QFIN Statistics 2 Unit 4 Team 6

Bazaluk_Va, Heldenberg, Kotzamanis, Lussetti, Schmidhammer Steger

29/03/2023

Exercise 61

a)

The given exercise involves i.i.d random variables $X_1, \ldots, X_n$ that follow a uniform distribution on the interval $[0, \theta]$, where θ is a fixed value. The parameter Θ, which follows a Pareto distribution, is also a random variable. Part (a) demonstrates that the posterior distribution is also a Pareto distribution, and provides the expressions for the marginal, joint, and prior densities as follows:

$$fX_i|\Theta(x_i|\theta) = \begin{cases} 1/\theta & x_i \in [0, \theta] \\ 0 & \text{else} \end{cases}, \quad f_{X|\Theta} = \theta^{-n} I_{(x \leq \theta)}, \quad f_\Theta(\theta) = \begin{cases} \alpha\beta^\alpha\theta^{-(\alpha+1)} & \theta \geq \beta \\ 0 & \text{else} \end{cases}$$

The reason for using the indicator function is to ensure that the x_i values do not exceed the given θ. Despite the fact that θ is not known, it is necessary to impose this restriction. As a result, the posterior distribution will be derived as follows:

$$f_{\Theta|X}(\theta|x) = \frac{f_{X|\theta}(x|\theta) f_\Theta(\theta)}{\int f_{X|\theta}(x|\theta) f_\Theta(\theta)} \propto \theta^{-n} I_{(x \leq \theta)} \alpha\beta^\alpha\theta^{-(\alpha+1)} I_{(\theta \geq \beta)} \propto \theta^{-(n+\alpha+1)} I_{max(\beta,x) \leq \theta}$$

Hence, the posterior also corresponds to a Pareto distribution:

$$\Theta|X \sim Pareto(\alpha + n, max(\beta, x_n))$$

We get $Pareto(12, 14)$.

b)

The posterior mean will be calculated as the expectation:

$$\int_{\hat{\beta}}^{\infty} \theta\hat{\alpha}\hat{\beta}^{\hat{\alpha}}\theta^{-(\hat{\alpha}+1)} d\theta = \hat{\alpha}\hat{\beta}^{\hat{\alpha}} \int_{\hat{\beta}}^{\infty} \theta^{-\hat{\alpha}} d\theta = \frac{\hat{\alpha}\hat{\beta}}{\hat{\alpha} - 1}$$

, with $\hat{\alpha} > 1$. Hence we get for our case:

$$\hat{\theta}_{Post} = \frac{(\alpha + n)max(\beta, x_n)}{\alpha + n - 1}, \quad \text{with } \alpha + n > 1$$

$$\Rightarrow \hat{\theta}_{Post} = \frac{12 * 14}{11} = 15.2727$$

c)

It is known that the Pareto density function is unimodal and decreasing, which implies that the highest posterior density (HPD) interval for the parameter θ will be an interval that spans from the lower to the upper bound of the posterior density. The upper bound of the interval can be determined by taking the maximum of β and x_n, and the desired probability mass in the interval will also be taken into account. Therefore, we can equate the cumulative distribution function (CDF) of the Pareto distribution with the desired level (e.g., 95) to obtain the interval limits.

$$\int_{\widetilde{\beta}}^{q} \widetilde{\alpha}\widetilde{\beta}^{\widetilde{\alpha}}\theta^{-(\widetilde{\alpha}+1)}d\theta = 1-\alpha \Rightarrow \widetilde{\alpha}\widetilde{\beta}^{\widetilde{\alpha}}\left[-\frac{1}{\widetilde{\alpha}}\theta^{-\widetilde{\alpha}}\right]_{\widetilde{\beta}}^{q} = 1-\alpha \Rightarrow 1-\frac{\widetilde{\beta}^{\widetilde{\alpha}}}{q^{\widetilde{\alpha}}} = 1-\alpha$$

$$\frac{\widetilde{\beta}^{\widetilde{\alpha}}}{q^{\widetilde{\alpha}}} = \alpha \Rightarrow q = \frac{\widetilde{\beta}}{\widetilde{\alpha}^{1/\widetilde{\alpha}}} \Rightarrow HPD = \left[\widetilde{\beta}, \frac{\widetilde{\beta}}{\widetilde{\alpha}^{1/\widetilde{\alpha}}}\right]$$

With our data set (and 95% interval) we obtain: $HPD = [14, 17.97]$

d)

We will compute the probability of the events $0 \leq \theta \leq 15$ and $\theta \geq 15$ by integrating the posterior over the corresponding intervals:

$$P(0 \leq \theta \leq 15) = \int_{14}^{15} 12 * 14^{12}\theta^{-(12+1)}d\theta = 12 * 14^{12} * \left(-\frac{1}{12}15^{-12} + \frac{1}{12}14^{-12}\right) = 1 - \left(\frac{14}{15}\right)^{12} = 0.56304$$

Since the problem requires computing two complementary probabilities, we can infer that the posterior probability of the null hypothesis being true is equal to 0.56304, while the alternative hypothesis has a posterior probability of 0.43696. As a result, we can conclude that the null hypothesis is more likely to be true a posteriori.

Exercise 62

Let $X_1, ..., X_n$ be an i.i.d. sample from the geometric distribution. We know that for a given value of p, the variables are geometrically distributed with exactly the parameter $p \in [0, 1]$. This unknown parameter is a random variable itself, denoted by P, which follows a Beta distribution with parameters α and β. We say $X = (X_1, ..., X_n)$ and $x = (x_1, ..., x_n)$, we can write the marginal and joint densities of the data for a given value of P = p then the marginal distribution (defined also as prior distribution) is defined as followed:

$$f(x_i|p) = p(1-p)^x, x_i$$

Beta distribution (source: wikipedia) is defined as: $f(x; \alpha, \beta) = constant \cdot x^{\alpha-1}(1-x)^{\beta-1}$ where constant=

$$\frac{1}{B(\alpha, \beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

Given that X values are i.i.d, then joint distribution is the product of marginal distribution:

$$lik(p|x_1, x_2, ...x_n) = \prod_{i=1}^{n} f(x_i|p) = p^n(1-p)^{\sum_{i=1}^{n} x_i}$$

From slide 6 lecture 3, we are given the definition for posterior distribution where

$$f_{P|X}(p|x) \propto f(x|p)f_P(p)$$

$$constant \cdot \propto p^n(1-p)^{\sum_{i=1}^n x_i} p^{\alpha-1}(1-p)^{\beta-1} = constant \cdot p^{\alpha+n-1}(1-p)^{\beta+\sum_{i=1}^n x_i - 1}$$

$$\implies P|X \sim Beta(\alpha+n, \beta + \sum_{i=1}^n x_i)$$

Hence we proved the Beta is conjugate to geometric data.

In the Bayesian paradigm, all information about P is contained in the posterior, and we can estimate the paramter by the *posterior mean* and *posterior mode* of this distribution. Beta distribution with parameters α and β, by definition (source:Wikipedia) has the following mean:

$$E(P) = \frac{\alpha}{\alpha+\beta}$$

Hence, the posterior Bayes estimate of p, denoted

$$\hat{p}_{POST} = \frac{\alpha+n}{\alpha+n+\beta+\sum_{i=1}^n x_i}$$

part b)

Fisher information $I(p)$ is defined as \$I(p) =E((l'(p))^2)=-E(l' '(p)). \$ Geometrically distributed random variable X can be differentiated twice:

$$l(p) = log(p) + xlog(1-p)$$

$$l'(p) = \frac{1}{p} - \frac{x}{1-p}$$

$$l''(p) = -\frac{1}{p^2} - \frac{x}{(1-p)^2}$$

We know that the expectation of a geometric r.v. X is given by $E(X|p) = \frac{1-p}{p}$. Using this result in the calculation approach via the second derivative, we obtain by making use of the linearity of the expectation operator:

$$I(p) = -E(-\frac{1}{p^2} - \frac{x}{(1-p)^2}) = \frac{1}{p^2} + \frac{1}{(1-p)^2}E(X) = \frac{1}{p^2} + \frac{1}{(1-p)^2} \cdot \frac{(1-p)}{p} = \frac{1}{p^2}(1 + \frac{p}{(1-p)} = p^{-2}(1-p)^{-1}$$

Therefore the Jeffreys prior is:

$$\pi(p) \propto \sqrt{p^{-2}(1-p)^{-1}} = p^{-1}(1-p)^{-\frac{1}{2}}$$

This is an improper prior, since it corresponds to a Beta distribution with parameters $\alpha = 0$, $\beta = 0.5$. This doesn't integrate to 1. That's why the Beta distribution is only defined for $\alpha > 0$. It delivers a proper posterior, but we can derive the posterior distribution in this case as: (based on above result) $P|X \sim$ Beta $(n, \frac{1}{2} + \sum_{i=1}^n x_i)$

Consequently, the posterior mean is given by:

$$\hat{p}_{POST} = \frac{n}{n+0.5+\sum_{i=1}^n x_i}$$

Exercise 63

We suppose that X is a discrete r.v. with

$$\mathbb{P}(X=0) = 2\theta/3, \qquad \mathbb{P}(X=1) = \theta/3, \qquad \mathbb{P}(X=2) = 2(1-\theta)/3, \qquad \mathbb{P}(X=3) = (1-\theta)/3$$

We also have 10 independent observations: (3, 0, 2, 1, 3, 2, 1, 0, 2, 1).

