Statistics 2 Unit 4 Team 8

Nikolaos Kornilakis Rodrigo Viale Jakub Trnan Aleksandra Daneva
Luis Diego Pena Monge

2024-04-17

Exercise 61

a)

Let $X \sim U[0, \theta]$, and assume as prior for $\theta \sim Pareto(\alpha = 7, \beta = 4)$. Let's first prove that the Pareto provides a conjugate prior to the uniform distribution in general:

$$f_{X|\theta}(X|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{I}_{\{X_i \in [0,\theta]\}} = \left(\frac{1}{\theta}\right)^n \mathbb{I}_{\{max(X_i) \in [0,\theta]\}}$$

$$f_\theta(\theta|\alpha, \beta) \propto \theta^{-(\alpha+1)} \mathbb{I}_{\{\beta \leq \theta\}}$$

We therefore have that:

$$f_{\theta|X}(\theta|X) \propto \theta^{-(\alpha+1)} \mathbb{I}_{\{\beta \leq \theta\}} \left(\frac{1}{\theta}\right)^n \mathbb{I}_{\{max(X_i) \in [0,\theta]\}}$$

$$= \theta^{-(\alpha+n+1)} \mathbb{I}_{\{max(\beta, max(X_i)) \leq \theta\}}$$

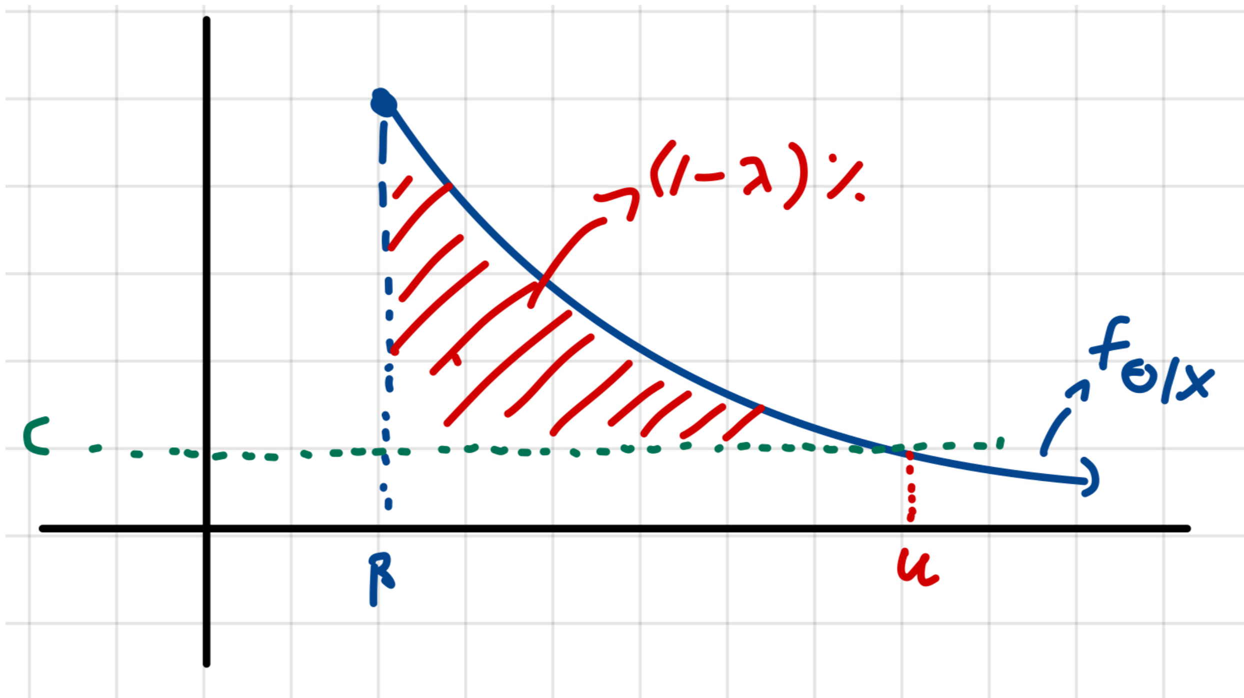
Which is exactly the kernel of a Pareto distribution with parameters $\alpha_{post} = \alpha + n$ and $\beta_{post} = max\{\beta, max(X_i : i = 1, ..., n)\}$. Replacing the given values for the exercise, and noting that the maximum observation is 14, we obtain that our posterior distribution is $Pareto(\alpha = 12, \beta = 14)$.

b)

Recall that for a Pareto distribution with parameters α, β, the expected value is given by $\frac{\alpha\beta}{\alpha-1}$. Then, the posterior mean ot θ is given by:

$$E[\theta|X] = \frac{12 \cdot 14}{12 - 1} = 168$$

To visualize our solution, let's first plot what it is we want to calculate:



c)

