Question 1

Suppose that the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ are observations of the independent random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, where each random variable each follow a Poisson distribution with parameter $\lambda > 0$. Recall that the if random variable $X \sim \text{Pois}(\lambda)$, then X has the probability mass function

$$p_X(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & \text{if } x \in \{0, 1, 2, \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

Find the maximum likelihood estimate of λ given the data $\mathbf{x}$.

Solution to Question 1

Since the random variables are independent, the likelihood can be written as

$$L(\theta|\mathbf{x}; \lambda) = f(\mathbf{x}|\theta; \lambda) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{e^{-n\lambda} \prod_{i=1}^{n} \lambda^{x_i}}{\prod_{i=1}^{n} x_i!} = \frac{e^{-n\lambda} \lambda^{n\overline{x}}}{\prod_{i=1}^{n} x_i!} = \frac{e^{-n\lambda} \lambda^{n\overline{x}}}{\prod_{i=1}^{n} x_i!}$$

In order to find the maximum likelihood estimate of λ, it may be easier to work with the log-likelihood:

$$\begin{aligned} \log L(\lambda|\mathbf{x}) &= \log\left(e^{-n\lambda}\right) + \log\left(\lambda^{n\overline{x}}\right) - \log\left(\prod_{i=1}^{n} x_i!\right) \\ &= -n\lambda + n\overline{x} \log \lambda - \sum_{i=1}^{n} \log(x_i!) \end{aligned}$$

To find the maximum, we compute the derivative

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \log L(\lambda|\mathbf{x}) = -n + \frac{n\overline{x}}{\lambda}.$$

Setting the derivative equal to zero,

$$\begin{aligned} -n + \frac{n\overline{x}}{\lambda} &= 0 \\ \Rightarrow n\lambda &= n\overline{x} \\ \Rightarrow \lambda &= \overline{x}. \end{aligned}$$

We need to check that this is a maximum, so compute the second derivative

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} \log L(\lambda|\mathbf{x}) = -\frac{n\overline{x}}{(\lambda)^2} < 0$$

for any value of $\lambda > 0$, and so the second derivative is negative when $\lambda = \overline{x}$, so there is (at least) a local maximum at $\lambda = \overline{x}$.

To show it is a global maximum, we must check the value of the likelihood at the limits of λ at 0 and ∞. Simplifying the likelihood slightly by setting $a = n\overline{x} \geq 0$ (since each $x_i \geq 0$) and $b = \prod_{i=1}^{n} x_i! \geq 0$ (since each $x_i \geq 0$ and $k! \geq 1$ for $k \geq 0$), we have

$$L(\lambda|\mathbf{x}) = \frac{e^{-n\lambda} \lambda^{n\overline{x}}}{\prod_{i=1}^{n} x_i!} = \frac{\lambda^a}{b e^{n\lambda}}$$

Then, first we calculate

$$\lim_{\lambda \to 0} L(\lambda|\mathbf{x}) = \frac{0}{b \cdot 1} = 0$$

If we tried to compute the limit $\lim_{\lambda \to \infty} L(\lambda|\mathbf{x})$ naively, then we would have an expression of the form $\frac{\infty}{\infty}$. However, we can use L'Hospital's rule to evaluate this limit correctly. We note that a must be a nonnegative integer, since each x_i is a nonnegative integer and so $n\overline{x} = \sum x_i$ is a nonnegative integer. Therefore,

$$\begin{aligned} \lim_{\lambda \to \infty} L(\lambda|\mathbf{x}) &= \lim_{\lambda \to \infty} \frac{\lambda^a}{be^{n\lambda}} \\ &= \lim_{\lambda \to \infty} \frac{a!}{bn^a e^{n\lambda}} \qquad \text{(after taking } a \text{ derivatives)} \\ &= 0. \end{aligned}$$

Therefore, $\widehat{\lambda} = \overline{x}$ is indeed a global maximum, and so is the maximum likelihood estimate of λ.

Question 2

Suppose you are on a gameshow and win a prize: there are two containers, labelled A and B, that are each filled with 1000 banknotes, where each banknote is either a £10 note, a £20 note or a £50 note. The prize is that you can choose to keep one of the containers. Before you choose a container, you are told the distributions of the banknotes in each of the two containers is different. Writing the sample space as $\Omega = \{10, 20, 50\}$, and considering the parameter θ as identifying the distribution, $\theta \in \{1, 2\}$, you are given the information that the two distributions are summarised by the following table, where f_θ is the probability mass function of the distribution when the parameter value is θ:

	$\omega = 10$	$\omega = 20$	$\omega = 50$
$f_1(\omega)$	0.3	0.4	0.3
$f_2(\omega)$	0.3	0.1	0.6

Furthermore, you are allowed to sample from one of the containers before making our choice: you can pull exactly one banknote out of exactly one of the containers, and then choose to keep either that container or the other container. We plan to use maximum likelihood estimation to aid us in choosing which container to keep (and you would like to keep the container with the most money).