(a) Find the method of moments estimate of θ

To apply the method of moments we have to follow the steps below: 1) We derive the expectation of X as a function of the parameter θ:

$$\mathbb{E}(X) = 0 \cdot \frac{2\theta}{3} + 1 \cdot \frac{\theta}{3} + 2 \cdot \frac{2(1-\theta)}{3} + 3 \cdot \frac{1-\theta}{3} = -2\theta + \frac{7}{3}$$

2) We now invert this function to find θ as a function of the first moment:

$$\hat{\theta}_{MOM} = -\frac{1}{2}(\bar{x} - \frac{7}{3})$$

3) We derive the mean for the sample using the observations:

$$\bar{x} = \frac{3+0+2+1+3+2+1+0+2+1}{10} = 1.5$$

4) We find the MOM estimate for θ:

$$\hat{\theta}_{MOM} = -\frac{1}{2}(1.5 - \frac{7}{3}) = 0.4167$$

(b) Find an approximate standard error of your estimate

In this part of the task we can use the CLT and the fact that:

$$\frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{n\to\infty} N(0,1)$$

Now we need to find the second moment of the random vatiable:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$\mathbb{E}(X^2) = 0^2 \cdot \frac{2\theta}{3} + 1^2 \cdot \frac{\theta}{3} + 2^2 \cdot \frac{2(1-\theta)}{3} + 3^2 \cdot \frac{1-\theta}{3} = -\frac{16}{3}\theta + \frac{17}{3}$$

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = -\frac{16}{3}\theta + \frac{17}{3} - (-2\theta + \frac{7}{3})^2 = -4\theta^2 + 4\theta + \frac{2}{9}$$

Now we can see that:

$$\bar{x} \sim N(-2\theta + \frac{7}{3}, \frac{-4\theta^2 + 4\theta + \frac{2}{9}}{n})$$

But we also know the relationship between the theta MOM estimate and the mean, which can help us see how the theta bar is distributed and, consequently, its SE estimate:

$$\hat{\theta}_{MOM} = -\frac{1}{2}(\bar{x} - \frac{7}{3})$$

$$\hat{\theta}_{MOM} \sim N(\theta, \frac{1}{4}\frac{-4\theta^2 + 4\theta + 2/9}{n})$$

$$\hat{SE}(\hat{\theta}_{MOM}) = \sqrt{\frac{1}{4}\frac{-4\hat{\theta}^2{}_{MOM} + 4\hat{\theta}_{MOM} + 2/9}{n}} = \sqrt{\frac{1}{4}\frac{-4 \cdot (5/12)^2 + 4 \cdot (5/12) + 2/9}{10}} = \sqrt{\frac{43}{1440}} \sim 0.1728$$

(c) What is the maximum likelihood estimate of θ?

We know that each observation happened in the sample s times. As the observations are independent:

$$lik(\theta) = \left(\frac{2\theta}{3}\right)^{s_0} \cdot \left(\frac{\theta}{3}\right)^{s_1} \cdot \left(\frac{2(1-\theta)}{3}\right)^{s_2} \cdot \left(\frac{1-\theta}{3}\right)^{s_3}$$

$$\ell(\theta) = s_0 log\left(\frac{2\theta}{3}\right) + s_1 log\left(\frac{\theta}{3}\right) + s_2 log\left(\frac{2(1-\theta)}{3}\right) + s_3 log\left(\frac{1-\theta}{3}\right)$$

Now we differentiate and set equal to 0:

$$\ell'(\theta) = s_0\frac{3}{\theta} + s_1\frac{3}{\theta} + s_2\frac{3}{1-\theta} + s_3\frac{3}{1-\theta} = 0 \implies \theta = \frac{s_0+s_1}{s_0+s_1+s_2+s_3} = \frac{s_0+s_1}{n}$$

We remember from the sample that $s_0 = 2, s_1 = 3$ and n=10. This way, we get:

$$\hat{\theta}_{MLE} = \frac{5}{10} = 0.5$$

(d) What is an approximate standard error of the MLE?

In this part of the task we can apply Fisher information to find the assymptotic variance of an MLE of theta parameter. We know that:

$$I(\theta) = \mathbb{E}((\ell'(\theta))^2) = -\mathbb{E}(\ell''(\theta))$$

We already know the ℓ' and we have to differentiate it again to derive Fisher information:

$$\ell'(\theta) = s_0\frac{3}{\theta} + s_1\frac{3}{\theta} + s_2\frac{3}{1-\theta} + s_3\frac{3}{1-\theta}$$

$$\ell''(\theta) = -\frac{3}{\theta^2}(s_0+s_1) - \frac{3}{(1-\theta)^2}(s_2+s_3)$$

$$I(\theta) = -\mathbb{E}\left(-\frac{3}{\theta^2}(s_0+s_1) - \frac{3}{(1-\theta)^2}(s_2+s_3)\right) = \frac{3}{\theta^2}(\mathbb{E}(s_0)+\mathbb{E}(s_1)) + \frac{3}{(1-\theta)^2}(\mathbb{E}(s_2)+\mathbb{E}(s_3))$$

Now we somehow have to fill in the expectation values for all s. For this, we assume that $S_i \sim B(n, \mathbb{P}(X=i))$, as each s means how many time each observation appears in the sample. Bearing this in mind, we can proceed:

$$I(\theta) = \frac{3}{\theta^2}(\mathbb{E}(s_0)+\mathbb{E}(s_1)) + \frac{3}{(1-\theta)^2}(\mathbb{E}(s_2)+\mathbb{E}(s_3)) =$$

$$= \frac{3}{\theta^2}\left(n\frac{2\theta}{3} + n\frac{\theta}{3}\right) + \frac{3}{(1-\theta)^2}\left(n\frac{2(1-\theta)}{3} + n\frac{1-\theta}{3}\right) = 3\frac{n}{\theta} + 3\frac{n}{1-\theta} = \frac{3n}{\theta(1-\theta)}$$

Now we can proceed to finding an approximate standard error of the MLE:

$$\hat{SE}(\hat{\theta}_{MLE}) = \frac{1}{\sqrt{nI(\hat{\theta}_{MLE})}} = \frac{1}{\sqrt{n \cdot \frac{3n}{\hat{\theta}_{MLE}(1-\hat{\theta}_{MLE})}}} =$$

$$= \frac{\sqrt{\hat{\theta}_{MLE}(1-\hat{\theta}_{MLE})}}{\sqrt{3}n} = \frac{\sqrt{0.5\cdot 0.5}}{\sqrt{3}\cdot 10} = 0.0289$$

(e) If the prior distribution of Θ is uniform on [0,1], what is the posterior density? Plot it. What is the mode of the posterior?