That means that if we want to find the $(1-\lambda)\%$ HPD for θ, we need to find a value of u, c such that:

$$F_{\theta|X}(u|X) - (u-\beta)\cdot c = 1-\lambda$$

Notice three things: first the previous equality is possible due to the form of $f_{\theta|X}$, specifically the fact that the density function is not only continuous but non-increasing in its whole domain. This guarantees that the HPD is given by an interval and that the start of the interval must be exactly β. Second, by construction we must have $c = f_{\theta|X}(u|X)$, then:

$$F_{\theta|X}(u|X) - (u-\beta)\cdot f_{\theta|X}(u|X) = 1-\lambda$$

And third, if you were to expand this expression (which we did), this is not an analytically solvable equation. Therefore, we utilize numeric root-finding algorithms to compute the value of u. Specifically, we use Newton's method, and note that the function we want to find the root for and its first derivative are, respectively:

$$\begin{aligned}
\mathcal{F}(u) &= F_{\theta|X}(u|X) - (u-\beta)\cdot f_{\theta|X}(u|X) - (1-\lambda)\\
\mathcal{F}'(u) &= F'_{\theta|X}(u|X) - [f_{\theta|X}(u|X) + uf'_{\theta|X}(u|X)] + \beta f'_{\theta|X}(u|X)\\
&= (\beta-u)f'_{\theta|X}(u|X)
\end{aligned}$$

```r
# Parameters
alpha <- 12
beta <- 14

# Significance level
lambda <- 0.05

# CDF
F.theta <- function(u) 1 - (beta/u)^alpha

# PDF
```

```r
f.theta <- function(u) alpha*beta^alpha / u^(alpha+1)

# Derivative of PDF
f.prime.theta <- function(u) -alpha*(alpha+1)*beta^alpha*u^(-alpha-2)

# Target function
f.target <- function(u) F.theta(u) - (u-beta)*f.theta(u) - (1 - lambda)

# Derivative of target function
f.prime.target <- function(u) (beta - u)*f.prime.theta(u)

newtonMethod <- function(x0, f, fprime, tol, iter = 0){
  if(abs(f(x0)) < tol){
    return( c(x = x0, y = f(x0), iter = iter) )
  }else{
    Recall( x0 - f(x0)/fprime(x0), f, fprime, tol, 1+iter)
  }
}

(root <- newtonMethod(15, f.target, f.prime.target, tol = 1e-6))
```

```
##              x              y           iter
##  2.047801e+01 -5.180760e-09  6.000000e+00
```

We see that the root is found at $u = 20.47801$, let's verify this:

```r
library(EnvStats)

u <- root[1]
c <- dpareto(u, shape = alpha, location = beta)

ppareto(u, shape = alpha, location = beta) - (u - beta)*c
```

```
##    x
## 0.95
```

Which verifies the value found numerically. Therefore, a 95% HPD interval for θ is given by $[0, 20.47801]$

d)

We proceed to test the hypotheses. For that we will use the previously defined function for the CDF of our posterior distribution, and the facts that:

$$P(0 \leq \theta \leq 15) = F_{\theta|X}(15|X) - F_{\theta|X}(14|X)P(\theta > 15) = 1 - P(0 \leq \theta \leq 15)$$

Notice the first equality comes from the definition of the Pareto distributions and our posterior parameters. In other words, we have that:

$$\begin{aligned} P(0 \leq \theta \leq 15) &= P(\{0 \leq \theta < 14\} \cup \{14 \leq \theta \leq 15\}) \\ &= P(0 \leq \theta < 14) + P(14 \leq \theta \leq 15) \\ &= 0 + P(14 \leq \theta \leq 15) \end{aligned}$$

```r
p_H0 <- F.theta(15) - F.theta(14)
p_H1 <- 1 - p_H0
```

```r
c(H0 = p_H0, H1 = p_H1)
```

```
##        H0        H1
## 0.5630404 0.4369596
```

We observe therefore that we cannot reject the null hypothesis as it is the more likely one given the posterior obtained with the observed realizations of the random variable.

Exercise 62

a)

Let $X_1, ..., X_n \overset{iid}{\sim} f(x|p) = p(1-p)^x$. Suppose p's prior is given by $p \sim Beta(\alpha, \beta)$. Then we have that:

$$f(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$$

$$f_{X|p}(X|p) = \prod_{i=1}^{n} f(X_i|p) = p^n(1-p)^{\sum_{i=1}^n X_i}$$

Then the posterior distribution proportional to the product of the previous two, as follows:

$$\begin{aligned} f_{p|X} &\propto f(p) \cdot f_{X|p}(X|p) \\ &= p^{\alpha-1+n}(1-p)^{\beta-1+\sum_{i=1}^n X_i} \\ &= p^{\alpha+n-1}(1-p)^{\beta+\sum_{i=1}^n X_i-1} \end{aligned}$$

Which is exactly the kernel for a beta distribution with parameters $\alpha_{posterior} = \alpha + n$ and $\beta_{posterior} = \beta + \sum_{i=1}^n X_i$. This means that the Bayes estimator with respect to squared error is given by the posterior expectation, which (since the posterior is beta) is:

$$E[p|X] = \frac{\alpha + n}{\beta + \sum_{i=1}^n X_i}$$

b)

We start by calculating the Fisher information for the parameter p:

$$
\begin{aligned}
I(p) &= E\left[\left(\frac{d}{dp}ln(f(X|p))\right)^2\right] \\
&= E\left[\left(\frac{d}{dp}ln(p(1-p)^X)\right)^2\right] \\
&= E\left[\left(\frac{d}{dp}ln(p) + Xln(1-p))\right)^2\right] \\
&= E\left[\left(\frac{1}{p} - \frac{X}{1-p}\right)^2\right] \\
&= \frac{1}{p^2} - \frac{2}{p(1-p)}E[X] + \frac{1}{(1-p)^2}E[X^2]
\end{aligned}
$$