(a) What is the maximum likelihood estimate (MLE) for θ if you sample $\omega = 50$?

(b) What is expected amount of money in each container?

(c) If you chose to sample from container A and pulled out a £50, would you prefer to keep container A or container B? (Supposing you want to keep the container with the most money.)

(d) Again, suppose you choose to sample from container A and pull out a £50. How many times more (or less) plausible/likely is it that container A is the container with the most money?

(e) What is the MLE for θ if you sample $\omega = 20$?

(f) Suppose you choose to sample from container A and pull out a £20. Would you choose to keep container A or container B?

(g) Again, suppose you choose to sample from container A and pull out a £20. How many times more (or less) plausible/likely is it that container A is the container with the most money?

(h) What is the MLE for θ if you sample $\omega = 10$?

(i) If you chose to sample from container A and pulled out a £10, would you choose to keep container A or container B?

(j) Is the MLE always unique?

Solution to Question 2

Part (a):

Since

$$L(\theta = 2|\omega = 50) = 0.6 > 0.3 = L(\theta = 1|\omega = 50),$$

the MLE is given by $\widehat{\theta} = 2$.

Part (b):

Consider the container cointaining banknotes with distribution $\theta = 1$. Let X_1 be the value of a banknote in this container, and let Y_1 be the amount of money in this container. Then

$$\begin{aligned}
\mathrm{E}(Y_1) = \mathrm{E}(1000X_1) = 1000\mathrm{E}(X_1) &= 1000 \sum_{\omega \in \{10,20,50\}} \omega \cdot f_1(\omega) \\
&= 1000(10 \cdot 0.3 + 20 \cdot 0.4 + 50 \cdot 0.3) = 1000(3 + 8 + 15) \\
&= 26000
\end{aligned}$$

So the expected amount in the container with $\theta = 1$ is £23,000.

Now consider the container cointaining banknotes with distribution $\theta = 2$. Let X_2 be the value of a banknote in this container, and let Y_2 be the amount of money in this container. Then

$$\begin{aligned}
\mathrm{E}(Y_2) = \mathrm{E}(1000X_2) = 1000\mathrm{E}(X_2) &= 1000 \sum_{\omega \in \{10,20,50\}} \omega \cdot f_1(\omega) \\
&= 1000(10 \cdot 0.3 + 20 \cdot 0.1 + 50 \cdot 0.6) = 1000(3 + 2 + 30) \\
&= 35000
\end{aligned}$$

So the expected value of the amount of money in the container with distribution $\theta = 2$ is £35,000, and so is the most valuable container.

Part (c):

From Part (a) the MLE is $\widehat{\theta} = 2$. In other words container A, which we have sampled from, is most likely to be the container with distribution $\theta = 2$, i.e. the container with the most money, so we would prefer to keep container A.

Part (d):

Since

$$L(\theta = 2|\omega = 50) = 0.6 = 2 \times 0.3 = 2L(\theta = 1|\omega = 50),$$

we could say that it is **twice as likely** that container A is the container with the most money.

Part (e):

$$L(\theta = 1|\omega = 20) = 0.4 > 0.1 = L(\theta = 2|\omega = 20),$$

the MLE is given by $\widehat{\theta} = 1$.

Part (f):

From Part (e) the MLE is $\widehat{\theta} = 1$, so container A is most likely to be the container with distribution $\theta = 1$, i.e. the **not** container with the most money (and the most), so we would prefer to take container B.

Part (g):

Since

$$L(\theta = 1|\omega = 20) = 0.4 = 4 \times 0.1 = 4L(\theta = 2|\omega = 20),$$

we could say that it is **four times less likely** that container A is the container with the most money, as opposed to container B.

Part (h):

$$L(\theta = 1|\omega = 10) = 0.3 = L(\theta = 2|\omega = 10)$$

Therefore, in this case the MLE is given by both $\widehat{\theta} = 1$ and $\widehat{\theta} = 2$.

Part (i):

From Part (h), pulling out a £10 is equally likely to be from the container with distribution $\widehat{\theta} = 1$as it is to be from the container with distribution $\widehat{\theta} = 2$. Therefore, statistically, given the information we have, deciding to keep A or rather take B are equally good decisions.

Part (j):

As Part (i) shows, there can be situations where the MLE is not unique.