To find the posterior density for the parameter Θ, we apply the Bayesian framework:

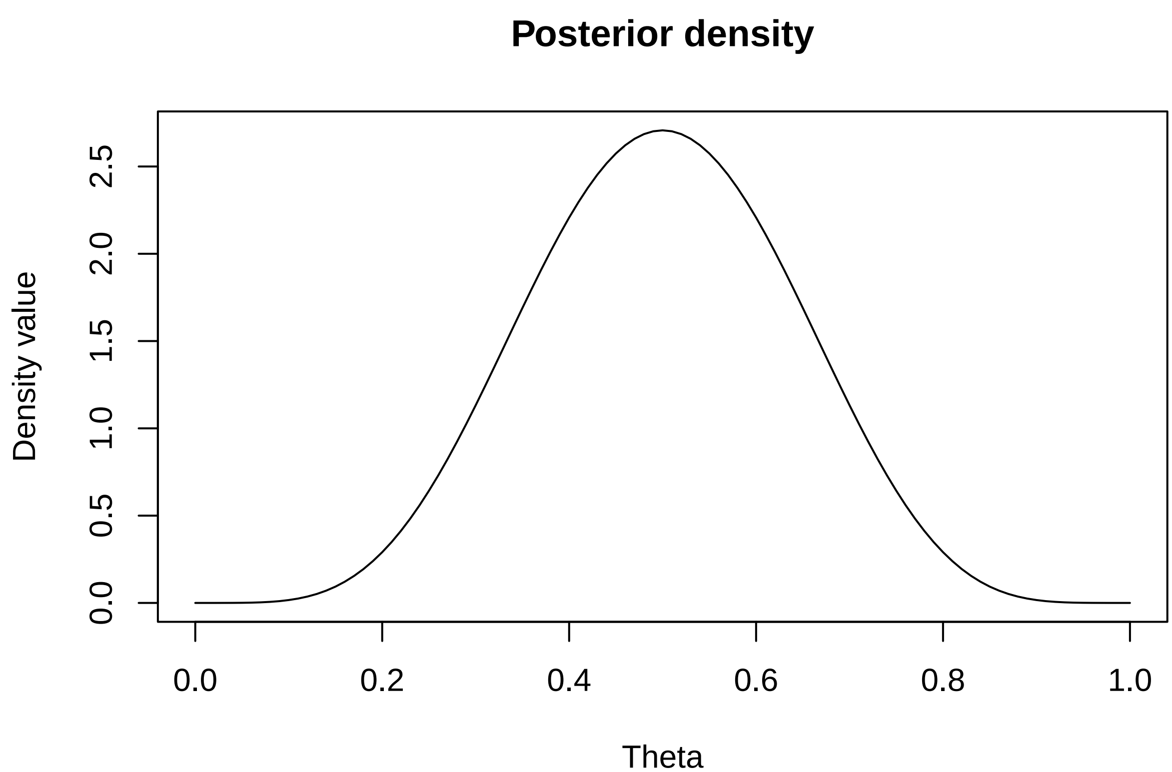
$$\Theta \sim \mathbb{U}[0,1] \implies f_\Theta(\theta) = 1$$
$$f_{X|\Theta}(x|\theta) = \left(\frac{2\theta}{3}\right)^{s_0} \cdot \left(\frac{\theta}{3}\right)^{s_1} \cdot \left(\frac{2(1-\theta)}{3}\right)^{s_2} \cdot \left(\frac{1-\theta}{3}\right)^{s_3}$$
$$f_{\Theta|X}(\theta|x) = \frac{f(x|\theta)f_\Theta(\theta)}{\int_0^1 f(x|\theta)f_\Theta(\theta)d\theta} = \frac{3^{-n}2^{s_0+s_2}\theta^{s_0+s_1}(1-\theta)^{s_2+s_3}}{3^{-n}2^{s_0+s_2}\int_0^1 \theta^{s_0+s_1}(1-\theta)^{s_2+s_3}} = \frac{\theta^{s_0+s_1}(1-\theta)^{s_2+s_3}}{\int_0^1 \theta^{s_0+s_1}(1-\theta)^{s_2+s_3}}$$

We introduce two parameters-helpers: $a = s_0 + s_1 + 1$ and $b = s_2 + s_3 + 1$. The posterior density is found as follows:

$$f_{\Theta|X}(\theta|x) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)} \implies \Theta|X \sim Beta(a,b)$$

We know that $a = s_0 + s_1 + 1 = 6$ and $b = s_2 + s_3 + 1 = 6$ and now we can plot the posterior density:

```
plotpostden<-function(x) dbeta(x,6,6)
curve(plotpostden,0,1,xlab="Theta",ylab="Density value",main="Posterior density")
```



We see that the MLE and the mode equal both to 0.5.

Exercise 65

$X_1, ..., X_n$ are i.i.d. $N(\mu, \sigma^2)$

Let $\theta = (\mu, \sigma)$.

a)

If μ is known, what is the MLE of σ?

We use the likelihood function and we maximize σ.

$$
\begin{aligned}
lik(\theta) &= f(x_1, \ldots, x_n | \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\
\ell(\theta) &= -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \\
\frac{\partial \ell(\theta)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^{n}(x_i - \mu)^2 = 0 \implies \\
\hat{\sigma} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2}
\end{aligned}
$$

b)

If σ is known, what is the MLE of μ?

We use the same likelihood function, but this time we maximize μ instead of σ.

$$
\begin{aligned}
lik(\theta) &= f(x_1, \ldots, x_n | \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\
\ell(\theta) &= -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \\
\frac{\partial \ell(\theta)}{\partial \mu} &= \frac{1}{\sigma^2} \cdot \left(\sum_{i=1}^{n} x_i - n \cdot \mu\right) = 0 \implies \\
\hat{\mu} &= \frac{1}{n} \cdot \sum_{i=1}^{n} X_i
\end{aligned}
$$

c)

In the case above, where σ is knows, does any other unbiased estimate of μ have smaller variance?

First, we need to calculate the Fisher information:

$$
\begin{aligned}
I(\mu) &= -\mathbb{E}\left(\frac{\partial^2 \ell(\mu)}{\partial^2 \mu}\right) \\
\log f(x|\mu) &= \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x-\mu)^2}{2\sigma^2} \\
\frac{\partial^2 \log f(x|\mu)}{\partial^2 \mu} &= -\frac{1}{\sigma^2} \\
I(\mu) &= \mathbb{E}\left(\frac{1}{\sigma^2}\right) = \frac{1}{\sigma^2}
\end{aligned}
$$

Therefore, the Cramer-Rao lower bound is:

[$\frac{1}{n \cdot I(\mu)} = \frac{\sigma^2}{n}$.]

This is also the variance of $\bar{X}$, which is the MLE for μ.
This means that there exist no unbiased estimator for μ with lower variance than the variance of $\hat{\mu}_{MLE}$.

Exercise 67

a)

$$
Y_i = \beta x_i + \epsilon_i, \qquad i = 1, \ldots, n, \qquad \epsilon_i \overset{iid}{\sim} N(0, \sigma^2)
$$

Above we have a simple linear regression model without an intercept, which tells us that $\mathbb{E}(Y_i|\beta x_i) = \beta x_i$, but also, when reorganizing also implies that $\epsilon_i = Y_i - \beta x_i$, which we can use to derive a density function for an individual Y_i:

$$
f(y_i|\beta x_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(y_i - \beta x_i)^2}{2\sigma^2}\right)
$$

Taking the product now (iid) and applying the logarithm to it we acquire the likelihood and the log-likelihood function:

$$
\begin{aligned}
lik(\beta, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{(y_i - \beta x_i)^2}{2\sigma^2}\right) \\
\ell(\beta, \sigma) &= -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta x_i)^2
\end{aligned}
$$

Now to obtain the MLE for β, we have to differentiate the log-likelihood with respect to β and set the equation to zero:

$$
\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= -\frac{2}{2\sigma^2}\sum_{i=1}^{n}(-x_i)(y_i - \beta x_i) = 0 \\
\beta \sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i y_i \\
\hat{\beta}_{MLE} &= \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}
\end{aligned}
$$

To figure out if β is unbiased, we need to compute the expectation of our $\hat{\beta}_{MLE}$

$$\begin{aligned}
\mathbb{E}(\hat{\beta}_{MLE}) &= \mathbb{E}\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}\right) = \mathbb{E}\left(\frac{\sum_{i=1}^n x_i(\beta x_i + \epsilon_i)}{\sum_{i=1}^n x_i^2}\right) \\
&= \mathbb{E}\left(\frac{\beta\sum_{i=1}^n x_i^2 + \sum_{i=1}^n \epsilon_i x_i}{\sum_{i=1}^n x_i^2}\right) = \mathbb{E}\left(\frac{\beta\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2}\right) + \mathbb{E}\left(\frac{\sum_{i=1}^n \epsilon_i x_i}{\sum_{i=1}^n x_i^2}\right) \\
&= \mathbb{E}(\beta) + \mathbb{E}\left(\frac{\sum_{i=1}^n \epsilon_i x_i}{\sum_{i=1}^n x_i^2}\right) = \beta
\end{aligned}$$

In the beginning we swapped y_i for the corresponding equation $(\beta x_i + \epsilon_i)$. The last step holds as well since $\mathbb{E}(\epsilon_i) = 0 \forall i$. We can conclude from that, that $\hat{\beta}_{MLE}$ is indeed unbiased.