Recall that for a geometric we have $E[X] = \dfrac{1-p}{p}$ and $Var[X] = \dfrac{1-p}{p^2}$

$$
\begin{aligned}
&= \frac{1}{p^2} - \frac{2}{p(1-p)}\frac{1-p}{p} + \frac{1}{(1-p)^2}\left(\frac{1-p}{p^2} + \frac{(1-p)^2}{p^2}\right) \\
&= (1-p)^{-1}p^{-2}
\end{aligned}
$$

Then we have that our noninformative prior is proportional to:

$$
f(p) \propto \sqrt{I(p)} = (1-p)^{-1/2}p^{-1}
$$

Which (abusing slightly notation) corresponds to the kernel of a $Beta(1/2, 0)$ random variable. Then we have that our posterior follows:

$$
\begin{aligned}
f_{p|X}(p|X) &\propto (1-p)^{-1/2}p^{-1}p^n(1-p)^{\sum_{i=1}^n X_i} \\
&= (1-p)^{\sum_{i=1}^n X_i + 1/2 - 1}p^{n-1}
\end{aligned}
$$

Which corresponds to the kernel of a $Beta(n, \sum_{i=1}^n X_i + 1/2)$ distribution. This means that the posterior mean (Bayes estimator) is given by:

$$
E[p|X] = \frac{n}{n + \sum_{i=1}^n X_i + \frac{1}{2}}
$$

Exercise 63

a)

Notice that:

$$
\begin{aligned}
E[X] &= 0 \cdot \frac{2\theta}{3} + 1 \cdot \frac{\theta}{3} + 2 \cdot \frac{2(1-\theta)}{3} + 3 \cdot \frac{1-\theta}{3} \\
&= \frac{-6\theta + 7}{3}
\end{aligned}
$$

Since the expected value is the first moment, which we estimate via $\bar{X}$, our method of moments estimator for θ is given by:

$$\theta_{MoM} = \frac{7 - 3\bar{X}}{6}$$

Given our sample:

```
sample <- c(3, 0, 2, 1, 3, 2, 1, 0, 2, 1)
Xbar <- mean(sample)
(theta_mom <- (7 - 3*Xbar)/6)
```

```
## [1] 0.4166667
```

b)

To calculate an approximation for the standard error, notice that the MoM estimation relies on sample means. This screams for the use of central limit theorem, which is what we do next.

$$\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \to N(0,1) \Rightarrow \bar{X} \to N(\mu, \sigma^2/n)$$

To do so, we first find the MoM estimator for the variance:

$$\begin{aligned} E[X^2] &= 0^2 \cdot \frac{2\theta}{3} + 1^2 \cdot \frac{\theta}{3} + 2^2 \cdot \frac{2(1-\theta)}{3} + 3^2 \cdot \frac{1-\theta}{3} \\ &= \frac{17 - 16\theta}{3} \end{aligned}$$

Then:

$$Var[X] = E[X^2] - E[X]^2 = \frac{17 - 16\theta}{3} - \left(\frac{-6\theta + 7}{3}\right)^2 = \frac{2}{9} + 4\theta - 4\theta^2$$

Then by CLT, asymptotically:

$$\bar{X} \sim N\left(\frac{-6\theta + 7}{3}, \frac{2}{9n} + \frac{4\theta}{n} - \frac{4\theta^2}{n}\right)$$

Then:

$$\theta_{MoM} = \frac{7 - 3\bar{X}}{6} \sim N\left(\theta, \frac{1}{4}\left[\frac{2}{9n} + \frac{4\theta}{n} - \frac{4\theta^2}{n}\right]\right)$$

Which means we can approximate the standard error using the MoM as plug-in estimator:

$$SE(\theta_{MoM}) \approx \sqrt{\frac{1}{4}\left[\frac{2}{9n} + \frac{4\theta_{MoM}}{n} - \frac{4\theta_{MoM}^2}{n}\right]}$$

Given our sample:

```
n <- length(sample)
(se_mom <- sqrt(1/(4*n) * (2/9 + 4*theta_mom - 4*theta_mom^2)  ))
```

```
## [1] 0.1728037
```

c)

Notice that for any given sample with n_i observations of i, we have that the likelihood is:

$$L = \left(\frac{2\theta}{3}\right)^{n_0} \left(\frac{\theta}{3}\right)^{n_1} \left(\frac{2(1-\theta)}{3}\right)^{n_2} \left(\frac{(1-\theta)}{3}\right)^{n_3}$$

Then the log-likelihood is given by:

$$LL = n_0(ln(2\theta) - ln(3)) + n_1(ln(\theta) - ln(3)) + n_2(ln(2(1-\theta)) - ln(3)) + n_3(ln(1-\theta) - ln(3))$$

We maximize the log-likelihood:

$$\begin{aligned}
\frac{d}{d\theta}LL = 0 &\iff \frac{n_0}{\theta} + \frac{n_1}{\theta} - \frac{n_2}{1-\theta} - \frac{n_3}{1-\theta} = 0 \\
&\iff \frac{n_0 + n_1}{\theta} = \frac{n_2 + n_3}{1-\theta} \\
&\iff \theta = \frac{n_0 + n_1}{n_0 + n_1 + n_2 + n_3}
\end{aligned}$$

Then the MLE estimate for θ is $\theta_{MLE} = \frac{n_0+n_1}{n}$. For our sample:

```
(theta_mle <- sum(sample %in% c(0,1))/n)
```

```
## [1] 0.5
```

d)

