Solution Sheet 4

1. When the data values $x_1, \ldots, x_n$ are the observed values of a random sample of size n from a discrete distribution with probability mass function $p(x;\theta)$, the Likelihood function has the form $L(\theta) = p(x_1;\theta)p(x_2;\theta)\cdots p(x_n;\theta)$. Writing $\hat\theta$ for the maximum likelihood estimate $\hat\theta_{\text{mle}}$, then in regular cases $\hat\theta$ satisfies the likelihood equation $\sum_{i=1}^n \partial/\partial\theta \log p(x_i;\theta) = 0$.

   Here the observations are from a Poisson(θ) distribution, so

$$\begin{aligned}
p(x;\theta) &= e^{-\theta}\theta^x/x! \\
\log p(x;\theta) &= -\theta + x\log(\theta) - \log(x!) \\
\frac{\partial}{\partial\theta}\log p(x;\theta) &= -1 + \frac{x}{\theta} \\
\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta}\log p(x_i;\theta) &= \left(-1+\frac{x_1}{\theta}\right) + \cdots + \left(-1+\frac{x_n}{\theta}\right) = -n + \frac{\sum_1^n x_i}{\theta}
\end{aligned}$$

   so the maximum likelihood estimate $\hat\theta$ satisfies the likelihood equation

$$\begin{aligned}
-n + \frac{\sum_1^n x_i}{\hat\theta} &= 0 \\
\text{i.e.} \qquad \hat\theta = \frac{\sum_1^n x_i}{n} &= \bar{x}.
\end{aligned}$$

2. When the data values $x_1, \ldots, x_n$ are the observed values of a random sample of size n from a continuous distribution with probability density function $f(x;\theta)$, the Likelihood function has the form $L(\theta) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$. Writing $\hat\theta$ for the maximum likelihood estimate $\hat\theta_{\text{mle}}$, then in regular cases $\hat\theta$ satisfies the likelihood equation $\sum_{i=1}^n \partial/\partial\theta \log f(x_i;\theta) = 0$.

   Here, adapting the method in question 1 above to the continuous case, we have

$$\begin{aligned}
f(x;\theta) &= \theta x^{\theta-1} \\
\log f(x;\theta) &= \log(\theta) + (\theta-1)\log(x) \\
\frac{\partial}{\partial\theta}\log f(x;\theta) &= \frac{1}{\theta} + \log(x) \\
\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(x_i;\theta) &= \left(\frac{1}{\theta}+\log(x_1)\right) + \cdots + \left(\frac{1}{\theta}+\log(x_n)\right) = \frac{n}{\theta} + \sum_1^n \log(x_i)
\end{aligned}$$

   so the maximum likelihood estimate $\hat\theta$ satisfies the likelihood equation

$$\begin{aligned}
\frac{n}{\hat\theta} + \sum_1^n \log(x_i) &= 0 \\
\text{i.e.} \qquad \hat\theta &= -\frac{n}{\sum_1^n \log(x_i)}
\end{aligned}$$

   For the given data, $n = 5$ and $\sum_1^5 \log(x_i) = -5.314919$ so $\hat\theta = -5/(-5.314919) = 0.94075$.

3. (a) Here the observations are from a Binomial(K, θ) distribution, so following the method in question 1 above

$$
\begin{aligned}
p(x;\theta) &= \binom{K}{x} \theta^x (1-\theta)^{K-x} \\
\log p(x;\theta) &= \log\left[\binom{K}{x}\right] + x\log(\theta) + (K-x)\log(1-\theta) \\
\frac{\partial}{\partial\theta}\log p(x;\theta) &= \frac{x}{\theta} - \frac{(K-x)}{(1-\theta)} \\
\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta}\log p(x_i;\theta) &= \left(\frac{x_1}{\theta} - \frac{(K-x_1)}{(1-\theta)}\right) + \cdots + \left(\frac{x_n}{\theta} - \frac{(K-x_n)}{(1-\theta)}\right) \\
&= \frac{\sum_1^n x_i}{\theta} - \frac{(nK - \sum_1^n x_i)}{(1-\theta)}
\end{aligned}
$$