Since it is unbiased, the MSE is given by the variance of the estimate (bias=0):

$$\begin{aligned}
MSE(\hat{\beta}_{MLE}) &= \mathbb{V}(\hat{\beta}_{MLE}) \\
&= \mathbb{V}\left(\sum_{i=1}^n \left(\frac{x_i}{\sum_{i=1}^n x_i^2}\right) y_i\right) = \sum_{i=1}^n \frac{x_i^2}{(\sum_{i=1}^n x_i^2)^2}\mathbb{V}(y_i) = \frac{\sum_{i=1}^n x_i^2}{(\sum_{i=1}^n x_i^2)^2}\sigma^2 \\
&= \frac{\sigma^2}{\sum_{i=1}^n x_i^2}
\end{aligned}$$

b)

First, we recall the log-likelihood function from part 1:

$$\ell(\beta, \sigma) = -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2$$

Now we need to differentiate the function, like we did in 1), just with respect to σ and set the expression equal to zero:

$$\begin{aligned}
\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (y_i - \beta x_i)^2 = 0 \\
\Rightarrow n\sigma^2 = \sum_{i=1}^n (y_i - \beta x_i)^2 \\
\hat{\sigma}_{MLE} = \sqrt{\frac{1}{n}\sum_{i=1}^n (y_i - \beta x_i)^2}
\end{aligned}$$

c)

Again, we recall the log-likelihood function from part 1 and expand it as follows:

$$\begin{aligned}
\ell(\beta, \sigma) &= -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2 \\
&= -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2 + \frac{\beta}{\sigma^2}\sum_{i=1}^n x_i y_i - \frac{1}{2}\left(\frac{\beta}{\sigma}\right)^2\sum_{i=1}^n x_i^2
\end{aligned}$$

Now let $\theta = \frac{\beta}{\sigma}$ and differentiate the log-likelihood function by it and set the equation to zero:

$$\begin{aligned}
\ell(\beta, \sigma, \theta) &= -n\log(\sigma) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2 + \frac{\theta}{\sigma}\sum_{i=1}^{n} x_i y_i - \frac{\theta^2}{2}\sum_{i=1}^{n} x_i^2 \\
\frac{\partial \ell}{\partial \theta} &= \frac{\sum_{i=1}^{n} x_i y_i}{\sigma} - \theta \sum_{i=1}^{n} x_i^2 = 0 \\
\hat{\theta}_{MLE} &= \frac{\sum_{i=1}^{n} x_i y_i}{\hat{\sigma}_{MLE}\sum_{i=1}^{n} x_i^2} = \frac{\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \beta x_i)^2}} \\
&= \frac{\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} x_i)^2}}
\end{aligned}$$

d)

To derive the Fisher information matrix $I(\beta, \sigma)$, we first need to derive the necessary second order derivatives:

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2 \\
\frac{\partial^2 \ell}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n} y_i^2 + \frac{6\beta}{\sigma^4}\sum_{i=1}^{n} x_i y_i - \frac{3\beta^2}{\sigma^4}\sum_{i=1}^{n} x_i^2 \\
\frac{\partial^2 \ell}{\partial \beta \partial \sigma} &= -\frac{2\beta}{\sigma^3}\sum_{i=1}^{n} x_i^2 + \frac{2\beta}{\sigma^3}\sum_{i=1}^{n} x_i^2
\end{aligned}$$

Now, we need to find the elements of the Fisher matrix:

$$\begin{aligned}
I(\beta) &= -\mathbb{E}\left(-\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2\right) = \frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} \\
I(\sigma) &= -\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \sigma^2}\right) = -\frac{n}{\sigma^2} + \frac{3}{\sigma^4}\left(n\sigma^2 + \beta^2\sum_{i=1}^{n} x_i^2\right) - \frac{6\beta}{\sigma^4}\left(\beta\sum_{i=1}^{n} x_i^2\right) + \frac{3\beta^2}{\sigma^4}\sum_{i=1}^{n} x_i^2 \\
&= \frac{2n}{\sigma^2} \\
I(\beta, \sigma) = I(\sigma, \beta) &= -\mathbb{E}\left(-\frac{2\beta}{\sigma^3}\sum_{i=1}^{n} x_i^2 + \frac{2\beta}{\sigma^3}\sum_{i=1}^{n} x_i^2\right) = 0
\end{aligned}$$

Hence, we obtain the Fisher information matrix as:

$$I(\beta, \sigma) = \begin{bmatrix} \frac{\sum_{i=1}^{n} x_i^2}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

e)

We use the hint and denote the parameter vector $\theta = (\beta, \sigma)$. For an unbiased estimator T of $g(\theta) = \beta/\sigma$, we obtain under regularity conditions:

$$var(T) \geq (\nabla g(\theta))' I(\theta)^{-1} \nabla g(\theta)$$

The gradient of $g(\theta)$ is given by $\nabla g(\theta)' = \left(\frac{1}{\sigma}, -\frac{\beta}{\sigma^2}\right)$ and the Cramer-Rao lower bound for T is

$$var(T) \geq \begin{pmatrix} \frac{1}{\sigma} & -\frac{\beta}{\sigma^2} \end{pmatrix} \begin{bmatrix} \frac{\sigma^2}{\sum_{i=1}^n x_i^2} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix} \begin{pmatrix} \frac{1}{\sigma} \\ -\frac{\beta}{\sigma^2} \end{pmatrix} = \frac{1}{\sum_{i=1}^n x_i^2} + \frac{1}{2n}\left(\frac{\beta}{\sigma}\right)^2$$

Exercise 68

a)

T is considered sufficient for the parameter λ if the conditional distribution of $X_1, \ldots, X_n$ given $T = t$ is independent of λ. In the case of $X_1, \ldots, X_n$ being Poisson-distributed with parameter λ, we need to verify whether this property holds for $T = \sum_{i=1}^n X_i$. Therefore, the conditional distribution can be expressed as follows:

$$\mathbb{P}\left(X_1, \ldots, X_n \mid T = t\right) = \frac{\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n, T = t\right)}{\mathbb{P}(T = t)}$$

Since $T = \sum_{i=1}^n X_i$, T has Poisson distribution $T \sim$ Poisson $(n\lambda)$ and the marginal destribution is:

$$\mathbb{P}(T = t) = e^{-n\lambda} \frac{(n\lambda)^t}{t!}$$

The joint distribution:

$$\begin{aligned} \mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n, T = t\right) &= \begin{cases} \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}, & \text{if } \sum_{i=1}^n x_i = T \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} e^{-n\lambda} \frac{\lambda^T}{x_1!, \ldots, x_n!}, & \text{if } \sum_{i=1}^n x_i = T \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

We can define the conditional distribution of $X_1, \ldots, X_n$ given $T = t$:

$$\begin{aligned} \mathbb{P}\left(X_1, \ldots, X_n \mid T = t\right) &= \frac{\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n, T = t\right)}{\mathbb{P}(T = t)} \\ &= \frac{e^{-n\lambda}\lambda^T}{x_1!, \ldots, x_n!} \times \frac{T!}{e^{-n\lambda}(n\lambda)^T} \\ &= \frac{T!}{n^T x_1!, \ldots, x_n!} \end{aligned}$$

We observe that $\mathbb{P}\left(X_1, \ldots, X_n \mid T = t\right)$ does not depend on $\lambda \Rightarrow T\left(X_1, \ldots, X_n\right) = \sum_{i=1}^n X_i$ is a sufficient statistics for λ

b)

If X_1 is not sufficient statistics for λ, then the conditional distribution of $X_1, \ldots, X_n$ given $X_1 = x_1$ will depend on λ.

$$\begin{aligned}
\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n \mid X_1 = x_1\right) &= \frac{\mathbb{P}\left(X_1 = x_1, \ldots, X_n = x_n, X_1 = x_1\right)}{\mathbb{P}\left(X_1 = x_1\right)} \\
&= \frac{\prod_{i=1}^{n} \mathbb{P}\left(X_i = x_i\right)}{\mathbb{P}\left(X_1 = x_1\right)} \\
&= \prod_{i=2}^{n} \mathbb{P}\left(X_i = x_i\right) \\
&= \frac{e^{-(n-1)\lambda} \times \lambda^{\sum_{i=1}^{n} x_i}}{x_1!, \ldots, x_n!}
\end{aligned}$$

