Q1: The random variables $X_1 \dots, X_n$ are independent and sampled from a distribution whose probability density function, $f(x;\theta)$ is such that:

$$f(x;\theta) \propto exp(-\theta x^2)$$

where $x \geq 0$ is an unknown parameter and $f(x;\theta) = 0$ if $x < 0$.

a. Determine the full probability density function

By definition: X is a continous random variable if exist a non negative function that integrate to one on a defined support χ. This means that X have to verify these two conditions:

- $f(x;\theta) \geq 0$
- $\int_\chi f(x;\theta) = 1 \quad \forall x \in \chi$

In order to derive the full pdf, it seems reasonable to integrate $e^{-\theta x_i^2}$ on its support and then complete the pdf exploiting the fact that $\int_\chi f(x;\theta) = 1$.

$$\int_0^{+\infty} e^{-\theta x^2} dx$$

Let: $\theta x^2 = t$, $x = \sqrt{t/\theta}$, $dx = \frac{1}{2\sqrt{\theta t}}$, where $x \in [0, +\infty)$ and $t \in (0, +\infty)$. We can re-write the expression as:

$$\int_0^{+\infty} \frac{1}{2\sqrt{\theta t}} e^{-t} dt$$

rearrenging $\sqrt{t}$ as $t^{-1/2}$ we have.

$$\frac{1}{2\sqrt{\theta}} \int_0^{+\infty} t^{-1/2} e^{-t} dt$$

We can notice that under the intergal sign we have something that seems to be the well known Gamma function, that can be defined in this way

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

In addition, it is known that the gamma function with $\alpha = 1/2$ assumes a particular value, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Just leaving out the constant terms from the integral sign and recognizing the gamma function we obtain:

$$\frac{1}{2\sqrt{\theta}} \int_0^{+\infty} t^{-1/2} e^{-t} dt = \frac{1}{2\sqrt{\theta}} \sqrt{\pi}$$

We can now use this result to derive the full pdf:

$$
\int_0^{+\infty} e^{-\theta x_i^2} dx = \frac{1}{2\sqrt{\theta}}\sqrt{\pi}
$$
$$
\int_0^{+\infty} \frac{2\sqrt{\theta}}{\sqrt{\pi}} e^{-\theta x_i^2} dx = 1
$$

In which we have just divided by the terms on the right side, this last passage and recalling that the pdf of a continuous random variable have to integrato to 1, we can then write the **full pdf**.

$$
f(x;\theta) = \frac{2\sqrt{\theta}}{\sqrt{\pi}} e^{-\theta x^2}, \;\; x \geq 0, \theta > 0
$$

as the exponential function is always non negative, this expresson is non negative Thus, the conditions are fulfilled.

b. Determine the maximum likelihood estimator of θ

Given a $\mathbf{X} = (X_1, \ldots, X_n)^T$ a randomvector from $X \sim f(x_i;\theta)$ and the corresponding observed values $\mathbf{x} = (x_1, \ldots, x_n)^T$. The likelihood function can be written as the product of the marginal densities, as long as we are dealing with i.i.d. random variables.

$$
L(\theta;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta)
$$

In which we have denoted as $L(\theta;\mathbf{x})$ the likelihood function, where we have written θ first to emphasize the fact that it is a function of theta. With this method we want to find the value of θ that makes our observed data the most likely to have been observed, given the probability distribution of the sample, in a more mathematical way we are looking for:

$$
\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\theta;\mathbf{x})
$$

Where with Θ we denoted the parametric space of θ, the likelihood function can be expressed as:

$$
L(\theta;\mathbf{x}) = \prod_{i=1}^{n} \frac{2\sqrt{\theta}}{\sqrt{\pi}} e^{-\theta x_i^2} \text{ for } \theta > 0
$$
$$
L(\theta;\mathbf{x}) = \frac{2^n \theta^{n/2}}{\pi^{n/2}} e^{-\theta \sum_{i=1}^{n} x_i^2} \text{ for } \theta > 0
$$

The log-likelihood is:

$$\log L(\theta; \mathbf{x}) = n\log(2) + \frac{n}{2}\log\theta - \frac{n}{2}\log\pi - \theta\sum_{i=1}^{n} x_i^2$$

in this case the maximum can be obtained by solving:

$$\left.\frac{d\log L(\theta;\mathbf{x})}{d\theta}\right|_{\theta=\hat{\theta}} = 0$$

Then the frist derivative of the log-likelihood (the score function) is.

$$\frac{d\log L(\theta;\mathbf{x})}{d\theta} = \frac{n}{2\theta} - \sum_{=n}^{n} x_i^2$$

Setting this equation equal to zero and solving for $\hat{\theta}$ we obtain:

$$\begin{aligned}
\frac{n}{2\hat{\theta}} - \sum_{=n}^{n} x_i^2 &= 0 \\
\frac{n}{2\hat{\theta}} &= \sum_{i=1}^{n} x_i^2 \\
\hat{\theta}_{ML} &= \frac{n}{2\sum_{=n}^{n} x_i^2}
\end{aligned}$$

Thus the maximum likelihood **estimator** expressed in terms of random variables is:

$$\hat{\theta}_{ML} = \frac{n}{2\sum_{=n}^{n} X_i^2}$$

To be sure that we have obtained a maximum we have to check that the second derivative is less than zero. Thus, we have derived a maximum.

$$\frac{d^2\log L(\theta;\mathbf{x})}{d\theta^2} = -\frac{n}{2\theta^2}$$

As $n > 0$ and $\theta > 0$ the second derivative is negative.