so the maximum likelihood estimate $\hat\theta$ satisfies the likelihood equation

$$
\frac{\sum_1^n x_i}{\hat\theta} - \frac{(nK - \sum_1^n x_i)}{(1-\hat\theta)} = 0
$$

i.e. $$\textstyle\sum_1^n x_i - \hat\theta\sum_1^n x_i = \hat\theta nK - \hat\theta\sum_1^n x_i$$

i.e. $$\hat\theta = \textstyle\sum_1^n x_i/nK = \bar{x}/K$$

(b) For the given data, $n = 7, K = 10$ and $x_1 + \cdots + x_7 = 57$ so the maximum likelihood estimate is $\hat\theta = 57/70 = 0.81429$.

Let X denote the trainee's score in the examination. The probability the trainee will pass the examination is $P(X = 10 \text{ or } 9;\theta) = P(X = 10;\theta) + P(X = 9;\theta) = \theta^{10} + 10\theta^9(1-\theta)$. Thus the maximum likelihood estimate of the probability of passing is just $\hat\theta^{10} + 10\hat\theta^9(1-\hat\theta) = 0.4205$.

(c) In the lecture example, given data of 4, 5, and 1, there must have been in total 3 heads in 10 tosses (although in this experiment the '3' was fixed in advance, and the '10' the result of the random results of the coin tosses). If you analysed this as if the number of tosses was fixed in advance, and the 3 heads was the observed result, you could use part (a), with $K = 1$, $n = 10$ and $(x_1, x_2, \ldots, x_n) = (0,0,0,1,0,0,0,0,1,1)$ and by part (a) the mle for θ would again be 0.3.

4. Here the observations are from an Exponential(θ) distribution, so following the method outlined above adapted to the case of a continuous distribution,

$$
\begin{aligned}
f(x;\theta) &= \theta e^{-\theta x} \\
\log f(x;\theta) &= \log(\theta) - \theta x \\
\frac{\partial}{\partial\theta}\log f(x;\theta) &= \frac{1}{\theta} - x \\
\textstyle\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(x_i;\theta) &= \left(\frac{1}{\theta} - x_1\right) + \cdots + \left(\frac{1}{\theta} - x_n\right) = \frac{n}{\theta} - \sum_1^n x_i
\end{aligned}
$$

so the maximum likelihood estimate $\hat\theta$ satisfies the likelihood equation

$$
\frac{n}{\hat\theta} - \sum_1^n x_i = 0 \quad \text{i.e.} \quad \hat\theta = n/\sum_1^n x_i = 1/\bar{x}.
$$

For the given data, $\bar{x} = 3.812$, so $\hat\theta = 0.26233$.

(a) Since $X \sim \text{Exp}(\theta)$, the median of the distribution of the lifetimes of lamps in the population is the value m (which depends on θ) such that $P(X > m; \theta) = P(X \leq m; \theta) = \frac{1}{2}$. For an Exponential($\theta$ distribution, $P(X > x; \theta) = e^{-\theta x}$, so $e^{-\theta m} = \frac{1}{2}$, and $e^{\theta m} = 2$, giving $m = \log(2)/\theta$. Thus the maximum likelihood estimate of the median of the distribution is $\hat{m} = \log(2)/\hat{\theta} = 2.642$. Note that the sample median is 3.7.

(b) Since $P(X > x; \theta) = e^{-\theta x}$, the probability a randomly chosen lamp will survive beyond 10 hours is $P(X > 10; \theta) = \exp(-10\theta)$. Thus the maximum likelihood estimate of this survival probability is just $\exp(-10\hat{\theta}) = 0.0726$. Note that the sample proportion of values greater than 10 is $1/25 = 0.04$.