Apparently, the conditional distribution depends on $\lambda \Rightarrow X_1$ is not sufficient statistics for λ.

c)

The factorization theorem asserts that for $T(X_1, \ldots, X_n)$ to be sufficient for the parameter λ, it is both necessary and sufficient that the joint distribution of $X_1, \ldots, X_n$ can be factorized in the form:

$$f\left(x_1, \ldots, x_n \mid \lambda\right) = g\left[T\left(x_1, \ldots, x_n\right), \lambda\right] h\left(x_1, \ldots, x_n\right)$$

The density function is:

$$f\left(x_1, \ldots, x_n \mid \lambda\right) = \prod_{i=2}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \times \lambda^{\sum_{i=1}^{n} x_i} \times \prod_{i=1}^{n} \frac{1}{x_i!}$$

Factorize $f\left(x_1, \ldots, x_n \mid \lambda\right)$ as follows: $g\left[T\left(x_1, \ldots, x_n\right), \lambda\right] = e^{-n\lambda}\lambda^{T}$, which depends on T and λ $h\left(x_1, \ldots, x_n\right) = \prod_{i=1}^{n} \frac{1}{x_i!}$, which depends only on $x_1, \ldots, x_n$ As factorization is possible, $T\left(x_1, \ldots, x_n\right)$ is sufficient statistics for λ.

Exercise 69

Given random variables $X : X_1, X_2, ..., X_n$ and the joint density $f(x_1, \ldots, x_n; \theta)$, the statistics $Y(u(x_1, x_2, ..., X_n))$ is sufficient for parameter θ iff PDF can be factorised as following:

$$g(t(x_1, \ldots, x_n); \theta_n) h(x_1, \ldots, x_n) = f(x_1, \ldots, x_n | \theta_n) \ [Factorization\ theorem]$$

The joint distribution for i.i.d $X_i \sim Gamma(\alpha, \beta)$ is defined as:

$$t(x_1, \ldots, x_n | \alpha, \beta) = \prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i}$$

We want $f(x_1, \ldots, x_n) = F(\sum_{i=1}^{n} x_i, \prod_{i=1}^{n} x_i)$, plugging f and t in factorization theorem formula;

$$g(F(\sum_{i=1}^{n} x_i, \prod_{i=1}^{n} x_i), \alpha, \beta) h(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} * x_i^{\alpha-1} * e^{-\beta x_i}$$

So $h(x_1, \ldots, x_n) = 1$,

$g(F(\sum_{i=1}^{n} x_i, \prod_{i=1}^{n} x_i), \alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} * e^{-\beta \sum_{i=1}^{n} x_i} * (\prod_{i=1}^{n} x_i)^{\alpha-1}$.

$g(T(x_1, \ldots, x_n), \alpha, \beta)$ depends on $T(x_1, \ldots, x_n) = (\sum_{i=1}^{n} x_i, \prod_{i=1}^{n} x_i)$. Thus $\sum_{i=1}^{n} x_i$ and $\prod_{i=1}^{n} x_i$ are sufficient for (α, β).

Exercise 76

(a) Compute the likelihood ratio Λ for each possible of X value and order the x_i according to Λ.

The computations for the likelihood ratios look the following way:

$$LR_1 = \frac{\mathbb{P}(x_1|H_0)}{\mathbb{P}(x_1|H_A)} = \frac{0.2}{0.1} = 2$$

$$LR_2 = \frac{\mathbb{P}(x_2|H_0)}{\mathbb{P}(x_2|H_A)} = \frac{0.3}{0.4} = 0.75$$

$$LR_3 = \frac{\mathbb{P}(x_3|H_0)}{\mathbb{P}(x_3|H_A)} = \frac{0.3}{0.1} = 3$$

$$LR_4 = \frac{\mathbb{P}(x_4|H_0)}{\mathbb{P}(x_4|H_A)} = \frac{0.2}{0.4} = 0.5$$

Now we order x_i according to Λ:

$$1)x_3 : LR_3 = 3$$

$$2)x_1 : LR_1 = 2$$

$$3)x_2 : LR_2 = 3/4$$

$$4)x_4 : LR_4 = 1/2$$

(b) What is the likelihood ratio test of H_0 versus H_A at level $\alpha = .2$? What is the test at level $\alpha = .5$?

The formula that we use here for the probability of the type I error is:

$$\alpha = \mathbb{P}(acceptH_1|H_0) = \mathbb{P}(rejectH_0|H_0) = \mathbb{P}(LR < LR(x_i)|H_0)$$

Now we gather all the information we have:

X	H_0	LR	$\mathbb{P}(LR < LR(x_i)\|H_0)$
x_3	0.3	3	0.7
x_1	0.2	2	0.5
x_2	0.3	0.75	0.2
x_4	0.2	0.5	0

First, we solve for $\alpha = 0.2$:

$$\alpha = \mathbb{P}(acceptH_1|H_0) = \mathbb{P}(rejectH_0|H_0) = \mathbb{P}(LR < LR(x_i)|H_0) = \mathbb{P}(X = x_4|H_0) = 0.2$$

The acceptance region for the null hypothesis is $\{x_3, x_1, x_2\}$.

Next, we solve for $\alpha = 0.5$:

$$\alpha = \mathbb{P}(acceptH_1|H_0) = \mathbb{P}(rejectH_0|H_0) = \mathbb{P}(LR < LR(x_i)|H_0) = \mathbb{P}(X = x_4, X = x_2|H_0) = 0.5$$

The acceptance region for the null hypothesis is $\{x_3, x_1\}$.

(c) If the prior probabilities are $\mathbb{P}(H_0) = \mathbb{P}(H_1)$, which outcomes favor H_0?

The condition for accepting the hull hypothesis is:

$$\frac{\mathbb{P}(H_0|x)}{\mathbb{P}(H_1|x)} > 1$$

We have to check for which realizations the condition is met:

$$\frac{\mathbb{P}(H_0|x)}{\mathbb{P}(H_1|x)} = \frac{\mathbb{P}(x|H_0)\mathbb{P}(H_0)}{\mathbb{P}(x|H_1)\mathbb{P}(H_1)} \implies \frac{\mathbb{P}(x|H_0)}{\mathbb{P}(x|H_1)} > 1$$

The outcomes for which this holds are x_1 and x_3. For those, $\mathbb{P}(x|H_0) > \mathbb{P}(x|H_1)$.

(d) What prior probabilities correspond to the decision rules with $\alpha = .2$ and $\alpha = .5$?

We first find the prior probabilities for $\alpha = 0.2$ and then for $\alpha = 0.5$.

$\alpha = 0.2$: The condition for the acceptance of the null hypothesis is (acceptance region starts with x_2):

$$\frac{\mathbb{P}(H_0|x)}{\mathbb{P}(H_1|x)} = \frac{\mathbb{P}(x|H_0)\mathbb{P}(H_0)}{\mathbb{P}(x|H_1)\mathbb{P}(H_1)} > 1$$

Knowing the acceptance region we can compute:

$$\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} \cdot \frac{3}{4} > 1 \implies \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} > \frac{4}{3}$$

We also know the condition for the rejection of the null hypothesis:

$$\frac{\mathbb{P}(x|H_0)\mathbb{P}(H_0)}{\mathbb{P}(x|H_1)\mathbb{P}(H_1)} \leq 1$$

$$\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} \cdot \frac{1}{2} \leq 1 \implies \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} \leq 2$$

The final interval for $\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$ is $\left(\frac{4}{3}, 2\right]$

$\alpha = 0.5$: The condition for the acceptance of the null hypothesis is (acceptance region is $\{x_3, x_1\}$):

$$\frac{\mathbb{P}(H_0|x)}{\mathbb{P}(H_1|x)} = \frac{\mathbb{P}(x|H_0)\mathbb{P}(H_0)}{\mathbb{P}(x|H_1)\mathbb{P}(H_1)} > 1$$

Knowing the acceptance region we can compute:

$$\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} \cdot 2 > 1 \implies \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} > \frac{1}{2}$$

We also know the condition for the rejection of the null hypothesis:

$$\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} \cdot \frac{3}{4} \leq 1 \implies \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} \leq \frac{4}{3}$$

The final interval for $\frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)}$ is $\left(\frac{1}{2}, \frac{4}{3}\right]$

Exercise 77

a) FALSE

STATMENT: The significance level of a statistical test is equal to the probability that the null hypothesis H_0 is true.

