# MATH2033 Mathematical Statistics
# Assignment 5 Suggested Solutions

1. (a) For the method of moments, we first need to find the expected value of $X$, which is a discrete random variable with the probability distribution function as defined in the exercise.

Using the definition of the expected value of a discrete random variable, we have

$$E(X) = \sum_{k=0}^{3} k \cdot P(X = k) = \frac{\theta}{3} + \frac{4}{3} \cdot (1 - \theta) + (1 - \theta) = \frac{7}{3} - 2 \cdot \theta.$$

From the above expression, we can express $\theta$ as

$$\theta = \frac{1}{2} \cdot \left(\frac{7}{3} - E(X)\right) = \frac{7}{6} - \frac{1}{2} \cdot E(X)$$

Now, the method of moments simply suggests writing the sample mean $\bar{X}$ in place of $E(X)$, and that would be the method of moments estimate of $\theta$.

So, the desired estimate is

$$\hat{\theta} = \frac{7}{6} - \frac{1}{2} \cdot \bar{X} \tag{1}$$

The sample mean can be found as

$$\bar{X} = \frac{3+0+2+1+3+2+1+0+2+1}{10} = \frac{3}{2},$$

which would yield a method of moment estimate of $\theta$ :

$$\hat{\theta} = \frac{7}{6} - \frac{1}{2} \cdot \frac{3}{2} = \frac{5}{12} = 0.417.$$

(b) Let's first find the variance of $\hat{\theta}$. Since for any random variable $X$, and any $a, c \in \mathbb{R}$, $\mathrm{Var}(a \cdot X + c) = a^2 \cdot \mathrm{Var}(X)$, then the variance of $\hat{\theta}$ is

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}\left(\frac{7}{6} - \frac{1}{2} \cdot \bar{X}\right) = \frac{1}{4} \cdot \mathrm{Var}(\bar{X})$$

Next, since $X_1, \ldots, X_n$ are independent, then the variance of their sum is actually the sum of their variances, where $n$ is the sample size, which means that

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) = \frac{1}{n} \cdot \mathrm{Var}\left(X_1\right),$$

where the final equality holds because all $X_i$ are identically distributed, so they have the same variance.

The variance of $X_1$ can be found as

$$\operatorname{Var}(X_1) = E\left(X_1^2\right) - \left[E(X_1)\right]^2.$$

Therefore, using the definition of the expected value of a (function of a) discrete random variable, we have

$$E\left(X_1^2\right) = \sum_{k=0}^{3} k^2 \cdot P(X=k) = \frac{\theta}{3} + \frac{8}{3} \cdot (1-\theta) + 3 \cdot (1-\theta) = \frac{17}{3} - \frac{16}{3} \cdot \theta.$$

Finally, the variance of $X_1$ is

$$\operatorname{Var}(X_1) = \frac{17}{3} - \frac{16}{3} \cdot \theta - \left(\frac{7}{3} - 2 \cdot \theta\right)^2 = -4\theta^2 + 4\theta + \frac{2}{9},$$

which means that the variance of $\bar{X}$ is (here, $n = 10$ )

$$\operatorname{Var}(\bar{X}) = \frac{1}{10} \cdot \left(-4\theta^2 + 4\theta + \frac{2}{9}\right) = -\frac{2}{5} \cdot \theta^2 + \frac{2}{5} \cdot \theta + \frac{1}{45},$$

which, together with (**??**), finally gives us the variance of $\hat{\theta}$ :

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{4} \cdot \left(-\frac{2}{5} \cdot \theta^2 + \frac{2}{5} \cdot \theta + \frac{1}{45}\right) = -\frac{1}{10} \cdot \theta^2 + \frac{1}{10} \cdot \theta + \frac{1}{180}.$$

The estimated variance of $\hat{\theta}$ can be obtained by substituting $\theta$ with its method of moments estimate $\hat{\theta} = 0.417$, which would then yield

$$s_{\hat{\theta}}^2 = -\frac{1}{10} \cdot 0.417^2 + \frac{1}{10} \cdot 0.417 + \frac{1}{180} = 0.0299.$$