As the likelihood is continuous in θ and that the domain of the density function $\mathbf{X}$ does not depend on θ, we can express the Fisher information $\mathbf{I}(\theta)$ in this way:

$$E\left[-\frac{d^2\log L(\theta;\mathbf{X})}{d\theta^2}\right] = E\left[\left(\frac{d\log L(\theta;\mathbf{X})}{d\theta}\right)^2\right] = \mathbf{I}(\theta)$$

Then we have that the Fisher information is:

$$E\left[-\frac{d^2 \log L(\theta;\mathbf{X})}{d\theta^2}\right] = E\left[\frac{n}{2\theta^2}\right]$$

as it is everything constant in those square brackets we have

$$= \frac{n}{2\theta^2}$$

In conclusion we can write

$$\mathbf{I}(\theta) = \frac{n}{2\theta^2}$$

c. Determine the posterior distribution of θ

First of all, we have to determine tha Jeffrey's prior, we know that:

$$\pi(\theta)^J \propto \sqrt{\mathbf{I}(\theta)}$$

This expression tells us that the Jeffrey's prior is the square root of the fisher information.

Using the Bayes theorem to derive the posterior distribution we know that it can be expressed in the following way:

$$\text{Posterior distribution} \propto \text{Prior distribution} \times \text{Likelihood}$$
$$\pi(\theta|\mathbf{x}) \propto \pi(\theta) \times L(\theta;\mathbf{x})$$

We can derive the expression of the Jeffrey's prior taking the square root of the Fisher information:

$$\pi(\theta)^J \propto \sqrt{\mathbf{I}(\theta)} = \sqrt{\frac{n}{2\theta^2}}$$
$$= \sqrt{\frac{n}{2}}\theta^{-1}$$

We can now derive the posterior distribution:

$$\pi(\theta|\mathbf{x}) \propto \pi(\theta)^J \times L(\theta;\mathbf{x})$$
$$\pi(\theta|\mathbf{x}) \propto \sqrt{\frac{n}{2}}\theta^{-1} \times \frac{2^n \theta^{n/2}}{\pi^{n/2}} e^{-\theta \sum_{i=1}^n x_i^2} \text{ for } \theta \geq 0$$

We suppose to have observed a sample of size $n = 6$, $\mathbf{x} = (1.28, 3.19, 2.53, 4.14, 2.66, 1.71)^T$.

we can compute that $\sum_i x_i^2 = 45.35$.

$$\begin{aligned}\pi(\theta|\mathbf{x}) &\propto \sqrt{\frac{6}{2}}\theta^{-1} \times \frac{2^6\theta^{6/2}}{\pi^{6/2}}e^{-\theta(45.35)}\\ &\propto \theta^{\frac{6}{2}-1}e^{-\theta(45.35)}\end{aligned}$$

We recognize the kernel of a $Gamma(\alpha = 6/2, \beta = 45.35)$.

We can conclude saying that the posterior distribution is a $Gamma(\alpha = 6/2, \beta = 45.35)$.

Q2, a. Use the Rao-Blackwell theorem to find an unbiased estimator of p.

Suppose that $X_1, \ldots, X_n$ are independent and identically distributed random variables where $X \sim Geomtreic(p)$ for $i \in 1, \ldots, n$ is proposed the following estimator:

$$\hat{p} = \begin{cases} 1 \text{ for } X_1 = 1 \\ 0 \text{ for } X_1 \neq 1 \end{cases}$$

To answer this question we are going to use the following theorem

Theorem 1. *Let $X_1, \ldots, X_n$ be random variables where the distribution of $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ is parametrized by $\theta \in \Theta$. Suppose that the statistic $S(\boldsymbol{X})$ is sufficient for θ and that $V(\boldsymbol{X})$ is an estimator of $m(\theta)$ with $E[V(\boldsymbol{X})] < \infty$ for all $\theta \in \Theta$ then the estimator*

$$T(\boldsymbol{X}) = E[V(\boldsymbol{X})|S(\boldsymbol{X})]$$

is such that $MSE(T(\boldsymbol{X}), m(\theta)) \leq MSE(V(\boldsymbol{X}), m(\theta))$ $\forall \theta \in \Theta$

The first step is to assess is $\hat{p}$ is an unbiased estimator of p or not.

$$E(\hat{p}) = 1P(X_1 = 1) + 0P(X_1 \neq 1) = p(1-p)^{1-1} = p$$

$\hat{p}$ is un unbiased estimator of p. According with the Rao-Blackwell theorem we can construct a better estimator using a sufficient statistic for θ. In other word we want to find $T(\mathbf{X}) = E[V(\mathbf{X})|S(\mathbf{X}) = s]$. However, we have to find $S(\mathbf{X})$. The next step is then to find a sufficent statistic for θ, to carry out this task we can use the factorization criterion.