5. (a) The likelihood function is just the joint probability density function as a function of the unknown parameters μ_X, μ_Y and σ^2. Since all the random variables are independent, this is just the product of all the $n + m$ individual (marginal) probability density functions,
i.e. $f_X(x_1; \mu_X, \sigma^2) \cdots f_X(x_n; \mu_X, \sigma^2) \times f_Y(y_1; \mu_Y, \sigma^2) \cdots f_Y(y_m; \mu_Y, \sigma^2)$.
Thus the loglikelihood function is just
$l(\mu_X, \mu_Y, \sigma^2) = \sum_{i=1}^n \log f_X(x_i; \mu_X, \sigma^2) + \sum_{j=1}^m \log f_Y(y_j; \mu_Y, \sigma^2)$.

(b) Since there are three unknown parameters μ_X, μ_Y and σ^2, the maximum likelihood estimates are the joint solutions of the three equations $\partial/\partial\mu_X\ l(\mu_X, \mu_Y, \sigma^2) = 0$, $\partial/\partial\mu_Y\ l(\mu_X, \mu_Y, \sigma^2) = 0$, and $\partial/\partial\sigma\ l(\mu_X, \mu_Y, \sigma^2) = 0$. Now

$$\begin{aligned}
\frac{\partial}{\partial\mu_X}\log f_X(x;\mu_X,\sigma^2) &= \frac{(x-\mu_X)}{\sigma^2} & \frac{\partial}{\partial\mu_X}\log f_Y(y;\mu_Y,\sigma^2) &= 0 \\
\frac{\partial}{\partial\mu_Y}\log f_X(x;\mu_X,\sigma^2) &= 0 & \frac{\partial}{\partial\mu_Y}\log f_Y(y;\mu_Y,\sigma^2) &= \frac{(y-\mu_Y)}{\sigma^2} \\
\frac{\partial}{\partial\sigma}\log f_X(x;\mu_X,\sigma^2) &= -\frac{1}{\sigma}+\frac{(x-\mu_X)^2}{\sigma^3} & \frac{\partial}{\partial\sigma}\log f_Y(y;\mu_Y,\sigma^2) &= -\frac{1}{\sigma}+\frac{(y-\mu_Y)^2}{\sigma^3}
\end{aligned}$$

Thus, on summing the terms above, we get

$$\begin{aligned}
\frac{\partial}{\partial\mu_X} l(\mu_X,\mu_Y,\sigma^2) &= \textstyle\sum_{i=1}^n \displaystyle\frac{\partial}{\partial\mu_X}\log f_X(x_i;\mu_X,\sigma^2) + \sum_{j=1}^m \frac{\partial}{\partial\mu_X}\log f_Y(y_j;\mu_Y,\sigma^2) \\
&= \frac{(x_1-\mu_X)}{\sigma^2} + \cdots + \frac{(x_n-\mu_X)}{\sigma^2} + 0 + \cdots + 0 \\
&= \frac{\sum_{i=1}^n x_i - n\mu_X}{\sigma^2} \\
\text{while} \quad \frac{\partial}{\partial\mu_Y} l(\mu_X,\mu_Y,\sigma^2) &= \frac{\sum_{j=1}^m y_j - m\mu_Y}{\sigma^2} \\
\text{and} \quad \frac{\partial}{\partial\sigma} l(\mu_X,\mu_Y,\sigma^2) &= -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (x_i-\mu_X)^2}{\sigma^3} - \frac{m}{\sigma} + \frac{\sum_{j=1}^m (y_j-\mu_Y)^2}{\sigma^3}
\end{aligned}$$

(c) On inspection, we see that the equations for the maximum likelihood estimates of μ_X and μ_Y are exactly the same as they would be for individual samples from each distribution and give

$$\hat{\mu}_X = (x_1 + \cdots x_n)/n = \bar{x} \quad \text{and} \quad \hat{\mu}_Y = (y_1 + \cdots y_m)/m = \bar{y}.$$

However, both data sets contribute jointly to the maximum likelihood estimate of σ^2, and the likelihood equation for σ has solution

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{\sum_{i=1}^n (x_i-\hat{\mu}_X)^2 + \sum_{j=1}^m (y_j-\hat{\mu}_Y)^2}{n+m} \\
&= \frac{\sum_{i=1}^n (x_i-\bar{x})^2 + \sum_{j=1}^m (y_j-\bar{y})^2}{n+m} \\
&= \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2 + \sum_{j=1}^m y_j^2 - m\bar{y}^2}{n+m}.
\end{aligned}$$