To find a standard error approximation, we will use the asymptotic normality of the MLE. For that, let's calculate its Fisher information:

$$\begin{aligned}
I(\theta) &= -E\left[\frac{d^2}{d\theta^2}LL\right] \\
&= -E\left[\frac{d}{d\theta}\frac{n_0 + n_1}{\theta} - \frac{n_2 + n_3}{1-\theta}\right] \\
&= -E\left[\frac{-(n_0 + n_1)}{\theta^2} - \frac{n_2 + n_3}{(1-\theta)^2}\right] \\
&= \frac{1}{\theta^2}(E[n_0] + E[n_1]) + \frac{1}{(1-\theta)^2}(E[n_2] + E[n_3])
\end{aligned}$$

Notice in this case since the Fisher information is for a single observarion of the r.v, each n represents the indicator of observing that particular value.

$$\begin{aligned}
&= \frac{1}{\theta^2}\left(\frac{2\theta}{3} + \frac{\theta}{3}\right) + \frac{1}{(1-\theta)^2}\left(\frac{2(1-\theta)}{3} + \frac{1-\theta}{3}\right) \\
&= \frac{1}{\theta(1-\theta)}
\end{aligned}$$

Then, using the estimator as plug-in, asymptotically the standard error of the MLE estimator is given by:

$$SE(\theta_{MLE}) = \frac{1}{\sqrt{nI(\theta)}} = \sqrt{\frac{\theta_{MLE}(1-\theta_{MLE})}{n}}$$

For our sample:

```
(se_mle <- sqrt( theta_mle*(1 - theta_mle)/n ))
```

```
## [1] 0.1581139
```

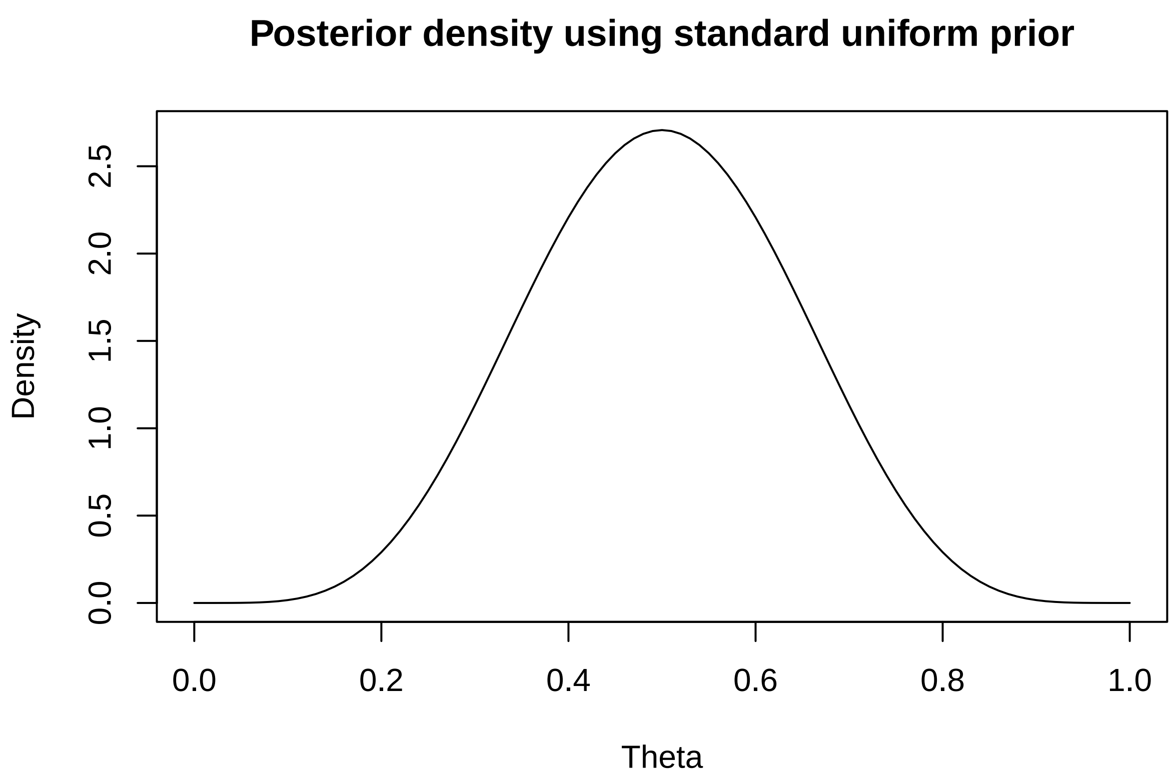
e)

Suppose θ has a standard uniform prior, then $f(\theta) = \mathbb{I}_{\theta\in[0,1]}$, and as seen before $f(X|\theta) = \left(\frac{2\theta}{3}\right)^{n_0}\left(\frac{\theta}{3}\right)^{n_1}\left(\frac{2(1-\theta)}{3}\right)^{n_2}\left(\frac{(1-\theta)}{3}\right)^{n_3}$. Then:

$$f_{\theta|X}(\theta|X) \propto \theta^{n_0+n_1}(1-\theta)^{n_2+n_3}\mathbb{I}_{\theta\in[0,1]}$$