Lastly, the estimated standard error of $\hat{\theta}$ is simply the square root of the above variance, so

$$s_{\hat{\theta}} = \sqrt{0.0299} = 0.1728.$$

(c) Let $n$ be the sample size, and let $X_1, \ldots, X_n$ be independent identically distributed random variables with the same probability distribution function (the one described in the exercise).
To find the MLE of $\theta$, we first define the likelihood function:

$$\begin{aligned}\mathcal{L}(\theta) &= f(x_1, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) \cdots f(x_n \mid \theta) \\ &= P(X = x_1 \mid \theta) \cdots P(X = x_n \mid \theta).\end{aligned}$$

Substituting $n = 10$, given values of $X$, and the definition of the probability distribution function of $X$ yields

$$\begin{aligned}\mathcal{L}(\theta) &= P(X=0 \mid \theta)^2 \cdot P(X=1 \mid \theta)^3 \cdot P(X=2 \mid \theta)^3 \cdot P(X=3 \mid \theta)^2 \\ &= \left(\frac{2}{3} \cdot \theta\right)^2 \cdot \left(\frac{1}{3} \cdot \theta\right)^3 \cdot \left(\frac{2}{3} \cdot (1-\theta)\right)^3 \cdot \left(\frac{1}{3} \cdot (1-\theta)\right)^2\end{aligned}$$

since the values 0 and 3 appeared two times in the sample while values 1 and 2 appeared three times.

It's easier to work with the natural logarithm of the given expression, so we define

$$\begin{aligned} l(\theta) &= \ln(\mathcal{L}(\theta)) \\ &= 2 \cdot \ln\left(\frac{2}{3} \cdot \theta\right) + 3 \cdot \ln\left(\frac{1}{3} \cdot \theta\right) + 3 \cdot \ln\left(\frac{2}{3} \cdot (1-\theta)\right) + 2 \cdot \ln\left(\frac{1}{3} \cdot (1-\theta)\right), \end{aligned}$$

and we need to find its global maximum on the interval $[0, 1]$ (where $\theta$ can take on values). The derivative of $l$ is

$$l'(\theta) = \frac{2}{\theta} + \frac{3}{\theta} - \frac{3}{1-\theta} - \frac{2}{1-\theta} = \frac{5 - 10 \cdot \theta}{\theta \cdot (1-\theta)}.$$

So

$$l'(\theta) = 0 \Longleftrightarrow \frac{5 - 10 \cdot \theta}{\theta \cdot (1-\theta)} = 0 \Longleftrightarrow 5 - 10 \cdot \theta = 0 \Longleftrightarrow \theta = \frac{1}{2}.$$

At that point, the likelihood function reaches its local maximum, but we need the global maximum, so let's check that the likelihood function (i.e. its natural logarithm) is strictly concave on $[0, 1]$.

The second derivative of $l$ is

$$l''(\theta) = \frac{-10 \cdot \theta \cdot (1-\theta) - (5 - 10 \cdot \theta) \cdot (1 - 2 \cdot \theta)}{[\theta \cdot (1-\theta)]^2} = \frac{-5 \cdot \left(2\theta^2 - 2\theta + 1\right)}{[\theta \cdot (1-\theta)]^2}.$$

The denominator of this expression is clearly always positive, and the discriminant of the quadratic equation in the numerator is $(-2)^2 - 4 \cdot 2 \cdot 1 = -4 < 0$, so that quadratic function does not have any real roots. Since, for instance, at $\theta = \frac{1}{2}$ it equals $\frac{1}{2} > 0$, then it's always strictly positive. However, multiplying the entire fraction by -5 means that the function $l''$ is always strictly negative, which then means that $l$ is a strictly concave function.

From the above, we can conclude that the point where the likelihood reaches its global maximum, i.e. the MLE of $\theta$, is

$$\hat{\theta} = \frac{1}{2}$$

2. (a) For the method of moments, we first need to find the expected value of $X$, which is a discrete random variable with the probability distribution function as defined in the exercise.

   Using the definition of the expected value of a discrete random variable, we have

$$E(X) = \sum_{k=1}^{2} k \cdot P(X = k) = 1 \cdot \theta + 2 \cdot (1-\theta) = 2 - \theta$$

   From the above expression, we can express $\theta$ as

$$\theta = 2 - E(X).$$

   Now, the method of moments simply suggests writing the sample mean $\bar{X}$ in place of $E(X)$, and that would be the method of moments estimate of $\theta$.

