3 \times 0.9^{15} + \binom{18}{2} 0.1^2 \times 0.9^{16} \right. \\ &\left. + \binom{18}{1} 0.1^1 \times 0.9^{17} + \binom{18}{0} 0.1^0 \times 0.9^{18} \right) \\ &= 1 - (0.070003 + 0.1680072 + 0.2835121 + 0.3001893 + 0.1500946) \\ &= 0.0281939 \end{aligned}$$

(c) What is the mean number of samples with pollutant during 123 samples?

Solution:
Recall that the mean of a binomial distributed random variable is $\mu = np$, therefore the mean number of samples with pollutant in 123 samples is

$$\mu = 123 \times 0.1 = 12.3$$

Question 8 – Soda Machine

The fill volume of a soda can in a soda-machine is normally distributed with a mean of 0.33 litres and a standard deviation of 0.01 litres.

(a) What is the probability a fill volume is less than 0.325 litres?

Solution:
Here we need to first standardise the values so that we can use the Standard Normal Distribution tables.

$$\begin{aligned} P(F < 0.325) &= P\left(\frac{F - \mu}{\sigma} < \frac{0.325 - 0.33}{0.01} \right) \\ &= P(Z < -0.5) \\ &= 0.3085 \end{aligned}$$

(b) If all cans less than 0.315 or greater than 0.335 litres are scrapped, what proportion of cans is scrapped?

Solution:
Here we need to add together the probability of getting a can with a fill less than 0.315 litres and the probability of getting a can with a fill greater than 0.335 litres. As these events are disjoint there is no intersection between them

$$\begin{aligned} P(F < 0.315 \cup F > 0.335) &= P(F < 0.315) + P(F > 0.335) \\ &= P(Z < -1.5) + P(Z > 0.5) \\ &= 0.0668 + 0.3085 \\ &= 0.3753 \end{aligned}$$

(c) What is the critical litre amount where 90% of cans fall below?

Solution:
Here we need to reverse the standardisation. The value for $P(Z < z) = 0.9$ is $z = 1.28$. Using the values given for μ, σ and our z we now calculate

$$\begin{aligned} \frac{x - 0.33}{0.01} &= 1.28 \\ x - 0.33 &= 0.0128 \\ x &= 0.3428 \end{aligned}$$

So the critical litre amount is 0.3428 litres.