Which is exactly the kernel for a beta distribution with parameters $\alpha = n_0 + n_1 + 1$ and $\beta = n_2 + n_3 + 1$. We proceed to plot it:

```
post <- function(x) dbeta(x,
                          shape1 = sum(sample %in% c(0,1))+1,
                          shape2 = sum(sample %in% c(2,3))+1)
curve(post,
      main = "Posterior density using standard uniform prior",
      xlab = "Theta",
      ylab = "Density")
```



Now, to calculate the mode of this distribution, notice that it's enough to look at and maximize its kernel (as everything else is constant):

$$\frac{d}{d\theta} f_{\theta|X}(\theta|X) \propto (\alpha-1)\theta^{\alpha-2}(1-\theta)^{\beta-1} - (\beta-1)(1-\theta)^{\beta-2}\theta^{\alpha-1} = 0 \iff (\alpha-1)\theta^{\alpha-2}(1-\theta)^{\beta-1} = (\beta-1)(1-\theta)^{\beta-2}\theta^{\alpha-1}$$
$$\iff (\alpha-1)(1-\theta) = (\beta-1)\theta$$
$$\iff \theta = \frac{\alpha-1}{\alpha+\beta-2}$$

Replacing with $\alpha = n_0 + n_1 + 1$ and $\beta = n_2 + n_3 + 1$, we get that the posterior mode is given by $\frac{n_0+n_1}{n_0+n_1+n_2+n_3}$, which corresponds to the MLE found above.

Exercise 65

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

a)

We start with the likelihood function:

$$\begin{aligned} L &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i-\mu)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\right). \end{aligned}$$

Taking the natural logarithm of the likelihood (log-likelihood), we have:

$$\begin{aligned} LL &= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2. \end{aligned}$$

To find the MLE of σ, we differentiate with respect to σ^2 and set the derivative to 0:

$$\frac{d}{d\sigma^2}LL = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(X_i-\mu)^2 = 0 \iff \sum_{i=1}^{n}(X_i-\mu)^2 = n\sigma^2$$
$$\iff \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2.$$

b)

This time, we maximize the log-likelihood with respect to *mu*:

$$\frac{d}{d\mu}LL = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0 \iff \sum_{i=1}^{n}X_i - n\mu = 0$$
$$\iff \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}X_i.$$

c)

Recall the Cramér-Rao lower bound that states if $\hat{T}$ is an unbiased estimator of θ, then:

$$\text{Var}(\hat{T}) \geq \frac{1}{nI(\theta)}$$

Recall that the MLE for μ is unbiased:

$$E(\hat{\mu}_{\text{MLE}}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \mu$$

Then, we can calculate its Fisher information as follows:

$$\begin{aligned}
I(\mu) &= -E\left(\frac{\partial^2}{\partial \mu^2} \ln f(X_i|\mu)\right) \\
&= -E\left(\frac{\partial^2}{\partial \mu^2}\left(-\frac{1}{2\sigma^2}(X_i - \mu)^2\right)\right) \\
&= -E\left(\frac{\partial}{\partial \mu}\left(\frac{X_i - \mu}{\sigma^2}\right)\right) \\
&= E\left(\frac{1}{\sigma^2}\right) \\
&= \frac{1}{\sigma^2}
\end{aligned}$$

Therefore, the variance of $\hat{T}$ is bounded by:

$$\text{Var}(\hat{T}) \geq \frac{\sigma^2}{n}$$

Now let's check $\hat{\mu}_{\text{MLE}}$'s variance:

$$\begin{aligned}
\text{Var}(\hat{\mu}_{\text{MLE}}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

Thus, the MLE for μ achieves the Cramér-Rao lower bound, implying that no other unbiased estimator of μ has a lower variance when σ is known.

Exercise 67

a)

Given the following model,

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

we can conclude that $\mathbb{E}\left(Y_i\right) = \beta x_i$ since the expectation of ϵ_i is 0. We can also express the ϵ in terms of Y and βx: $\epsilon_i = Y_i - \beta x_i$, and this expression is also normally distributed. Then, we can construct the density for Y_i:

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

Next, we derive the likelihood function:

$$\operatorname{lik}(\beta, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{\left(y_i - \beta x_i\right)^2}{\sigma^2}\right)$$

And by taking logs the log-likelihood:

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

Now, to get the MLE for β, we need set the derivative of the log-likelihood with respect to this parameter equal to zero:

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{\sigma^2} \sum_{i=1}^{n} \left(-x_i\right)\left(y_i - \beta x_i\right) = 0$$

$$\sum_{i=1}^{n} x_i^2 \beta - \sum_{i=1}^{n} x_i y_i = 0$$

$$\Rightarrow \hat{\beta}_{MLE} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Now, let us determine whether the MLE is biased. Note that the estimator is unbiased if $\mathbb{E}\left[\hat{\beta}_{MLE}\right] = \beta$:

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}_{MLE}\right] &= \mathbb{E}\left[\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}\right] = \mathbb{E}\left[\frac{\sum_{i=1}^{n} x_i(\beta x_i + \epsilon_i)}{\sum_{i=1}^{n} x_i^2}\right] \\
&= \mathbb{E}\left[\frac{\beta \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \epsilon_i x_i}{\sum_{i=1}^{n} x_i^2}\right] \\
&= \mathbb{E}\left[\frac{\beta \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2} + \frac{\sum_{i=1}^{n} \epsilon_i x_i}{\sum_{i=1}^{n} x_i^2}\right] \\
&= \mathbb{E}[\beta] + \mathbb{E}\left[\frac{\sum_{i=1}^{n} \epsilon_i x_i}{\sum_{i=1}^{n} x_i^2}\right] = \beta \qquad \text{(note that } \mathbb{E}[\epsilon_i] = 0\text{)}
\end{aligned}$$