   So, the desired estimate is

$$\hat{\theta} = 2 - \bar{X}$$

The sample mean can be found as

$$\bar{X} = \frac{1+2+2}{3} = \frac{5}{3},$$

which would yield a method of moment estimate of $\theta$ :

$$\hat{\theta} = 2 - \frac{5}{3} = \frac{1}{3} = 0.333.$$

(b) Let $n$ be the sample size, and let $X_1, \ldots, X_n$ be independent identically distributed random variables with the same probability distribution function (the one described in the exercise).

We define the likelihood function as:

$$\begin{aligned}\mathcal{L}(\theta) &= f(x_1, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) \cdots f(x_n \mid \theta) \\ &= P(X = x_1 \mid \theta) \cdots P(X = x_n \mid \theta)\end{aligned}$$

Substituting $n = 3$, given values of $X$, and the definition of the probability distribution function of $X$ yields

$$\mathcal{L}(\theta) = P(X = 1 \mid \theta) \cdot P(X = 2 \mid \theta)^2 = \theta \cdot (1-\theta)^2$$

since the value 1 appeared once in the sample, and value 2 appeared twice.

(c) To find the MLE of $\theta$, we need to find the point at which the likelihood function from $b$ ) part reaches its global maximum on the interval $[0,1]$ (where $\theta$ can take on values).

It's easier to work with the natural logarithm of the likelihood function, so we define

$$l(\theta) = \ln(\mathcal{L}(\theta)) = \ln(\theta) + 2 \cdot \ln(1-\theta),$$

and we need to find its global maximum on the interval $[0,1]$.

The derivative of $l$ is

$$l'(\theta) = \frac{1}{\theta} - \frac{2}{1-\theta} = \frac{1 - 3\cdot\theta}{\theta \cdot (1-\theta)}.$$

So

$$l'(\theta) = 0 \Longleftrightarrow \frac{1-3\cdot\theta}{\theta\cdot(1-\theta)} = 0 \Longleftrightarrow 1 - 3\cdot\theta = 0 \Longleftrightarrow \theta = \frac{1}{3}.$$

At that point, the likelihood function reaches its local maximum, but we need the global maximum, so let's check that the likelihood function (i.e. its natural logarithm) is strictly concave on $[0,1]$.

The second derivative of $l$ is

$$l''(\theta) = \frac{-3\cdot\theta\cdot(1-\theta) - (1-3\theta)\cdot(1-2\theta)}{[\theta\cdot(1-\theta)^2]} = \frac{-3\theta^2 + 2\theta - 1}{[\theta\cdot(1-\theta)]^2}.$$

The denominator of this expression is clearly always positive, and the discriminant of the quadratic equation in the numerator is $2^2 - 4\cdot(-3)\cdot(-1) = -8 < 0$, so that quadratic function does not have any real roots. Since, for instance, at $\theta = \frac{1}{2}$ it equals $-\frac{3}{4} < 0$, then it's always strictly negative. However, this means that $l$ is a strictly concave function.

From the above, we can conclude that the point where the likelihood reaches its global maximum, i.e. the MLE of $\theta$, is

$$\hat{\theta} = \frac{1}{3}.$$

3. (a) For the method of moments, we first need to find the expected value of $X$, which is a continuous random variable with the density function as defined in the exercise.
Using the definition of the expected value of a continuous random variable, we have

$$E(X)=\int_{-\infty}^{+\infty} x\cdot f(x)dx=\int_{-\infty}^{+\infty} \frac{1}{2\sigma}\cdot x\cdot e^{-\frac{|x|}{\sigma}}dx=0,$$

because the function under the integral is odd. This follows from the fact that the identity $x$ is an odd function, and the density function $f$ is clearly an even function, so their product must be an odd function.
So, we cannot use the first moment of $X$ to find the method of moments estimate of $\sigma$, which means that we have to proceed with higher moments. Let's find the second moment of $X$ :

$$\begin{aligned}
E\left(X^2\right)&=\int_{-\infty}^{+\infty} x^2\cdot f(x)dx=\int_{-\infty}^{+\infty}\frac{1}{2\sigma}\cdot x^2\cdot e^{-\frac{|x|}{\sigma}}dx\overset{(2)}{=}\frac{1}{\sigma}\cdot\int_0^{+\infty}x^2\cdot e^{-\frac{x}{\sigma}}dx\\
&=\sigma^2\cdot\int_0^{+\infty}y^2\cdot e^{-y}dy\quad\left[\text{use change of variables:}\ \begin{array}{cc} y=\frac{x}{\sigma} & x=0\Rightarrow y=0\\ dy=\frac{dx}{\sigma} & x=+\infty\Rightarrow y=+\infty\end{array}\right]\\
&=\sigma^2\cdot\Gamma(3)=\sigma^2\cdot 2!\\
&=2\sigma^2.
\end{aligned}$$