Factorization Criterion Suppose that $T(\mathbf{X})$ is a statistic formed from the sample $X_1, \ldots, X_n$ where the probability distribution of $X_1, \ldots, X_n$ is parametrized by $\theta \in \Theta$.

The statistic$T(\mathbf{X})$ is sufficient for θ if and only if the joint density of $f(\mathbf{x}, \theta)$ can be expressed as

$$f(\mathbf{x}, \theta) = g(T(\mathbf{X}), \theta)h(\mathbf{x})$$

where $g()$ and $h()$ are two non-negative functions.

Then, we compute the likelihood function of a geometric distribution with the hope to factorize the likelihood function in two non negative functions and find a sufficient statistic for θ.

Take $X_1, \ldots, X_n$ rv's i.i.d. with a geometric distribution, the likelihood function is:

$$
\begin{aligned}
L(p;\mathbf{x}) &= \prod_{i=1}^{n} p(1-p)^{x_i - 1} \\
&= p^n (1-p)^{\sum_i x_i - n}
\end{aligned}
$$

Using the factorization criterion we can factorize the likelihood in two non negative functions: $g(T(\mathbf{X}),\theta) = p^n(1-p)^{\sum_i X_i - n}$ and $h(\mathbf{x}) = 1$.

It follows immediately that $T(\mathbf{X}) = \sum_i^n X_i$ is a sufficient statistic for theta.

We have to prove the distribution of $T(\mathbf{X}) = \sum_i^n X_i$, we can use the moment generating function.

$$
\begin{aligned}
m_T(t) =& E[e^{Tt}] \\
=& E[e^{X_1 t}, \ldots, e^{X_1 t}] \\
& \text{as the random variables are independent, we have} \\
& = E[e^{X_1 t}] \ldots E[e^{X_n t}] \\
& \text{using the fact that are identically distributed we have} \\
& = E[e^{X_i t}]^n
\end{aligned}
$$

We can notice that the mgf of a geometric distribution is $m_X(t) = \frac{pe^t}{1-qe^t}$ where $q = 1-p$, it follows immediately that

$$
m_T(t) = E\left[e^{X_i t}\right]^n = \left(\frac{pe^t}{1-qe^t}\right)^n
$$

Which is the mgf of a negative binomial distribution. We can conclude that

$$
\sum_i^n X_i \sim NegBinomial(n,p)
$$

We can now use the Rao-blackwell theorem to find $E[\hat{p}|S(\mathbf{X}) = s]$, where $S(\mathbf{X}) \equiv T(\mathbf{X})$.

$$
\begin{aligned}
E[\hat{p}|S(\mathbf{X}) = s] &= 1P(X_1 = 1|S(\mathbf{X}) = s) \text{ for the Bayes' theorem} && (1) \\
&= \frac{P(X_1 = 1 \cap S(\mathbf{X}) = s)}{P(S(\mathbf{X}) = s)} && (2)
\end{aligned}
$$

We have just prooved the distribution of the denominator,we have now work on the numerator. To find a solution, we can think about that probability in terms of events: We know that the event $\{X_1 = 1; \sum_{i=1}^n X_i = s\}$ is equal to

$\{X_1 = 1; \sum_{i=2}^{n} X_i = s-1\}$.

We denote with $W = \sum_{i=2}^{n}$ a new random variable that has a negative binomial distribution with parameters $(n-1, p)$.

It is clear that as X_1 is out of the summation, we can argue that W and X_1 are independent, then we can write, knowing that:
$W \sim NegBin(n-1, p)$.

$$E[\hat{p}|S(\mathbf{X}) = s] = \frac{P(X_1 = 1)P((W(\mathbf{X}) = s-1)}{P(S(\mathbf{X}) = s)}$$

replacing the expressions we have

$$= \frac{p\binom{s-2}{n-2}p^{n-1}(1-p)^{s-n}}{\binom{s-1}{n-1}p^n(1-p)^{s-n}}$$

$$= \frac{p\binom{s-2}{n-2}p^{n-1}\cancel{(1-p)^{s-n}}}{\binom{s-1}{n-1}p^n\cancel{(1-p)^{s-n}}}$$

$$= \frac{\cancel{p}\binom{s-2}{n-2}\cancel{p^n}}{\cancel{p}\binom{s-1}{n-1}\cancel{p^n}}$$

$$= \frac{\frac{(s-2)!}{(n-2)!(s-2-n+2)!}}{\frac{(s-1)!}{(n-1)!(s-1-n+1)!}}$$

$$= \frac{\cancel{(s-2)!}}{(s-1)\cancel{(s-2)!}} \frac{(n-1)\cancel{(n-2)!}}{\cancel{(n-2)!}}$$

$$= \frac{n-1}{s-1}$$

Finally, the estimator obtained using the Rao-Blackwell theorem is:

$$\tilde{p} = \frac{n-1}{S-1}$$

Where S is a negative binomial with parameters (n, p).

b. Compute the $Var(\tilde{p})$ and hence assess whether or not $Var(\tilde{p})$ attains the Cramér-Rao lower bound for the variance of unbiased estimotors of p

The first thing that we want to assess is, if exists an unbiased estimator that attains the CR Lower bound. We can start from the Geometric distribution.