We can see that the MLE for β is unbiased. To calculate the MSE of the estimator, we use the fact that for an unbiased estimator, $MSE(\hat{\beta}_{MLE}) = \mathbb{V}(\hat{\beta}_{MLE})$. The $\hat{\beta}_{MLE}$ can be also written as $\beta + \frac{\sum_{i=1}^{n} \epsilon_i x_i}{\sum_{i=1}^{n} x_i^2}$, when we substitute y with $\beta x + \epsilon$ in the original expression for $\hat{\beta}_{MLE}$.

$$
\begin{aligned}
\mathbb{V}\left[\hat{\beta}_{MLE}\right] &= \mathbb{V}\left[\beta + \frac{\sum_{i=1}^{n} \epsilon_i x_i}{\sum_{i=1}^{n} x_i^2}\right] \\
&= 0 + \mathbb{V}\left[\frac{\sum_{i=1}^{n} \epsilon_i x_i}{\sum_{i=1}^{n} x_i^2}\right] \qquad \text{(using independence)} \\
&= \frac{1}{\left(\sum_{i=1}^{n} x_i^2\right)^2}\left(\sum_{i=1}^{n} x_i^2 \mathbb{V}\left[\epsilon_i\right]\right) \qquad \text{(note that } \mathbb{V}\left[\epsilon_i\right] = \sigma^2\text{)} \\
&= \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}
\end{aligned}
$$

and this is our MSE.

b)

Now we need to obtain the MLE for σ. We proceed similarly, find the derivative of the log-likelihood function and set it equal to 0 to maximize. Recall the log-likelihood function:

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

Differentiating w.r.t σ:

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

c)

Now, we need to find the MLE for $\frac{\beta}{\sigma}$. The ratio of β and σ that maximizes the log-likelihood function is determined by dividing the respective individual values that together maximize the log-likelihood, which were computed in part a) and b).

$$
\begin{aligned}
\left(\frac{\beta}{\sigma}\right)_{MLE} &= \frac{\hat{\beta}_{MLE}}{\hat{\sigma}_{MLE}} \\
&= \frac{\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\beta}x_i\right)^2}} \\
&= \frac{\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} x_j^2}x_i\right)^2}}
\end{aligned}
$$

d)

In this part we derive the Fisher information matrix $I(\beta,\sigma)$. Recall that the matrix under appropriate smoothness conditions is given by:

$$I(\beta,\sigma) = -E[\mathcal{H}(\ell(\beta,\sigma))]$$

To compute the Hessian matrix, we start by first order derivatives of log-likelihood w.r.t to both variables β and σ,

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}(-x_i)\left(y_i-\beta x_i\right) = -\frac{1}{\sigma^2}\left(\beta\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_iy_i\right)\\
\frac{\partial \ell}{\partial \sigma} &= -\frac{n}{\sigma}+\sigma^{-3}\sum_{i=1}^{n}\left(y_i-\beta x_i\right)^2
\end{aligned}$$

Then, we compute the second order derivatives,

$$\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta^2} &= -\frac{\sum_{i=1}^{n}x_i^2}{\sigma^2}\\
\frac{\partial^2 \ell}{\partial \sigma^2} &= \frac{n}{\sigma^2}-3\sigma^{-4}\sum_{i=1}^{n}\left(y_i-\beta x_i\right)^2\\
\frac{\partial^2 \ell}{\partial \beta\partial\sigma} &= \frac{\partial^2 \ell}{\partial \sigma\partial\beta} = 2\sigma^{-3}\left(\beta\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_iy_i\right)
\end{aligned}$$

Finally, we calculate the negative expectations to get the entries of the Fisher information matrix:

$$\begin{aligned}
-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \beta^2}\right] &= -\mathbb{E}\left[-\frac{\sum_{i=1}^{n}x_i^2}{\sigma^2}\right] = \frac{\sum_{i=1}^{n}x_i^2}{\sigma^2}\\
-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \sigma^2}\right] &= -\mathbb{E}\left[\frac{n}{\sigma^2}-3\sigma^{-4}\sum_{i=1}^{n}\left(y_i-\beta x_i\right)^2\right]\\
&= -\frac{n}{\sigma^2}+3\sigma^{-4}\mathbb{E}\left[\sum_{i=1}^{n}\epsilon_i^2\right]\\
&\text{(note that } \mathbb{E}(\epsilon^2)=\mathrm{Var}(\epsilon)+(\mathbb{E}(\epsilon))^2=\sigma^2+0^2=\sigma^2\text{)}\\
&= -\frac{n}{\sigma^2}+\frac{3n}{\sigma^2}=\frac{2n}{\sigma^2}\\
-\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \sigma\beta}\right] &= -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \beta\sigma}\right] = -\mathbb{E}\left[2\sigma^{-3}\left(\beta\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_iy_i\right)\right]\\
&= 2\sigma^{-3}\left(\beta\sum_{i=1}^{n}x_i^2-\mathbb{E}\left[\sum_{i=1}^{n}x_i\left(\beta x_i+\epsilon_i\right)\right]\right)\\
&= 2\sigma^{-3}\left(\beta\sum_{i=1}^{n}x_i^2-\beta\sum_{i=1}^{n}x_i^2+\mathbb{E}\left[\sum_{i=1}^{n}x_i\epsilon_i\right]\right) = 2\sigma^{-3}\cdot 0 = 0
\end{aligned}$$