(d) For the given data

$$n = 9, \quad \sum x_i = 104.0, \quad \sum x_i^2 = 1218.92$$
$$m = 10, \quad \sum y_i = 102.3, \quad \sum y_i^2 = 1071.69$$

giving

$$\hat{\mu_X} = \bar{x} = 11.5556 \qquad \hat{\mu_Y} = \bar{y} = 10.2300 \qquad \text{and} \qquad \hat{\sigma^2} = 2.2265.$$

6. (a) Since the probability that any given subject cheated on their tax return is τ and since θ denotes the probability that a randomly chosen subject will tick box 1, we have that $\theta = P(\text{Tick 1}|\text{Asked about tax}) \times P(\text{Asked about tax}) + P(\text{Tick 1}|\text{Asked to toss coin}) \times P(\text{Asked to toss coin}) = \tau \times 1/2 + 1/2 \times 1/2 = \tau/2 + 1/4$, so $\tau = 2\theta - 1/2$. Note that since $0 \leq \tau \leq 1$, the set of possible values for the parameter θ is $0.25 \leq \theta \leq 0.75$.

(b) Let $X_i = 1$ if the ith subject ticks box 1 and let $X_i = 0$ if the ith subject ticks box 0. Then $X_1, \ldots, X_n$ is a random sample from a Bernoulli(θ) distribution. Thus

$$\begin{aligned}
p(x;\theta) &= \theta^x(1-\theta)^{1-x} \\
\log p(x;\theta) &= x\log(\theta) + (1-x)\log(1-\theta) \\
\frac{\partial}{\partial\theta}\log p(x;\theta) &= \frac{x}{\theta} - \frac{(1-x)}{(1-\theta)} \\
\sum_{i=1}^n \frac{\partial}{\partial\theta}\log p(x_i;\theta) &= \left(\frac{x_1}{\theta} - \frac{(1-x_1)}{(1-\theta)}\right) + \cdots + \left(\frac{x_n}{\theta} - \frac{(1-x_n)}{(1-\theta)}\right) = \frac{\sum_1^n x_i}{\theta} - \frac{(n - \sum_1^n x_i)}{(1-\theta)}
\end{aligned}$$

This is clearly a decreasing function of θ, and so there is a unique solution $\tilde{\theta}$ to the equation

$$\begin{aligned}
\frac{\sum_1^n x_i}{\tilde{\theta}} - \frac{(n - \sum_1^n x_i)}{(1-\tilde{\theta})} &= 0 \\
\text{i.e.} \qquad \sum_1^n x_i - \tilde{\theta}\sum_1^n x_i &= \tilde{\theta}n - \tilde{\theta}\sum_1^n x_i \\
\text{i.e.} \qquad \tilde{\theta} &= \sum_1^n x_i/n = \bar{x}
\end{aligned}$$

The likelihood is increasing in θ for $\theta < \tilde{\theta}$, and decreasing for $\theta > \tilde{\theta}$. Thus the mle $\hat{\theta} = \tilde{\theta}$ if this is in the range $0.25 < \tilde{\theta} < 0.75$, but otherwise the likelihood is maximised at 0.25 or 0.75 accordingly. In summary

$$\hat{\theta} = \begin{cases} 0.25 & \text{if } \bar{x} < 0.25 \\ \bar{x} & \text{if } 0.25 < \bar{x} < 0.75 \\ 0.75 & \text{if } \bar{x} > 0.75 \end{cases}$$

For the given data, $\sum_1^n x_i/n = 8/20 = 0.4$, so the maximum likelihood estimate of θ is $\hat{\theta} = 0.4$. Since $\tau = 2\theta - 1/2$, the maximum likelihood estimate of τ is just $2\hat{\theta} - 1/2 = 0.3$. Note how this allows us to compute a maximum likelihood estimate for τ without ever knowing for each individual subject whether or not they cheated on their tax return, since we do not know which question an individual subject was asked to answer.