Given a $\mathbf{X} = (X_1, \ldots, X_n)^T \sim \textit{Geometric}(p)$ rv's i.i.d. supposing to have observed a sample $\mathbf{x} = (x_1, \ldots, x_n)$ we can write the likelihood function:

$$
\begin{aligned}
L(p;\mathbf{x}) &= \prod_{i=1}^{n} p(1-p)^{x_i-1} \\
&= p^n(1-p)^{\sum_i x_i - n}
\end{aligned}
$$

The log-Likleihood is:

$$
LogL(p,\mathbf{x}) = n\log p + \left(\sum_i x_i - n\right)\log 1 - p
$$

The maximum can be found solving

$$
\left.\frac{d\log L(\theta;\mathbf{x})}{d\theta}\right|_{\theta=\hat{\theta}} = 0
$$

Deriving respect to p the log-likelihood we obtain:

$$
\begin{aligned}
\frac{dLogL(\theta,\mathbf{x})}{d\theta} &= \frac{n}{p} - \frac{(\sum_i x_i - n)}{1-p} \\
&\frac{n}{\hat{p}} - \frac{(\sum_i x_i - n)}{1-\hat{p}} = 0 \\
&= \frac{(1-\hat{p})n - \hat{p}(\sum_i x_i - n)}{\hat{p}(1-\hat{p})} = 0 \\
&\frac{n \cancel{-pn} - p\sum_i x_i + \cancel{pn}}{p\cancel{(1-p)}} = 0 \\
&n - \hat{\sum}_i x_i = 0 \\
&\hat{p} = \frac{n}{\sum_i x_i}
\end{aligned}
$$

The ML estimator is $\hat{p} = \frac{1}{X}$, is

In order to verify that we have found a maximum, we have to check the sign of the second derivative. The score function is

$$
\frac{dLogL(\theta,\mathbf{x})}{d\theta} = \frac{n}{p} - \frac{(\sum_i x_i - n)}{1-p}
$$

The secod derivarive of the log-likelihood is:

$$\begin{aligned}\frac{d^2 LogL(\theta, \mathbf{x})}{d\theta^2} &= -\frac{n}{p^2} - \frac{(\sum_i x_i - n)}{(1-p)^2} \\ &= -\left[\frac{n}{p^2}\frac{(\sum_i x - n)}{(1-p)^2}\right] < 0\end{aligned}$$

Note how, the sign of the second derivative is determined by the negative sign in front of the parenthesis. All the variables inside the parenthesis are positive. We conclude that the estimator we found is in correspondence of a maximum

We can notice that the support of X does not depend on θ then:

$$E\left[-\frac{d^2 \log L(\theta; \mathbf{X})}{d\theta^2}\right] = E\left[\left(\frac{d \log L(\theta; \mathbf{X})}{d\theta}\right)^2\right] = \mathbf{I}(\theta)$$

It follows that:

$$E\left[-\frac{d^2 \log L(\theta; \mathbf{X})}{d\theta^2}\right] = E\left[\frac{n}{p^2} + \frac{(\sum_i X_i - n)}{(1-p)^2}\right]$$

Usign the linearity property of the expected value we have:

$$E\left[-\frac{d^2 \log L(\theta; \mathbf{X})}{d\theta^2}\right] = \left[\frac{n}{p^2} + \frac{(E(\sum_i X_i) - n)}{(1-p)^2}\right]$$

Using the fact that $X_i \sim Geometric(p)$ and the fact that are i.i.d.

$$\begin{aligned}&= \frac{n}{p^2} + \frac{n/p - n}{(1-p)^2} \\ &= \frac{n}{p^2} + \frac{n - pn}{p(1-p)^2} \\ &= \frac{n}{p^2} + \frac{n\cancel{(1-p)}}{p(1-p)^{\cancel{2}}} \\ &= \frac{n}{p^2} + \frac{n}{p(1-p)} \\ &= \frac{n}{p^2(1-p)}\end{aligned}$$

Cramer- Rao inequity:

If $X_1, \dots, X_n$ is a random sample generated by $X \sim f(x;\theta)$, under regularity conditions, for every ***unbiased*** *estimator T_n of θ we have that:*

$$Var(T_n) \geq \mathbf{I}(\theta)^{-1}$$

This means that if we guess that $\hat{p}$ is unbiased for p, we should have that

$$Var(\hat{p}) = \mathbf{I}(\theta)^{-1} = \frac{p^2(1-p)}{n}$$

To verify if this expression is the right one, we're going to use the following result:

The variance of $T(\boldsymbol{X})$ attains the Cramer- Rao lower bound if and only if $T(\boldsymbol{X})$ and $U(\theta, \boldsymbol{X})$ are linearly related

$$U(\theta, \mathbf{X}) = \mathbf{I}(\theta)(T\mathbf{X} - \theta)$$

We recall that the score equation that we have just derived is

$$\begin{aligned}
U(\theta, \mathbf{X}) &= \frac{n}{p} - \frac{(\sum_i X_i - n)}{(1-p)} \\
&= \frac{n(1-p) + p(\sum_i X_i - n)}{p(1-p)} \\
&= \frac{n - \cancel{np} - p\sum_i X_i + \cancel{np}}{p(1-p)} \\
&= \frac{1}{p(1-p)}\left[n - p\sum_i X_i\right] \\
&= \frac{\sum_i X_i}{p(1-p)}\left[\frac{n}{\sum_i X_i} - p\right]
\end{aligned}$$

From the last expression it is clear that the maximum likelihood estimator does not reach the CRLB, Since we cannot express the score as a linear combination of the fischer's information. We conclude that there is no expression for the Cramer Rao lower bound for this estimator.

We can confirm this result showing that the expected value of the ML estimator is different from p

$$E[\hat{p}_{ML}] = E\left[\frac{n}{\sum_i X}\right] \neq \frac{n}{E[\sum X_i]}$$

Thanks to the linearity property of the expectation we have

$$\frac{n}{E[\sum X_i]} = \frac{n}{\sum E[X_i]}$$

using that X_i are i.i.d. random variables we can write

$$= \frac{n}{n(1/p)} = p$$

To double check that the ML estimator for p is biased and then it cannot reach the CRLB, we can use the Jensen's inequality for convex functions. As is well known $1/x$ is a convex function, then.

$$E\left[\frac{n}{\sum_i X}\right] \geq \frac{n}{E[\sum X_i]} = p$$

We can conclude by saying that since the ML estimator does not reach the lower limit of cramer rao, we cannot compare $Var(\tilde{p})$ with the CRLB.

c. is $\tilde{p}$ the minimum variance unbiased estimator of p ?

To verify that $\tilde{p}$ is a MVUE, the first passage is to assess if is an unbiased estimator.

$$\begin{aligned}
E[\tilde{p}] &= E\left[\frac{n-1}{s-1}\right] \\
&= \sum_{s=n}^{\infty} \binom{s-1}{n-1} \frac{n-1}{s-1} p^n (1-p)^{n-r} \\
&= \sum_{s=n}^{\infty} \frac{\cancel{(s-1)}(s-2)!}{\cancel{(n-1)}(n-2)!(s-n)!} \frac{\cancel{n-1}}{\cancel{s-1}} p^n (1-p)^{s-n} \\
&= \sum_{s=n-1}^{\infty} \frac{(s+1-2)!}{(n-2)!(s+1-n)!} p^n (1-p)^{s+1-n} \\
&= \sum_{s=n-1}^{\infty} \frac{(s-1)!}{(n-2)!(s+1-n)!} p^{n+1-1} (1-p)^{s+1-n} \\
&= p \underbrace{\sum_{s=n-1}^{\infty} \frac{(s-1)!}{(n-2)!(s+1-n)!} p^{n-1} (1-p)^{s+1-n}}_{1} \\
&= p
\end{aligned}$$

We recognize a $NegBin(n-1,p)$ that by definition have to sum to 1. We can conclude that $\tilde{p}$ is an unbaised estimator of p.

To verify that $\tilde{p}$ is a MVUE we can use the following result.

Lehmann- Scheffé Theorem: *If $T(\boldsymbol{X})$ is a complete sufficient statistic for θ and $g(T(\boldsymbol{X}))$ is some function of $T(\boldsymbol{X})$ such that $E(g(T(\boldsymbol{X}))) = m(\theta)$, then*

$g(T(\boldsymbol{X}))$ *is the unique minimum variance unbiased estimator for* $m(\theta)$.

Using the Factorization criterio, we have already proved that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is a sufficent statistic for p, it is remained to prove the completeness.

We can recall the following important result about the completeness: *If a statistic has distribution that belongs to the exponential family of distributions then the statistics is complete with respect to the unknown distributional parameter.*

To verify that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is complete, we have to assess if it has a distribution that belongs to the exponential family, we know that $\sum_{i=1}^{n} X_i \sim NegBin(n,p)$. Thus, we can write the joint distribution for i.i.d. rv's as

$$\begin{aligned}
P(\mathbf{X}=\mathbf{x}) &= \prod_{i=1}^{m} \binom{x-1}{n-1} p^n (1-p)^{x-n} \\
&= p^{mn}(1-p)^{\sum_{i=1}^{m}(x_i-n)} \prod_{i=1}^{m} \binom{x-1}{n-1} \\
&\text{using exp() and log() we can write} \\
&= exp\{mn\log(p) + \sum_{i=1}^{m}(x_i-n)\log(1-p) + \log\binom{x-1}{n-1}\} \\
&= exp\{mn\log(p) + \sum_{i=1}^{m} x_i \log(1-p) - mn\log(1-p) + \log\binom{x-1}{n-1}\} \\
&= exp\{\sum_{i=1}^{m} x_i \log(1-p) + mn\log\left(\frac{p}{1-p}\right) + \log\binom{x-1}{n-1}\}
\end{aligned}$$