Collecting the results we form the matrix:

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

e)

In the last part, we need to find the Cramer-Rao lower bound for unbiased estimator of $\frac{\beta}{\sigma}$. We make use of the hint, namely, if T is unbiased for the scalar valued function $g(\theta) = \frac{\beta}{\sigma}$,$\theta = (\beta, \sigma)$, then under usual regularity conditions,

$$\mathbb{V}(T) \geqslant (\nabla g(\theta))' I(\theta)^{-1} \nabla g(\theta)$$

We first need the gradient of g(θ)

$$\nabla g(\beta, \sigma) = \left(\frac{\partial g}{\partial \beta}, \frac{\partial g}{\partial \sigma}\right) = (1/\sigma, -\beta/\sigma^2)$$

Now, we have everything neccessary to calculate the Cramer-Rao lower bound. Using the inverse of the Fisher information matrix from part d),

$$\mathbb{V}(T) \geqslant (1/\sigma, -\beta/\sigma^2) \begin{bmatrix} \frac{\sigma^2}{\sum_{i=1}^n x_i^2} & 0 \\ 0 & \frac{\sigma^2}{2n} \end{bmatrix} \begin{bmatrix} 1/\sigma \\ -\beta/\sigma^2 \end{bmatrix} = \frac{1}{\sum_{i=1}^n x_i^2} + \frac{-\beta^2}{\sigma^2 2n}$$

which is the Cramer-Rao lower bound.

Exercise 68

We have an i.i.d sample $X_1, X_2, \ldots, X_n$ from a Poisson distribution with mean λ, and let $T = \sum_{i=1}^n X_i$

a)

First we need to show that distribution $X_1, X_2, \ldots, X_n$ is independent of λ, i.e. T is sufficient for λ. The joint probability of a sample given T is

$$P(X_1, X_2, \ldots, X_n \mid T = t) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, T = t)}{P(T = t)}$$

First we determine the marginal distribution. Because T = $\sum_{i=1}^n X_i$ and $X_i \sim P(\lambda)$, the probability mass function of T is

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

For the joint distribution, provided that $\sum_{i=1}^n x_i = T$,

$$\begin{aligned} P(X_1 = x_1, \ldots, X_n = x_n, T = t) &= \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \qquad \text{(note that the x exponents sum up to T)} \\ &= e^{-n\lambda} \frac{\lambda^t}{x_1! \cdot \ldots \cdot x_n!} \end{aligned}$$

Finally, we return back to the initial expression for the conditional probability,

$$
\begin{aligned}
P(X_1, X_2, \ldots, X_n \mid T = t) &= \frac{P(X_1 = x_1, \ldots, X_n = x_n, T = t)}{P(T = t)} \\
&= \frac{t!}{(n\lambda)^t e^{-n\lambda}} \times e^{-n\lambda} \frac{\lambda^t}{x_1! \cdot \ldots \cdot x_n!} \\
&= \frac{t!}{n^t x_1! \cdot \ldots \cdot x_n!}
\end{aligned}
$$

Clearly, the conditional probability does not depend on λ, meaning that T is indeed a sufficient statistic for λ.

b)

Now, suppose X_1 is not sufficient statistic for λ. Then following conditional probability will depend on λ

$$
\begin{aligned}
P(X_1, X_2, \ldots, X_n \mid X_1 = x_1) &= \frac{P(X_1 = x_1, \ldots, X_n = x_n, X_1 = x_1)}{P(X_1 = x_1)} \\
&= \frac{\prod_{i=1}^n P(X_i = x_i)}{P(X_1 = x_1)} \\
&= \prod_{i=2}^n P(X_i = x_i) \\
&= \prod_{i=2}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\
&= e^{-\lambda(n-1)} \frac{\lambda^{\sum_{i=2}^n x_i}}{x_2! \cdot \ldots \cdot x_n!}
\end{aligned}
$$

In this case, the lambda the conditional distribution depends on λ, so X_1 is not sufficient statistics.

c)

We use the factorization theorem to show that T is sufficient statistics for λ. The theorem says that for T to be sufficient statistics for λ, the joint distribution can be factorized as following:

$$
f(x_i, \ldots, x_n) = g(T, \lambda) h(x_1, \ldots, x_n)
$$

The joint density needs to be factorized in a way such that g depends on T and λ and h on $x_1, \ldots, x_n$. Recall that our joint density is:

$$
\begin{aligned}
f(x_i, \ldots, x_n \mid \lambda) &= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\
&= e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!} \\
&= e^{-n\lambda} \lambda^T \prod_{i=1}^n \frac{1}{x_i!}
\end{aligned}
$$

We can see that the factorization is possible, with $g(T, \lambda) = e^{-n\lambda} \lambda^T$ and $h = \prod_{i=1}^n \frac{1}{x_i!}$.