In (2) we used a fact that the function under the integral is even (as a product of two even functions), so its integral over the entire $\mathbb{R}$ is double the integral over $[0,+\infty)$. So, from the above expression, we can express $\sigma$ as

$$\sigma^2=\frac{1}{2}\cdot E\left(X^2\right)\Longleftrightarrow\sigma=\sqrt{\frac{1}{2}\cdot E\left(X^2\right)}.$$

Now, the method of moments simply suggests writing the sample second moment in place of $E\left(X^2\right)$, and that would be the method of moments estimate of $\sigma$.
Remember that the sample second moment is defined as

$$\frac{1}{n}\cdot\sum_{i=1}^n X_i^2,$$

which gives us our desired estimate:

$$\hat{\sigma}=\sqrt{\frac{1}{2n}\cdot\sum_{i=1}^n X_i^2}.$$

(b) Let $n$ be the sample size, and let $X_1,\ldots,X_n$ be independent identically distributed random variables with the same density function (the one described in the exercise).
To find the MLE of $\sigma$, we first define the likelihood function:

$$\mathcal{L}(\sigma)=f\left(x_1,\ldots,x_n\mid\sigma\right)=f\left(x_1\mid\sigma\right)\cdots f\left(x_n\mid\sigma\right).$$

Substituting the definition of the density function of $X$ yields

$$\mathcal{L}(\sigma)=\frac{1}{2\sigma}\cdot e^{-\frac{|x_1|}{\sigma}}\cdots\frac{1}{2\sigma}\cdot e^{-\frac{|x_n|}{\sigma}}=\frac{1}{(2\sigma)^n}\cdot e^{-\frac{|x_1|+\cdots+|x_n|}{\sigma}}.$$

It's easier to work with the natural logarithm of the given expression, so we define

$$l(\sigma) = \ln(\mathcal{L}(\sigma)) = -n \cdot \ln(2\sigma) - \frac{|x_1| + \cdots + |x_n|}{\sigma},$$

and we need to find its global maximum on the interval $\langle 0, +\infty \rangle$ (where $\sigma$ can take on values).
The derivative of $l$ is

$$l'(\sigma) = -\frac{n}{\sigma} + \frac{|x_1| + \cdots + |x_n|}{\sigma^2}$$

So

$$l'(\sigma) = 0 \Longleftrightarrow -\frac{n}{\sigma} + \frac{|x_1| + \cdots + |x_n|}{\sigma^2} = 0 \Longleftrightarrow \sigma \cdot \sum_{i=1}^{n} |x_i| = n \cdot \sigma^2$$

$$\Longleftrightarrow \sigma = 0 \quad \text{or} \quad \sigma = \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i|$$

Clearly $\hat{\sigma} = 0$ is not a good estimate, since $\sigma$ cannot be 0 (look at the definition of the density $f$ and note that $\sigma$ is in the denominator), so we have that the MLE of $\sigma$ is

$$\hat{\sigma} = \frac{1}{n} \cdot \sum_{i=1}^{n} |X_i|$$

4. (a) Let $n$ be the sample size, and let $X_1, \ldots, X_n \sim N\left(\mu, \sigma^2\right)$ be independent identically distributed random variables with the same density function.
Remember that the density function of $X \sim N\left(\mu, \sigma^2\right)$ is

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

To find the MLE of $\sigma$, we first define the likelihood function:

$$\mathcal{L}(\sigma) = f\left(x_1, \ldots, x_n \mid \sigma\right) = f\left(x_1 \mid \sigma\right) \cdots f\left(x_n \mid \sigma\right).$$

Substituting the definition of the density function of $X$ yields

$$\begin{aligned}\mathcal{L}(\sigma) &= \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right)^n \cdot e^{-\frac{(x_1-\mu)^2 + \cdots + (x_n-\mu)^2}{2\sigma^2}}\end{aligned}$$

It's easier to work with the natural logarithm of the given expression, so we define

$$l(\sigma) = \ln(\mathcal{L}(\sigma)) = n \cdot \ln\left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right) - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} \left(x_i - \mu\right)^2,$$

and we need to find its global maximum on the interval $(0, +\infty)$ (where $\sigma$ can take on values).

