Practical 4 Solutions

Question 1

(a) For the expectation we have

$$\begin{aligned}
E(\bar{X}_n) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\
&= \frac{1}{n} n\mu \\
&= \mu.
\end{aligned}$$

(b) And for the variance

$$\begin{aligned}
Var(\bar{X}_n) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

Question 2

Let X be the random variable for systolic blood pressure of 5-year-old boys.

We are given that $X \sim N(94, 11^2)$.

In order to calculate probabilities for a normally distributed random variable, we first need to transform the random variable to a standard normal distribution (i.e. a normal distribution with a mean of zero and standard deviation of one). To do this, we use the transformation:

$$Z = \frac{X - \mu}{\sigma}.$$

In this example, $Z = (X - 94)/11$.

(a) $P(X < 70) = P(Z < (70 - 94)/11) = P(Z < -2.182) = \Phi(-2.182) = 0.015$ (using Neave tables).

(b)

$$\begin{aligned}
P(X > 100) &= 1 - P(X < 100) \\
&= 1 - P(Z < (100 - 94)/11) = 1 - P(Z < 0.545) = 1 - 0.7071 = 0.29.
\end{aligned}$$

(c) We need to find $P(80 < X < 100)$. Standard normal tables are given for areas of the curve on the left, i.e. for the probability that Z is less than certain values, hence we try to express all probabilities as 'probabilities less than...'.

$$
\begin{aligned}
P(80 < X < 100) &= P(X < 100) - P(X < 80) \\
&= P(Z < (100 - 94)/11) - P(Z < (80 - 94)/11) \\
&= \Phi(0.545) - \Phi(-1.273) \\
&= 0.7071 - 0.1015 \\
&= 0.60.
\end{aligned}
$$

(d) Let m be the systolic blood pressure that will put someone in the top 1% of the distribution: i.e. $P(X > m) = 0.01$, or $P(X < m) = 0.99$.

In terms of the standard normal distribution:

$$
\begin{aligned}
P(Z < (m - 94)/11) &= 0.99 \\
\Phi((m - 94)/11) &= 0.99
\end{aligned}
$$

But $\Phi(2.3263) = 0.99$. Hence we have

$$
\begin{aligned}
(m - 94)/11 &= 2.3263 \\
\Rightarrow m &= 2.3263 \times 11 + 94 = 119.6\text{mm Hg}.
\end{aligned}
$$

(z-values for some percentage points of the standard normal distribution are found on p. 20 of Neave).

So someone with a systolic blood pressure of 119.6mmHg or more will be in the top 1% of the distribution.

(e) Let t be the systolic blood pressure that will put someone in the bottom 10% of the distribution: i.e. $P(X < t) = 0.10$.

In terms of the standard normal distribution $P(Z < (t - 94)/11) = 0.10$.

$$
\begin{aligned}
\Phi((t - 94)/11) &= 0.10 \\
\text{But } \Phi(-1.2816) &= 0.10 \text{ (since } \Phi(+1.2816) = 0.90) \\
\text{Hence } (t - 94)/11 &= -1.2816 \\
m &= -1.2816 \times 11 + 94 = 79.9\text{mm Hg}.
\end{aligned}
$$

Question 3

(a) The random variable X is the number of fits that the patient has up until t days. We are told that fits occur at a rate of λ per day. Thus $X \sim \text{Poisson}(\lambda t)$.

(b) T is a random variable that is the time to the next fit. The event that the first fit happens before time t, $(T \leq t)$, is equivalent to the event that there is at least one fit before t, i.e. $(X \geq 1)$.

(c) From (b), we see that

$$
\begin{aligned}
F(t) &= \Pr(T \leq t) \\
&= \Pr(X \geq 1) \\
&= 1 - \Pr(X = 0)
\end{aligned}
$$

We know that $X \sim \text{Poisson}(\lambda t)$ and so

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

Then

$$\begin{aligned} F(t) &= 1 - \Pr(X = 0) \\ &= 1 - e^{-\lambda t} \end{aligned}$$

(d) We now use the relationship between $F(t)$ and $f(t)$ to compute $f(t)$ as follows:

$$\begin{aligned} f(t) &= \frac{d}{dt}F(t) \\ &= \frac{d}{dt}(1 - e^{-\lambda t}) \\ &= \lambda e^{-\lambda t} \end{aligned}$$

This shows that the time between independent Poisson events has an Exponential distribution. This result will be used in other courses on the MSc where there is interest in the time to some event like remission for cancer studies or time to death. Outside of medical statistics it is widely used in queuing theory applications such as waiting on hold in call centres!

Additional: Question 4

(a)

If the value of $f(x)$ represented a probability, then the probability of X taking any value between 0 and 1 would be 1, meaning that it would be certain to take that value: this is clearly not the case, so the value of $f(x)$ cannot represent a probability. With a continuous density function probabilities are represented by areas under the curve, not by the vertical axis value.