Question 3

Suppose that for every batch of lightbulbs produced in a factory an unknown proportion θ are defective. Suppose that a random sample of n lightbulbs is taken from a batch, and for $i = 1, 2, \ldots, n$ let the random variable $X_i = 1$ if the ith lightbulb is defective and let $X_i = 0$ otherwise. We assume that the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are independent, and we observe $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

(a) What distribution can we use to model each X_i?

(b) Given your answer in (a), write down the probability mass/density function $f(x_i|\theta)$.
Hint: the p.m.f./p.d.f. should be a polynomial in θ.

(c) Given your answer in (b), compute down the likelihood function $L(\theta|\mathbf{x})$.

(d) Compute the maximum likelihood estimate for θ, given the observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

(e) Write down the maximum likelihood estimator for θ, given the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

Solution to Question 3

Note that Example 8.5.7 of the lecture notes is very similar but below is a more complete solution (in the notes, I neglected to compute second derivative, and skipped over how we obtained the likelihood for the Bernoulli distribution).

Part (a):

Since the random variable X_i is equal to 1 with (unknown) probability θ, and 0 otherwise (i.e. with probability $1-\theta$), we can model $X_i \sim \text{Bern}(\theta)$.

Part (b):

$$f(x_i|\theta) = \begin{cases} \theta^{x_i} (1-\theta)^{1-x_i}, & \text{if } x_i \in \{0, 1\}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that this is equivalent to:

$$f(x_i|\theta) = \begin{cases} \theta, & \text{if } x_i = 1, \\ 1-\theta, & \text{if } x_i = 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is the probability mass function of a random variable following a Bernoulli distribution with parameter θ.

Part (c):

Since the X_i are assumed to be independent, the joint probability mass function of $\mathbf{X}$ is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\left(\sum_{i=1}^{n} x_i\right)} (1-\theta)^{\left(\sum_{i=1}^{n} (1-x_i)\right)} = \theta^{n\overline{x}} (1-\theta)^{n-n\overline{x}}$$

Then,

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \theta^{n\overline{x}} (1-\theta)^{n-n\overline{x}}.$$

Part (d):

Instead of trying to maximise the likelihood directly, it will be easier to maximise the log-likelihood.

$$\begin{aligned} \log L(\theta|\mathbf{x}) &= \log\left(\theta^{n\overline{x}} (1-\theta)^{n-n\overline{x}}\right) \\ &= \log\left(\theta^{n\overline{x}}\right) + \log\left((1-\theta)^{n-n\overline{x}}\right) \\ &= n\overline{x} \log\theta + (n-n\overline{x}) \log(1-\theta) \end{aligned}$$

For the moment, assume that the x_i are not all 0 or not all 1. Taking the derivative,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\log L(\theta|\mathbf{x}) = \frac{n\overline{x}}{\theta} + \frac{n-n\overline{x}}{1-\theta}(-1) = \frac{n\overline{x}}{\theta} - \frac{n-n\overline{x}}{1-\theta}$$

Setting the derivative equal to 0,

$$0 = \frac{n\overline{x}}{\theta} - \frac{n-n\overline{x}}{1-\theta} = \frac{n\overline{x}(1-\theta)-(n-n\overline{x})\theta}{\theta(1-\theta)}$$

$$\begin{aligned}
&\Rightarrow 0 = n\overline{x}(1-\theta)-(n-n\overline{x})\theta = n\overline{x} - n\overline{x}\theta - n\theta + n\overline{x}\theta \\
&\Rightarrow 0 = n(\overline{x}-\theta) \\
&\Rightarrow \theta = \overline{x}
\end{aligned}$$

To check if this is a maximum or a minimum, we need to compute the second derivative evaluated at this value, $\theta = \overline{x}$:

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log L(\theta|\mathbf{x}) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{n\overline{x}}{\theta} - \frac{n-n\overline{x}}{1-\theta}\right] \\
&= -\frac{n\overline{x}}{\theta^2} - \left(-\frac{n-n\overline{x}}{(1-\theta)^2}(-1)\right) \\
&= -\frac{n\overline{x}}{\theta^2} - \frac{n-n\overline{x}}{(1-\theta)^2} \\
&= \frac{n}{\theta^2(1-\theta)^2}\left(-\overline{x}(1-2\theta+\theta^2)-(1-\overline{x})\theta^2\right) \\
&= \frac{n}{\theta^2(1-\theta)^2}\left(-\overline{x}+2\overline{x}\theta-\theta^2\right)
\end{aligned}$$