CORRECTION:
The significance level of a statistical test is the conditional probability of rejecting the null hypothesis H_0, when the null hypothesis is actually true, which is a type I error.

b) FALSE

STATEMENT: If the significance level of a test is decreased, the power would be expected to increase.

CORRECTION:
If the significance level of a test is decreased, the power would be expected to decrease as well. The test rejects the null hypothesis H_0 with lower probability, under the null hypothesis H_0. The power of a test is 1 minus the probability of rejecting the null hypothesis H_0, under the conditional probability for a given alternate hypothesis H_A, meaning that the power should decrease, since the test rejects less often, if it has a lower level.

c) FALSE

STATEMENT: If a test is rejected at the significance level α, the probability that the null hypothesis is true equals α.

CORRECTION:
As stated in *a*): The significance level is the probability that the null hypothesis H_0 is rejected, when it is true and not the probability that the null hypothesis is true.We do not evaluate the probability of the null hypothesis H_0 being true with the significance level α.

d) FALSE

STATEMENT: The probability that the null hypothesis is falsely rejected is equal to the power of the test.

CORRECTION:
The probability that the null hypothesis is falsely rejected is not equal to the power of the test. The power of the test is the probability that the null hypothesis H_0 is correctly rejected.

e) FALSE

STATEMENT: A type I error occurs when the test statistic falls in the rejection region of the test.

CORRECTION:
It depends. When the test statistic falls in the rejection region of the test, the test procedure rejects the null hypothesis. This may lead to a right decision, if the null hypothesis is false, or to a wrong decision (type I error), if the null hypothesis is true. We do not know which, just from the rejection region of the test.

f) FALSE

STATEMENT: A type II error is more serious than a type I error.

CORRECTION:
Type I and type II errors are both serious, but type I error is more serious.In Neyman-Pearson testing, there is a size limit of the type I error to find the test that minimizes the type II error subject to the constraint.Since the size of the type I error is controlled, this is a reason to say that type I has priority.

g) FALSE

STATEMENT: The power of a test is determined by the null distribution of the test statistic.

CORRECTION:
The power of a test is not determined by the null distribution of the test statistic, since it is the conditional probability of correctly rejecting the null hypothesis H_0, when the alternate hypothesis H_A is true. The null hypothesis H_0 is the probability distribution of the test statistic, when the null hypothesis H_0 is true. The power of a test is a function, which is computed for every distribution in the alternate hypothesis H_A .

h) TRUE

STATEMENT: The likelihood ratio is a random variable.

It is the ratio of the conditional pmf/pdf of the data (X) under H_0 and H_A and is a function of the data which is a random variable.

Exercise 78

We have to consider the introductory coin tossing example from the lecture and suppose that instead of tossing the coin 10 times, the coin was tossed until a head came up and the total number of tosses X was recorded.

From the lecture we know that coin 0 has a probability of heads equal to 0.5 , while coin 1 has a probability of heads equal to 0.7 .

The example goes as follows: Someone chooses one of the coins, tosses it until heads comes up and tells us the total number of tosses X it took to land heads, while not telling which coin was picked. For this we formulated following hypotheses:

$$\begin{aligned} H_0 &: \text{ coin 0 was tossed,} \\ H_1 &: \text{ coin 1 was tossed.} \end{aligned}$$

Moreover, it is clear that X is geometrically distributed:

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad k = 1, \dots, n$$

a)

If the prior probabilities are equal, we have to find which outcomes favor H_0 and which favor H_1.

As we mentioned in Exercise 76, the outcomes which favor the null hypothesis are those for which the posterior probability is greater than 1. Additionally, since the prior probabilities are equal, we only have to check for the likelihood ratio being greater than 1.

$$\frac{\mathbb{P}(H_0 \mid X)}{\mathbb{P}(H_1 \mid X)} = \frac{\mathbb{P}(H_0)\,\mathbb{P}(X \mid H_0)}{\mathbb{P}(H_1)\,\mathbb{P}(X \mid H_1)} = \frac{\mathbb{P}(X \mid H_0)}{\mathbb{P}(X \mid H_1)} = \Lambda(X) > 1,$$

Since X is a geometric random variable:

$$\begin{aligned}
\mathbb{P}(X = k \mid H_0) &= (1-0.5)^{k-1} * 0.5 = 0.5^k \\
\mathbb{P}(X = k \mid H_1) &= (1-0.7)^{k-1} * 0.7 = 0.7 * 0.3^{k-1}
\end{aligned}$$

Therefore, the outcomes which favor H_0 are

$$\begin{aligned}
\frac{0.5^k}{0.7 * 0.3^{k-1}} &> 1 \\
0.5^k &> 0.7 * 0.3^{k-1} \\
\left(\frac{0.5}{0.3}\right)^k &> \frac{0.7}{0.3} \\
k * \ln\left(\frac{5}{3}\right) &> \ln\left(\frac{7}{3}\right) \\
k &> \frac{\ln\left(\frac{7}{3}\right)}{\ln\left(\frac{5}{3}\right)} \\
k &> 1.659
\end{aligned}$$

From this we can conclude that, since k has to be an integer, all events $X \geq 2$ favor the null hypothesis, and events $X \leq 1$ favor the alternative hypothesis.

b)

Now we have to suppose $\mathbb{P}(H_0)/\mathbb{P}(H_1) = 10$ and find the outcomes which favor H_0.

Similarly to Part a), we can conclude that the events which would favor the null hypothesis are those which satisfy:

$$\frac{\mathbb{P}(H_0 \mid X)}{\mathbb{P}(H_1 \mid X)} = \frac{\mathbb{P}(H_0)\,\mathbb{P}(X \mid H_0)}{\mathbb{P}(H_1)\,\mathbb{P}(X \mid H_1)} > 1,$$

Since we assume $\mathbb{P}(H_0)/\mathbb{P}(H_1) = 10$, the above is equivalent to

$$\frac{\mathbb{P}(X \mid H_0)}{\mathbb{P}(X \mid H_1)} > \frac{1}{10}$$

Again, since X is a geometric random variable:

$$\begin{aligned}
\mathbb{P}(X = k \mid H_0) &= (1-0.5)^{k-1} * 0.5 = 0.5^k \\
\mathbb{P}(X = k \mid H_1) &= (1-0.7)^{k-1} * 0.7 = 0.7 * 0.3^{k-1}
\end{aligned}$$

Therefore, the outcomes which favor H_0 are

$$
\begin{aligned}
\frac{0.5^k}{0.7 * 0.3^{k-1}} &> 0.1 \\
0.5^k &> 0.07 * 0.3^{k-1} \\
\left(\frac{0.5}{0.3}\right)^k &> \frac{0.07}{0.3} \\
k * \ln\left(\frac{5}{3}\right) &> \ln\left(\frac{7}{30}\right) \\
k &> \frac{\ln\left(\frac{7}{30}\right)}{\ln\left(\frac{5}{3}\right)} \\
k &> -2.849
\end{aligned}
$$

From this we can conclude that, since k has to be a non-negative integer, all possible events favor the null hypothesis.

c)

Next we have to identify the significance level of the test that rejects H_0 if $X \geq 8$.

Due to the definition of the test, the decision function is

$$
d(X) = I_{\text{rejection region}}\,(x) \begin{cases} 1, & X \geq 8 \\ 0, & X \leq 7 \end{cases}
$$

We know that the significance level α is the expected value of $d(X)$ under the probability law specified by H_0.

$$
\begin{aligned}
\alpha &= \mathbb{E}_0(d(X)) \\
&= 1 * \mathbb{P}_0(X \geq 8) + 0 * \mathbb{P}_0(X \leq 7) \\
&= \mathbb{P}_0(X \geq 8) \\
&= \sum_{k=8}^{\infty} \mathbb{P}_0(X = k) \\
&= \sum_{k=8}^{\infty} (1-0.5)^{k-1} * 0.5 = \sum_{k=8}^{\infty} 0.5^k = 0.5^8 \sum_{k=8}^{\infty} 0.5^{k-8} = 0.5^8 * \sum_{k=0}^{\infty} 0.5^k \\
&= 0.5^8 * \frac{1}{1-0.5} \\
&= 0.5^7 = 0.0078125
\end{aligned}
$$

Therefore, $\alpha = 0.0078125$, since under the null hypothesis, X is a geometric random variable with parameter $p = 0.5$.

d)

Lastly, we have to determine the power of this test.