(b)

Recall the following form of the probability density function for a Uniform(a, b) random variable:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b; \\ 0, & \text{otherwise} \end{cases}$$

Therefore for a Uniform(0,1) random variable:

$$f(x) = \begin{cases} 1, & \text{for } 0 \leq x \leq 1; \\ 0, & \text{otherwise} \end{cases}$$

We find $E(X)$ and $Var(X)$ using the definitions from the lecture

$$E(X) = \int_0^1 x f(x) dx = \int_0^1 x \cdot 1 dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

$$E(X^2) = \int_0^1 x^2 f(x) dx = \int_0^1 x^2 \cdot 1 dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

(c) using (b)

First note the fact

$$Y \sim \text{Uniform}(a, b) \quad \Rightarrow \quad Y = (b-a)X + a \quad \text{where} \quad X \sim \text{Uniform}(0,1)$$

Once we have made this transformation, the rest is easy:

$$\begin{aligned}
E(Y) &= E((b-a)X + a) \\
&= (b-a)E(X) + a \\
&= \frac{1}{2}(b-a) + a && \text{(from (b))} \\
&= \frac{1}{2}(a+b)
\end{aligned}$$

$$\begin{aligned}
Var(Y) &= Var((b-a)X + a) \\
&= (b-a)^2 Var(X) \\
&= \frac{1}{12}(b-a)^2 && \text{(from (b))}
\end{aligned}$$

(c) without using (b)

$$\begin{aligned}
E(Y) &= \int_a^b y f(y) dy \\
&= \int_a^b y \frac{1}{b-a} dy \\
&= \frac{1}{b-a} \int_a^b y dy \\
&= \frac{1}{b-a} \left[\frac{y^2}{2} \right]_a^b \\
&= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} \\
&= \frac{1}{b-a} \cdot \frac{(b-a)(b+a)}{2} \\
&= \frac{b+a}{2}
\end{aligned}$$

$$
\begin{aligned}
E(Y^2) &= \int_a^b y^2 f(y)dy \\
&= \int_a^b y^2 \frac{1}{b-a} dy \\
&= \frac{1}{b-a} \int_a^b y^2 dy \\
&= \frac{1}{b-a} \left[\frac{y^3}{3} \right]_a^b \\
&= \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} \\
&= \frac{1}{b-a} \cdot \frac{(b-a)(b^2+ab+a^2)}{3} \\
&= \frac{(b^2+ab+a^2)}{3}
\end{aligned}
$$

$$
\begin{aligned}
Var(Y) &= E(Y^2) - (E(Y))^2 \\
&= \frac{(b^2+ab+a^2)}{3} - \left(\frac{b+a}{2} \right)^2 \\
&= \frac{(b^2+ab+a^2)}{3} - \frac{b^2+2ab+a^2}{4} \\
&= \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12} \\
&= \frac{b^2-2ab+a^2}{12} \\
&= \frac{1}{12}(b-a)^2
\end{aligned}
$$

Additional: Question 5

(a) From the definition in the lecture a function $f(y)$ is a pdf if $f(y) \geq 0$ and $\int_{-\infty}^{\infty} f(y)dy = 1$. So for our particular

$$f(y) = ky^2(1-y),\ 0 < y < 1$$

we can see that $f(y) \geq 0$ for $0 < y < 1$ and $k > 0$. So we just need to verify the value of k such that the integral equals 1.

$$
\begin{aligned}
\int_0^1 ky^2(1-y)dy &= \int_0^1 k(y^2 - y^3)dy \\
&= k\left(\left[\frac{y^3}{3} - \frac{y^4}{4} \right]_0^1 \right) \\
&= k\left(\frac{1}{3} - \frac{1}{4} \right) \\
&= k\left(\frac{1}{12} \right)
\end{aligned}
$$

So $k = 12$ and hence the function $f(y) = 12y^2(1-y)$, $0 < y < 1$ is a pdf.

(b)

Using the definition of a cdf we have

$$\begin{aligned}
F(y) &= \Pr(Y \leq y) \\
&= \int_0^y 12t^2(1-t)dt \\
&= 12\left(\left[\frac{t^3}{3} - \frac{t^4}{4}\right]_0^y\right) \\
&= 12\left(\frac{y^3}{3} - \frac{y^4}{4}\right) \\
&= 12\left(\frac{4y^3 - 3y^4}{12}\right) \\
&= y^3(4-3y)
\end{aligned}$$

(c)

$$\begin{aligned}
\Pr(Y < 0.2) &= F(0.2) \\
&= (0.2)^3(4 - 3(0.2)) \\
&= 0.0272
\end{aligned}$$

$$\begin{aligned}
\Pr(Y > 0.6) &= 1 - \Pr(Y \leq 0.6) \\
&= 1 - F(0.6) \\
&= 1 - (0.6)^3(4 - 3(0.6)) \\
&= 0.5248
\end{aligned}$$

The more eagle eyed students with a probability and statistics background would have recognised $f(y)$ as being the pdf of a Beta distribution.