With:

$$\begin{aligned}
b(\theta) &= n\log\left(\frac{p}{1-p}\right) \\
a(\theta) &= \log(1-p) \\
T(\mathbf{X}) &= \sum X_i \\
c(\mathbf{x}) &= \log\binom{x-1}{n-1}
\end{aligned}$$

Thus, $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is a complete and sufficient statistic. Using the Lehman- Sheffè theorem we can conclude saying that as $T = \frac{n-1}{X-1}$ is a function of a sufficient and complete statistic, as it has been proved, $T = \frac{n-1}{X-1}$ is the unique minimum variance unbiased estimator for p .

Q3. The random variables $X_1, \ldots, X_n$ are independent and sampled from a distribution with probability density function:

$$\begin{cases} \frac{\lambda\alpha^\lambda}{x^{\lambda+1}} & \text{if } x \geq \alpha \\ 0 & \text{otherwise} \end{cases}$$

Where $\lambda > 1$ is known but $\alpha > 0$ is an unknown parameter.

a). Find the ML estimator of α and call it $\tilde{\alpha}$.

First of all, let's write down the likelihood function, using the fact that all $X_1, \ldots, X_n$ have the same distribution and are independent, we know that under these conditions the joint density can be expressed as the product of the marginals. Let's guess to have observed $\mathbf{x}$ a realization of the random vector $\mathbf{X}$ we can write the likelihood function as:

$$\begin{aligned} L(\alpha, \lambda; \mathbf{x}) &= \prod_{i=1}^{n} \frac{\lambda\alpha^\lambda}{x_i^{\lambda+1}} \mathbb{1}_{(0<\alpha\leq x_i)} \\ &= \frac{\lambda^n \alpha^{n\lambda}}{\prod_i x_i^{\lambda+1}} \mathbb{1}_{(0<\alpha\leq x_i)} \end{aligned}$$

We have used the $\mathbb{1}_{(0<\alpha\leq x_i)}$ to stress that the likelihood is a function of α as we are considering λ as known. And that the likelihood is valid only if $\alpha \in (0, x_i]$. If we consider $x_{(1)} = min(x_1, \ldots, x_n)$ we can rewrite the likelihood function in this way:

$$L(\alpha, \lambda; \mathbf{x}) = \frac{\lambda^n \alpha^{n\lambda}}{\prod_i x_i^{\lambda+1}} \mathbb{1}_{(0<\alpha\leq x_{(1)})}$$

To find the ML estimator, we can not compute the derivative and set it equal to zero, we have to think carefully about the value of α that maximixes that likelihood function. This function is monotonic, as α increases the function grows. Said that, it seems reasonable to choice the greatest possible value that α can assume, in this case is $x_{(1)}$. The maximum likelihood estimator for α in terms of random variables is $X_{(1)}$.

$$\tilde{\alpha} = X_{(1)}$$

b) Is $\tilde{\alpha}$ unbiased for α ?

It is obvious that $\tilde{\alpha}$ is the maximum of the values that α can take, this suggests that $\tilde{\alpha}$ may be a biased estimator of the parameter. However, to anwer this question, we're going to compute the $E[\tilde{\alpha}] = E[X_{(1)}]$. First of all we have to find the distribution of $X_{(1)}$. Let V the minimum, $V > v$ if and only if $X_i > v$ $\forall i$.

$$
\begin{aligned}
1 - F_V(v) =& P_V(V \geq v) \\
&= P(x_1 \geq v), \ldots, P(X_n \geq v) \\
\text{Using the fact that are i.i.d.} \\
&= [1 - F(v)]^n \\
1 - F_V(v) =& = [1 - F(v)]^n \\
F_V(v) =& 1 - [1 - F(v)]^n
\end{aligned}
$$

We have to differentiate to obtain the pdf.

$$
\frac{dF_V(v)}{dv} = f_V(v) = nf(v)[1 - F(v)]^{n-1}
$$

to derive this expression we need the pmf of X.

$$
\begin{aligned}
F(x_o) &= P(X \leq x_o) \\
&= \int_\alpha^{x_o} \frac{\lambda\alpha^\lambda}{x^{\lambda+1}} dx \\
&= \lambda\alpha \int_\alpha^{x_o} \frac{1}{x^{\lambda+1}} dx \\
&= \lambda\alpha \int_\alpha^{x_o} x^{-(\lambda+1)} dx \\
&= \lambda\alpha \left(\frac{x^{-(\lambda+1)+1}}{-(\lambda+1)+1} \right)\Bigg|_\alpha^{x_o} \\
&= \lambda\alpha \left(\frac{x^{-\lambda}}{-\lambda} \right)\Bigg|_\alpha^{x_o} \\
&= -\frac{\cancel{\lambda}}{\cancel{\lambda}} \alpha^\lambda \left[x_o^{-\lambda} - \alpha^{-\lambda} \right] \\
&= 1 - \frac{\alpha^\lambda}{x_o^\lambda}
\end{aligned}
$$