The power of the test, $1 - \beta$, is the expected value of the decision function $d(X)$ from Part c), under the probability law specified by H_1.

$$
\begin{aligned}
1-\beta &= \mathbb{E}_1(d(X)) \\
&= 1 * \mathbb{P}_1(X \geq 8) + 0 * \mathbb{P}_1(X \leq 7) \\
&= \mathbb{P}_1(X \geq 8) \\
&= \sum_{k=8}^{\infty} \mathbb{P}_1(X = k) \\
&= \sum_{k=8}^{\infty} (1-0.7)^{k-1} * 0.7 = \frac{0.7}{0.3} \sum_{k=8}^{\infty} 0.3^k = \frac{0.7}{0.3} * 0.3^8 \sum_{k=8}^{\infty} 0.3^{k-8} = 0.7 * 0.3^7 * \sum_{k=0}^{\infty} 0.3^k \\
&= 0.7 * 0.3^7 * \frac{1}{1-0.3} \\
&= 0.3^7 = 0.0002187
\end{aligned}
$$

Thus, $1-\beta = 0.0002187$, since X is a geometric random variable with the parameter $p = 0.7$ under the alternative hypothesis.

Exercise 79

The likelihood function given λ is given by

$$
f(x_1, \dots, x_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{x_1+\dots+x_n}}{\prod_i^n x_i!} e^{-n\lambda}
$$

We have to find the likelihood ratio for testing $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda = \lambda_A$ where $\lambda_A > \lambda_0$. The ratio is given by

$$
LR = \frac{\lambda_0^{x_1+\dots+x_n} e^{-n\lambda_0}}{\lambda_A^{x_1+\dots+x_n} e^{-n\lambda_A}} = \left(\frac{\lambda_0}{\lambda_A}\right)^{x_1+\dots+x_n} e^{-n(\lambda_0-\lambda_A)}
$$

We need to find

$$
P(\text{rejecting } H_0|H_0) = P(LR < c|H_0) = \alpha
$$

As stated in the instructions, the sum of independent Poisson random variables are again Poisson distributed. I.e $S = X_1 + \dots + X_n \sim \text{Pois}(n\lambda)$. Since λ_0/λ_A and $e^{-n(\lambda_0-\lambda_A)}$ are constants, the LR is just a transformation of the random variable S, from which we know its distribution. Thus we can compute $P(LR < c|H_0) = P(S < f(c)) = \alpha$ for a fixed α by computing $Q_S(\alpha) = f(c)$ and solve for c.

Next we show that the obtained test is UMP for $H_0 : \lambda = \lambda_0$ versus $H_A : \lambda > \lambda_0$. We reject the H_0 if we have

$$
\begin{gathered}
\left(\frac{\lambda_0}{\lambda_A}\right)^{x_1+\dots+x_n} e^{-n(\lambda_0-\lambda_A)} < c \\
n\bar{x}(\log(\lambda_0) - \log(\lambda_A)) - n(\lambda_0 - \lambda_A) < \log(c) \\
\bar{x}(\log(\lambda_0) - \log(\lambda_A)) < \log(c)/n + \lambda_0 - \lambda_A := c_1
\end{gathered}
$$

For $\lambda_A > \lambda_0$

$$
\bar{x} > \frac{c_1}{\log(\lambda_0) - \log(\lambda_A)}
$$

For all $\lambda_A > \lambda_0$ $\log(\lambda_0) - \log(\lambda_A) < 0$ and the test rejects H_0 for $\bar{x} > c_1$. c_1 itselfs depends on α, λ_0 and n but not not on λ_A. So for all $\lambda_A > \lambda_0$ all alternative test reject the H_0 and are the same. The test ist thus UMP.

Exercise 80

Neyman-Pearson Lemma states that given null hypothesis H_0 and an alternative hypothesis H_1, and the test that rejects H_0 whenever the likelihood ratio is less than c at significance level α. $P(\Lambda < c|H_0) = \alpha$

We derive

$$\Lambda = \frac{f(x|H_0)}{f(x|H_A)} = \frac{\prod_{i=1}^n \theta_0 \gamma^{\theta_0}/x_i^{\theta_0+1}}{\prod_{i=1}^n \theta_A \gamma^{\theta_A}/x_i^{\theta_A+1}} = \frac{\prod_{i=1}^n \theta_0 \gamma^{\theta_0} x_i^{\theta_A+1}}{\prod_{i=1}^n \theta_A \gamma^{\theta_A} x_i^{\theta_0+1}}$$

Now we set $\Lambda < c$

$$\frac{\prod_{i=1}^n \theta_0 \gamma^{\theta_0} x_i^{\theta_A+1}}{\prod_{i=1}^n \theta_A \gamma^{\theta_A} x_i^{\theta_0+1}} < c$$

$$\left(\frac{\theta_0}{\theta_A}\right)^n \prod_{i=1}^n \left(\frac{x_i}{\gamma}\right)^{\theta_A-\theta_0} < c$$

$$(\theta_A - \theta_0)\sum_i^n \log(x_i/\gamma) < \log\left(c\left(\frac{\theta_A}{\theta_0}\right)^n\right)$$

$$\sum_i^n \log\left(\frac{x_i}{\gamma}\right) < \log\left(c\left(\frac{\theta_A}{\theta_0}\right)^n\right)/(\theta_A - \theta_0) = c_1$$

We then determine the probability of $\log(X/\gamma) \leq x$

$$P(\log(X/\gamma)) \leq x) = P(X \leq e^x\gamma).$$

finding out that *$log(X/\gamma) \sim f(x,\lambda)$ exponential distribution*. From the text we know that X follows a pareto distribution, then

$$P(X \leq e^x\gamma) = \left(1 - \frac{\gamma}{e^x\gamma}\right)^\theta \cdot I_{e^x\gamma\geq\gamma} = 1 - e^{-\theta x} \cdot I_{x\geq 0}\ (\textit{exponential distribution})$$

To sum up, $\log(X/\gamma) \implies$ exponentially distributed, $\sum_i^n \log\left(\frac{x_i}{\gamma}\right) \sim \Gamma(n,\theta)$. Let $K = \sum_{i=1}^n \log(X/\gamma)$ we have

$$\alpha = P(K < c_1|H_0) = F_{\theta_0}(c_1) \iff c_1 = F_{\theta_0}^{-1}(\alpha)$$

The most powerful level α test given H_0 is one where H_0 is rejected when $S < F_{\theta_0}^{-1}(\alpha)$ where $F_{\theta_0}^{-1}$ is the quantile function of a $\Gamma(n,\theta_0)$ distribution. We reject H_0 hypothesis when Λ is smaller than $F_{\theta_0}^{-1}$ quantile.

Part b)

Neyman Pearson test is the most powerful test for all $\lambda_A > \lambda_0$ from *part a*, where we rejects H_0 if $\sum_i^n \log(X_i/\gamma) < F_{\theta_0}^{-1}(\alpha)$ [$F_{\theta_0}^{-1}$ is the quantile function of a $\Gamma(n,\theta_0)$ distribution]. The latter is independent from θ_A, but is related to α, θ_0, n but As a consequence we get the same test for all $\lambda_A > \lambda_0$ and thus the test is UMP. The power function can be derived by

$$P(\text{rejecting } H_0|H_A) = P_A(S < c_1) = F_{\theta_A}(c_2) \iff c_2 = F_{\theta_A}^{-1}(\alpha)$$

where $F_{\theta_A}^{-1}$ is the quantile function of a $\Gamma(n,\theta_A)$ distribution.