Now,

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log L(\theta|\mathbf{x})\right|_{\theta=\overline{x}} = \frac{n}{\theta^2(1-\theta)^2}\left(\overline{x}^2-\overline{x}\right) = \frac{n\overline{x}}{\theta^2(1-\theta)^2}\left(\overline{x}-1\right)$$

Therefore, the sign of the second derivative is the same as the sign of $x(\overline{x}-1)$. Since all $x_i \in \{0,1\}$, and we assumed that the x_i are not all 0 and not all 1, then

$$\begin{aligned}
&0 < \overline{x} < 1 \\
\Rightarrow &\overline{x}(\overline{x}-1) < 0 \\
\Rightarrow &\left.\log L(\theta|\mathbf{x})\right|_{\theta=\overline{x}} < 0
\end{aligned}$$

which shows that $\theta = \overline{x}$ is a maximum.

Now, we need to check the cases that (i) all the $x_i = 0$ and (ii) all the $x_i = 1$. If all the $x_i = 0$, then

$$\log L(\theta|\mathbf{x}) = (n-n\overline{x})\log(1-\theta) = n\log(1-\theta)$$

since $\overline{x} = 0$. Note that as θ increases, this function decreases, and there is no local maximum in the interval $\theta \in (0,1)$. Therefore, the maximum occurs at $\theta = 0$, and therefore $\theta = 0 = \overline{x}$.

For the other case, suppose that all $x_i = 1$. Then $\overline{x} = 1$ and

$$\log L(\theta|\mathbf{x}) = n\overline{x}\log\theta = n\log\theta$$

This is an increasing function of θ, with no local maximum in the interval $(0, 1)$. Therefore, the maximum occurs at $\theta = 1 = \overline{x}$.

Therefore, in all cases, $\widehat{\theta} = \overline{x}$ is the maximum likelihood estimate for θ. If we were being very careful, and wanted to emphasise that this depends on observing the data $\mathbf{x}$, we would write $\widehat{\theta}(\mathbf{x}) = \overline{x}$.

Part (e):

From Part (d), the maximum likelihood estimate is $\widehat{\theta}(\mathbf{x}) = \overline{x}$. Therefore, the maximum likelihood **estimator** is $\widehat{\theta}(\mathbf{X}) = \overline{X}$.

Question 4

Suppose that the random variables $X_1, X_2, \dots, X_n$ are independent and identically distributed according to a uniform distribution on the closed interval $[0, \theta]$, for some parameter $\theta > 0$, where the exact value of the parameter θ is unknown. Given that $\mathbf{X} = (X_1, X_2, \dots X_n)$ is observed as $\mathbf{x} = (x_1, x_2, \dots, x_n)$, find the maximum likelihood estimator of θ by doing the following steps:

(a) Write down the probablity density function $f(x_i|\theta)$ for observation x_i, for $i = 1, 2, \dots, n$.

(b) Derive the likelihood $L(\theta|\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, \dots, x_n)$.

(c) Identifying any conditions the maximum likelihood estimate of θ must satisfy, find $\widehat{\theta}$, the maximum likelihood esitmate of θ.

(d) Given your answer in Part (c), write down the maximum likelihood **estimator** of θ.

Solution to Question 3

Part (a):

The probability density function $f(x_i|\theta)$ for observation x_i, where $i \in \{1, 2, \dots n\}$, has the form

$$f(x_i|\theta) = \begin{cases} \frac{1}{\theta}, & \text{if } 0 \le x_i \le \theta, \\ 0, & \text{otherwise.} \end{cases}$$

Part (b):

Since the random variables $X_1, X_2, \dots, X_n$ are independent, the joint probability density function $f(\mathbf{x}|\theta)$ of $X_1, X_2, \dots, X_n$ can be written as

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{if } 0 \le x_i \le \theta, \text{ for all } i \in \{1, 2, \dots n\}, \\ 0, & \text{otherwise.} \end{cases}$$

And therefore $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$, where $f(\mathbf{x}|\theta)$ is defined above.

Part (c):

Since the joint likelihood $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$, the maximum likelihood estimate of θ must be a value of θ such that (i) $x_i \le \theta$ for $i \in \{1, 2, \dots n\}$ and (ii) this value maximises $1/\theta^n$ among all possible values for θ.

Since $1/\theta^n$ is a decreasing function of θ, the maximum likelihood estimate will be the smallest value of θ such that $x_i \le \theta$ for $i \in \{1, 2, \dots n\}$.

Therefore, the maximum likelihood estimate of θ is $\widehat{\theta} = \max\{x_1, x_2, \dots, x_n\}$.

Part (d):

The maximum likelihood estimator is $\widehat{\theta} = \max\{X_1, X_2, \dots, X_n\}$.