To double check that the last expression is really the pmf, we have to derive respect to x_o.

$$
\begin{aligned}
\frac{\partial F(x_o)}{\partial x_o} &= 0 - (-\lambda) \frac{\alpha^\lambda}{x^{\lambda+1}} \\
&= \frac{\lambda\alpha^\lambda}{x^{\lambda+1}}
\end{aligned}
$$

We can now define the pdf of $X_{(1)}$ as:
$f_{X_{(1)}}(x_{(1)}) = nf_X(x)[1 - F_X(x)]^{n-1}$.

$$\begin{aligned}
f_{X_{(1)}} &= \frac{n\lambda\alpha^\lambda}{x^{\lambda+1}}\left[1 - \left(1 - \frac{\alpha^\lambda}{x^\lambda}\right)\right]^{n-1} \\
&= \frac{n\lambda\alpha^\lambda}{x^{\lambda+1}}\left[\frac{\alpha^\lambda}{x^\lambda}\right]^{n-1} \\
&= \frac{n\lambda\alpha^\lambda}{x^{\lambda+1}}\frac{\alpha^{\lambda(n-1)}}{x^{\lambda(n-1)}} \\
&= \frac{n\lambda\cancel{\alpha^\lambda}\alpha^{n\lambda}\cancel{\alpha^{-\lambda}}}{\cancel{x^\lambda}xx^{n\lambda}\cancel{x^{-\lambda}}} \\
&= \frac{n\lambda\alpha^{n\lambda}}{x^{n\lambda+1}}
\end{aligned}$$

Thus, after some semplifications we end up with the pdf of $X_{(1)}$, we can immediately notice that this pdf is equal to the pdf of $\mathbf{X}$ with different parameters, instead of λ we have $n\lambda$. Now we can compute the $E[\tilde{\alpha}] = E[X_{(1)}]$, and then assess if $\tilde{\alpha}$ is an biased estimator of α or not.

$$\begin{aligned}
E[X_{(1)}] &= \int_\alpha^{+\infty} x\frac{n\lambda\alpha^{n\lambda}}{x^{n\lambda+1}}dx \\
&= n\lambda\alpha^{n\lambda}\int_\alpha^{+\infty} \cancel{x}\frac{1}{x^{n\lambda+\cancel{1}}}dx \\
&= \int_\alpha^{+\infty} \frac{n\lambda\alpha^{n\lambda}}{x^{n\lambda}}dx \\
&= \int_\alpha^{+\infty} \frac{n\lambda\alpha^{n\lambda+1-1}}{x^{n\lambda}}dx
\end{aligned}$$

we add and subtract 1 from the exponent of α

$$= \alpha\int_\alpha^{+\infty} \frac{n\lambda\alpha^{n\lambda-1}}{x^{n\lambda}}dx$$

we can now divide and multiply by $(n\lambda - 1)$ to obtain

$$= \frac{\alpha n\lambda}{(n\lambda - 1)}\underbrace{\int_\alpha^{+\infty} \frac{(n\lambda-1)\alpha^{n\lambda-1}}{x^{n\lambda}}dx}_{1}$$

when we recognize the pdf of the min. with parameters $((n\lambda - 1), \alpha)$.

$$E[X_{(1)}] = \frac{\alpha n\lambda}{(n\lambda - 1)}$$

We then can conclude that $\tilde{\alpha}$ is a biased estimator of α.

c) A Bayesian statistician assumes that the prior distribution for

α is given by $\alpha \sim U[0, \alpha]$, $\alpha \in R^+$. Is asked to determine the posterior distribution of α.

We know that using the Baye's theorem we can define the posterior distribution as:

Posterior distribution $\propto$ Prior distribution $\times$ Likelihood.

$$\pi(\alpha|\mathbf{x}) \propto \frac{1}{a}\mathbb{1}_{(0\leq\alpha\leq a)} \frac{\lambda^n \alpha^{n\lambda}}{\prod_i x_i^{\lambda+1}} \mathbb{1}_{(0<\alpha\leq x_{(1)})}$$

We indicate as $\Phi = min(x_{(1)}, a)$. Then we can write:

$$\pi(\alpha|\mathbf{x}) \propto \frac{1}{a} \frac{\lambda^n \alpha^{n\lambda}}{\prod_i x_i^{\lambda+1}} \mathbb{1}_{(0<\alpha\leq\Phi)}$$

We recall that to write down the posterior distribution we have to find the normilizing constant, the idea is to integrate on the entire support of this r.v., let the result equal to one and divide by the normilizing constant. We omit the indicator function in order to simplify the expressions.