Exercise 81

To show that the likelihood ratio test of $H_0 : \theta = \theta_0$ versus $H_A : \theta = \theta_A$ is a function of T, we can use the factorization theorem. The principle behind it is that one can express $f(x_1, ..., x_n|\theta)$ as $g(t(x_1, ..., x_n), \theta) \cdot h(x_1, ..., x_n)$. This will change how the likelihood ration formula looks like the following way:

$$LR = \frac{f(x_1, ..., x_n|\theta_0)}{f(x_1, ..., x_n|\theta_A)} = \frac{g(t(x_1, ..., x_n), \theta_0) \cdot h(x_1, ..., x_n)}{g(t(x_1, ..., x_n), \theta_A) \cdot h(x_1, ..., x_n)} = \frac{g(T, \theta_0)}{g(T, \theta_A)}$$

The last entity here is just a function of T, for which we know the distribution under the null hypothesis, which also gives us the probability os a type I error:

$$\alpha = \mathbb{P}(reject H_0|H_0) = \mathbb{P}_0\left(\frac{g(T, \theta_0)}{g(T, \theta_A)} < c\right)$$

This way, this formula can used with the known distribution of T to find the corresponding c. The alternative to this is deriving the distribution of $G = \frac{g(T,\theta_0)}{g(T,\theta_A)}$ and only then finding c.

Exercise 82

Let $X \sim N(0, \sigma^2)$ and let consider testing H0 : $\sigma = \sigma_0$ versus HA : $\sigma = \sigma_A$, where $\sigma_A > \sigma_0$.

a)

What is the likelihood ratio as a function of x? What values favor H0? What is the rejection region of a level α test?

The likelihood ratio for $H_0 : \sigma = \sigma_0$ and $H_A : \sigma = \sigma_A$ is

$$\Lambda = \frac{f(x \mid \sigma = \sigma_0)}{f(x \mid \sigma = \sigma_A)} = \exp\left(-\frac{x^2}{\sigma_0^2} + \frac{x^2}{\sigma_A^2}\right) = \exp\left(x^2\left(\sigma_A^{-2} - \sigma_0^{-2}\right)\right)$$

Since $\sigma_A > \sigma_0$, which implies $\frac{1}{\sigma_A^2} < \frac{1}{\sigma_0^2} \implies \sigma_A^{-2} - \sigma_0^{-2} < 0$ and the likelihood ratio is decreasing in x^2. Under H_0, the random variable $\frac{X^2}{\sigma_0^2}$ is Chi-squared distributed with 1 degree of freedom.

$$\alpha = \mathbb{P}\left(X^2 > c \mid \sigma = \sigma_0\right) = \mathbb{P}\left(\frac{X^2}{\sigma_0^2} > \frac{c}{\sigma_0^2} \mid \sigma = \sigma_0\right) = 1 - F_1\left(\frac{c}{\sigma_0^2}\right) \implies c = F_1^{-1}(1-\alpha)\sigma_0^2$$

b)

Now we have a random sample $X_1, \ldots, X_n$.

What is the likelihood ratio as a function of x? What values favor H0? What is the rejection region of a level α test?

The likelihood ratio for $H_0 : \sigma = \sigma_0$ and $H_A : \sigma = \sigma_A$ is:

$$\Lambda = \frac{f(x \mid \sigma = \sigma_0)}{f(x \mid \sigma = \sigma_A)} = \exp\left(\sum_{i=1}^{n} X_i^2\left(-\sigma_0^{-2} + \sigma_A^{-2}\right)\right)$$

Analogously, since under H_0 the random variable $\frac{\sum X_i^2}{\sigma_0^2}$ is Chi-squared distributed with $k = n$ degrees of freedom.

$$\alpha = 1 - F_n\left(\frac{c}{\sigma_0^2}\right) \implies c = F_n^{-1}(1-\alpha)\sigma_0^2$$

Again, we can see that this test does not depend on σ_A. Therefore, it is uniformly most powerful (UMP) for testing $H_0 : \sigma = \sigma_0$ versus $H_A : \sigma > \sigma_0$.

c)

Is the test in the previous question uniformly most powerful (UMP) for testing $H_0 : \sigma = \sigma_0$ versus $H_A : \sigma > \sigma_0$.

YES!

Exercise 83

a)

We continue to use the well-established likelihood ratio test, which assumes that the observed sample values have non-zero probability values.

$$\begin{aligned}
&LR = \frac{f(x \mid H_0)}{f(x \mid H_A)} = \frac{1}{1/2} = 2 \\
&\alpha = \Pr(\text{ reject } H_0 \mid H_0) = \Pr_0(LR < c) = \Pr_0(2 < c) \\
&\alpha \stackrel{!}{=} 0 \Rightarrow c \leq 2 \\
&1 - \beta = \Pr(\text{ take } H_A \mid H_A) = 0
\end{aligned}$$

Through our analysis, we realized that if the rejection threshold c is less than or equal to 2, then the null hypothesis H_0 will not be rejected, but the alternative hypothesis H_A will also not be accepted, resulting in zero power.

b)

$$\begin{aligned}
\alpha = \Pr(\text{ reject } H_0 \mid H_0) = \Pr(X \leq c \mid \theta = 1) = \Pr_0(X \leq c) = c \\
1 - \beta = \Pr(\text{ take } H_A \mid H_A) = \Pr(\text{ reject } H_0 \mid \theta = 2) = \Pr_A(X \leq c) = \tfrac{c}{2}
\end{aligned}$$

c)

$$\begin{aligned}
\alpha = \Pr(\text{ reject } H_0 \mid H_0) = \Pr(1 - c \leq X \leq 1) = F_{X,H_0}(1) - F_{X,H_0}(1-c) = 1 - (1-c) = c \\
1 - \beta = \Pr(\text{ reject } H_0 \mid H_A) = \Pr(1 - c \leq X \leq 1) = F_X, H_A(1) - F_{X,H_A}(1-c) = 1/2 - \tfrac{1-c}{2} = \tfrac{c}{2}
\end{aligned}$$

d)

As demonstrated earlier, the likelihood ratio is unaffected by both the sample x and the significance level α. As a result, the LRT establishes a distinct region of rejection by specifying the threshold for rejection (c). This implies that the region of rejection will remain identical for any sample and level of significance.

e)

All previously mentioned tests will have their size and power interchanged since they will be based on the contradicting hypothesis. However, the LRT will maintain the same unique rejection region as before, even when the hypotheses are interchanged.

Exercise 84

a)

The density of the uniform distribution in the interval [0,1] is obtained when we set the parameter θ to 1, that is, $f(x|\theta = 1) = 1$.

b)

To test if the data comes from the uniform distribution, we will use as our null hypothesis:

$H_0 : \theta = 1$ and as alternative hypothesis

$H_1 : \theta \neq 1$.

So the generalized likelihood ratio is:

$\Lambda = \frac{\theta_0^n x^{n(\theta_0 - 1)}}{\sup_{\theta \in R} \theta^n \prod_{i=1}^n x_i^{\theta-1}} = \frac{1}{\hat{\theta}_{MLE}^n \prod_{i=1}^n x_i^{\hat{\theta}_{MLE}-1}}$

Then we will find the maximum likelihood estimator:

$lik(\theta) = \theta^n \prod_{i=1}^n x_i^{\theta-1}$

Then we will take the logarithm:

$l(\theta) = nlog(\theta) + (\theta - 1)\sum_{i=1}^n log(x_i)$

Then we will take the derivative:

$l'(\theta) = \frac{n}{\theta} + \sum_{i=1}^n log(x_i) = 0$

$$\hat{\theta}_{MLE} = \frac{-n}{\sum_{i=1}^n log(x_i)} = \frac{1}{\frac{1}{n}\sum_{i=1}^n -log(x_i)}$$

As the data size approaches infinity, the χ^2 distribution becomes equivalent to the -2 times the logarithm of the generalized likelihood ratio. Hence:

$$log(\Lambda) = -nlog(\hat{\theta}_{MLE}) + (\hat{\theta}_{MLE} - 1)\sum_{i=1}^n log(x_i)$$

$$-2log(\Lambda) = 2nlog(\hat{\theta}_{MLE}) - 2(\hat{\theta}_{MLE} - 1)\sum_{i=1}^n log(x_i)$$

We substitute the previous expression:

$$-2log(\Lambda) = 2nlog(\frac{1}{\frac{1}{n}\sum_{i=1}^n -log(x_i)}) - 2(\frac{1 - \frac{1}{\frac{1}{n}\sum_{i=1}^n -log(x_i)})}{\frac{1}{n}\sum_{i=1}^n -log(x_i))}\sum_{i=1}^n log(x_i) =$$

$$2 - 2nlog(\frac{1}{n}\sum_{i=1}^{n} -log(x_i)) - 2(\frac{1}{n}\sum_{i=1}^{n} -log(x_i))$$

Finally,we can see that if this is bigger than $\chi^2_{1,1-\alpha}$, then we can reject the null hypothesis with a significance level of α