The derivative of $l$ (with respect to $\sigma$ ) is

$$l'(\sigma) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 .$$

So

$$\begin{aligned} l'(\sigma) = 0 &\Longleftrightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \Longleftrightarrow \sum_{i=1}^{n} (x_i - \mu)^2 = n \cdot \sigma^2 \\ &\Longleftrightarrow \sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 \Longleftrightarrow \sigma = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu)^2}. \end{aligned}$$

Therefore, the MLE for $\sigma$ is

$$\hat{\sigma} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \mu)^2}.$$

(b) Let $n$ be the sample size, and let $X_1, \ldots, X_n \sim N\left(\mu, \sigma^2\right)$ be independent identically distributed random variables with the same density function.
To find the MLE of $\mu$, we first define the likelihood function:

$$\mathcal{L}(\mu) = f\left(x_1, \ldots, x_n \mid \mu\right) = f\left(x_1 \mid \mu\right) \cdots f\left(x_n \mid \mu\right).$$

Substituting the definition of the density function of $X$ yields

$$\begin{aligned} \mathcal{L}(\mu) &= \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right)^n \cdot e^{-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}} \end{aligned}$$

It's easier to work with the natural logarithm of the given expression, so we define

$$l(\mu) = \ln(\mathcal{L}(\mu)) = n \cdot \ln\left(\frac{1}{\sigma \cdot \sqrt{2\pi}}\right) - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 ,$$

and we need to find its global maximum on $\mathbb{R}$ (where $\mu$ can take on values).
The derivative of $l$ (with respect to $\mu$ ) is

$$l'(\mu) = \frac{1}{\sigma^2} \cdot \left(\sum_{i=1}^{n} x_i - n \cdot \mu\right).$$

So

$$\begin{aligned} l'(\mu) = 0 &\Longleftrightarrow \frac{1}{\sigma^2} \cdot \left(\sum_{i=1}^{n} x_i - n \cdot \mu\right) = 0 \Longleftrightarrow \sum_{i=1}^{n} x_i - n \cdot \mu = 0 \\ &\Longleftrightarrow \mu = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i. \end{aligned}$$

Therefore, the MLE for $\mu$ is

$$\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

5. (a) Suppose $X$ is a Bernoulli random variable with parameter $p \in [0,1]$. Let

$$f(x;p) = p^x(1-p)^{1-x}.$$

Thus

$$\begin{aligned}
\log f(x;p) &= x \log p + (1-x)\log(1-p) \\
\frac{\partial \log f(x;p)}{\partial p} &= \frac{x}{p} - \frac{1-x}{1-p} \\
\frac{\partial^2 \log f(x;p)}{\partial p^2} &= -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}.
\end{aligned}$$

Clearly,

$$\begin{aligned}
I(p) &= -E\left[\frac{-X}{p^2} - \frac{1-X}{(1-p)^2}\right] \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{(1-p)} = \frac{1}{p(1-p)},
\end{aligned}$$

which is larger for $p$ values close to zero or one.

(b) Let

$$f(x;\lambda) := e^{-\lambda}\lambda^x/x!, \quad x = 0,1,\ldots.$$

Then

$$\begin{aligned}
\frac{\partial \log f(x;\lambda)}{\partial \lambda} &= \frac{\partial}{\partial \lambda}(x\log\lambda - \lambda - \log x!) \\
&= \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}.
\end{aligned}$$

Accordingly,

$$I(\lambda) = E\left[\left(\frac{\partial \log f(X;\lambda)}{\partial \lambda}\right)^2\right] = \frac{E(X-\lambda)^2}{\lambda^2} = \frac{\sigma^2}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.$$

(c) This is the beta distribution with parameters $\theta$ and 1, which we denote by beta$(\theta,1)$. The derivative of the log of $f$ is

$$\frac{\partial \log f}{\partial \theta} = \log x + \frac{1}{\theta}.$$

From this we have $\partial^2 \log f/\partial\theta^2 = -\theta^{-2}$. Hence the information is $I(\theta) = \theta^{-2}$.