$$\begin{aligned}
\int_0^{\Phi} \pi(\theta|\mathbf{x}) d\alpha &= \int_0^{\Phi} \frac{1}{a} \frac{\lambda^n \alpha^{n\lambda}}{\prod_i x_i^{\lambda+1}} \; d\alpha \\
&= \frac{\lambda^n}{a \prod_i x_i^{\lambda+1}} \int_0^{\Phi} \alpha^{n\lambda} \; d\alpha \\
&= \frac{\lambda^n}{a \prod_i x_i^{\lambda+1}} \left[\left(\frac{\alpha^{n\lambda+1}}{n\lambda+1} \right) \Bigg|_0^{\Phi} \right] \\
&= \frac{\lambda^n}{a \prod_i x_i^{\lambda+1}} \frac{\Phi^{n\lambda+1}}{n\lambda+1}
\end{aligned}$$

We recall that a pdf to be a valid pdf have to satisfy $\int_{\mathcal{X}} f(\mathbf{x}, \theta) d\mathbf{x} = 1$, and be non negative. This last expression is the normilizing constant, now to obtain the posterior distribution we have just to divide the proportional posterior by the constant to obtain the exact formulation of the posterior.

$$\pi(\alpha|\mathbf{x}) = \frac{\frac{1}{a}\frac{\lambda^n \alpha^{n\lambda}}{\prod_i x_i^{\lambda+1}}}{\frac{\lambda^n}{a\prod_i x_i^{\lambda+1}}\frac{\Phi^{n\lambda+1}}{n\lambda+1}}$$

$$= \frac{1}{\cancel{a}} \frac{\cancel{\lambda^n}\alpha^{n\lambda}}{\cancel{\prod_i x_i^{\lambda+1}}} \times \frac{\cancel{a}\cancel{\prod_i x_i^{\lambda+1}}(n\lambda+1)}{\cancel{\lambda^n}\Phi^{n\lambda+1}}$$

$$= \frac{\alpha^{n\lambda}(n\lambda+1)}{\Phi^{n\lambda+1}}$$

$$\pi(\alpha|\mathbf{x}) = \frac{\alpha^{n\lambda}(n\lambda+1)}{\Phi^{n\lambda+1}}\mathbb{1}_{(0<\alpha\leq\Phi)}$$

This is the posterior distribution of α.

d) Assume that $\lambda = 4$ and $n = 8$ and $\mathbf{x} = (14.8, 7.2, 5.9, 5.4, 6.1, 5.2, 9.4, 5.5)$ and a prior of α such that $\alpha \sim U[0, 6]$ determine the Bayes' estimate of α where the loss function is given by:

$$L(\alpha, d) = (d - \alpha)^2$$

We want find the value of d that minimizes

$$E_{\alpha|\mathbf{x}}[(\alpha - d)^2]$$

Expanding the square of this expected value

$$E_{\alpha|\mathbf{x}}[(\alpha - d)^2] = E_{\alpha|\mathbf{x}}[(\alpha^2 - 2d\alpha + d^2)] \tag{3}$$

using the linearity property of the expectation (4)

$$= E_{\alpha|\mathbf{x}}[\alpha^2] - 2dE_{\alpha|\mathbf{x}}[\alpha] + d^2 \tag{5}$$

We know that $E_{\alpha|\mathbf{x}}[\alpha^2] = Var_{\alpha|\mathbf{x}}(\alpha) + \{E_{\alpha|\mathbf{x}}[\alpha]\}^2$ using this expression in (5) we have

$$E_{\alpha|\mathbf{x}}[(\alpha - d)^2] = Var_{\alpha|\mathbf{x}}(\alpha) + \{E_{\alpha|\mathbf{x}}[\alpha]\}^2 - 2dE_{\alpha|\mathbf{x}}[\alpha] + d^2$$
$$= Var_{\alpha|\mathbf{x}}(\alpha) + \big(E_{\alpha|\mathbf{x}}[\alpha] - d\big)^2$$

Thus, the last expression in minimized when $d = E_{\alpha|\mathbf{x}}[\alpha]$, the posterior mean.

We have seen that the value that minimizes this loss function is the posterior mean $E_{\alpha|\mathbf{x}}[\alpha]$. In this case $min(x_{(1)}), 6)$ is $x_1 = 5.2$. We can then compute the posterior mean.

$$
\begin{aligned}
E_{\alpha|\mathbf{x}}[\alpha] &= \int_0^{x_{(1)}} \alpha \frac{\alpha^{n\lambda}(n\lambda+1)}{x_{(1)}^{n\lambda+1}} \, d\alpha \\
&= \frac{(n\lambda+1)}{x_{(1)}^{n\lambda+1}} \int_0^{x_{(1)}} \alpha^{n\lambda+1} \, d\alpha \\
&= \frac{(n\lambda+1)}{x_{(1)}^{n\lambda+1}} \frac{x_{(1)}^{n\lambda+2}}{n\lambda+2} \\
&= \frac{(n\lambda+1)}{(n\lambda+2)} x_{(1)}
\end{aligned}
$$

plugging in the values we have:

$$
E_{\alpha|\mathbf{x}}[\alpha] = \frac{4*8+1}{4*8+2} 5.2 = 5.04706
$$

The value that minimizes that loss function is 5.04706.