Exercise 69

First, to show that $\prod_{i=1}^n X_i$ and $\sum_{i=1}^n X_i$ are sufficient statistics for gamma distribution, recall that a joint gamma distribution is

$$f(x_i, \dots, x_n \mid \alpha, \beta) = \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i}$$

Now, let $(T_1, T_2) = (\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$. We can try to use the factorization theorem from the last exercise, by re-arranging the joint density in a following way:

$$\begin{aligned} f(x_i, \dots, x_n \mid \alpha, \beta) &= \prod_{i=1}^n \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} \\ &= \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} e^{-\beta \sum_{i=1}^n x_i} \left(\prod_{i=1}^n x_i \right)^{\alpha-1} && \text{(recall } T_1, T_2\text{)} \\ &= \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} e^{-\beta T_2} T_1^{\alpha-1} \end{aligned}$$

Then the expression above is consistent with the factorization theorem:

$$g((T_1, T_2), \alpha, \beta) h(x_1, \dots, x_n) = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} e^{-\beta T_2} T_1^{\alpha-1}$$

where

$$g((T_1, T_2), \alpha, \beta) = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n} e^{-\beta T_2} T_1^{\alpha-1}$$

and

$$h(x_1, \dots, x_n) = 1$$

Therefore, the statistics $\prod_{i=1}^n X_i$ and $\sum_{i=1}^n X_i$ are sufficient for (α, β)

Exercise 76

a)

Recall that $\Lambda_i = \frac{P(x_i|H_0)}{P(x_i|H_A)}$. Then we have the following:

$$\begin{aligned} \Lambda_1 &= \frac{0.2}{0.1} = 2 \\ \Lambda_2 &= \frac{0.3}{0.4} = \frac{3}{4} \\ \Lambda_3 &= \frac{0.3}{0.1} = 3 \\ \Lambda_4 &= \frac{0.2}{0.4} = \frac{1}{2} \end{aligned}$$

Then the ordering is given by $\Lambda_3 > \Lambda_1 > \Lambda_2 > \Lambda_4$

b)

Given a level α, we have that:

$$\alpha = P(\text{Reject } H_0|H_0) = P(\Lambda < \Lambda_i|H_0)$$

Now, conditional on H_0 we have the following:

$$\begin{aligned}
P(\Lambda < \Lambda_3|H_0) &= P(x \in \{x_1, x_2, x_4\}|H_0) = 0.2 + 0.3 + 0.2 &&= 0.7 \\
P(\Lambda < \Lambda_1|H_0) &= P(x \in \{x_2, x_4\}|H_0) = 0.3 + 0.2 &&= 0.5 \\
P(\Lambda < \Lambda_2|H_0) &= P(x \in \{x_4\}|H_0) &&= 0.2 \\
P(\Lambda < \Lambda_4|H_0) &= P(x \in \emptyset|H_0) &&= 0
\end{aligned}$$

Then we have that taking $\alpha = 0.2$, we reject the null in $\{x_4\}$ and accept in $\{x_1, x_2, x_3\}$. However, taking $\alpha = 0.5$, we reject the null in $\{x_4, x_2\}$ and accept in $\{x_1, x_3\}$.

c)

Note we reject H_0 when

$$\frac{P(x|H_0)}{P(x|H_1)} < \frac{P(H_A)}{P(H_0)} = 1$$

Notice that $\Lambda_2, \Lambda_4 < 1$ and $\Lambda_1, \Lambda_3 > 1$, so we reject at $\{x_2, x_4\}$ and accept at $\{x_1, x_3\}$. In other words, x_1 and x_3 favor H_0.

d)

Note that we reject H_0 when:

$$P(x|H_0)P(H_0) \leq P(x|H_A)P(H_A) \iff \frac{P(x|H_0)P(H_0)}{P(x|H_A)P(H_A)} \leq 1 \iff \Lambda_x \frac{P(H_0)}{P(H_A)} \leq 1$$

Taking $\alpha = 0.2$, in b) we saw we reject in $\{x_4\}$, then:

$$\Lambda_4 \frac{P(H_0)}{P(H_A)} \leq 1 \iff \frac{1}{2}\frac{P(H_0)}{P(H_A)} \leq 1 \iff \frac{P(H_0)}{P(H_A)} \leq 2$$

Now we accept in $\{x_1, x_2, x_3\}$, and recall $\Lambda_2 < \Lambda_1 < \Lambda_3$, so our lower bound is defined by Λ_2:

$$\Lambda_2 \frac{P(H_0)}{P(H_A)} > 1 \iff \frac{3}{4}\frac{P(H_0)}{P(H_A)} > 1 \iff \frac{P(H_0)}{P(H_A)} > \frac{4}{3}$$

Therefore for $\alpha = 0.2$ we need $\frac{P(H_0)}{P(H_A)} \in \left(\frac{4}{3}, 2\right]$.

Following a similar logic for $\alpha = 0.5$, in b) we saw we reject in $\{x_4, x_2\}$, and since $\Lambda_2 > \Lambda_4$ then:

$$\Lambda_2 \frac{P(H_0)}{P(H_A)} \leq 1 \iff \frac{3}{4}\frac{P(H_0)}{P(H_A)} \leq 1 \iff \frac{P(H_0)}{P(H_A)} \leq \frac{4}{3}$$

And we accept in $\{x_1, x_3\}$, and recall $\Lambda_1 < \Lambda_3$, so:

$$\Lambda_1 \frac{P(H_0)}{P(H_A)} > 1 \iff 2\frac{P(H_0)}{P(H_A)} > 1 \iff \frac{P(H_0)}{P(H_A)} > \frac{1}{2}$$

Therefore for $\alpha = 0.5$ we need $\frac{P(H_0)}{P(H_A)} \in \left(\frac{1}{2}, \frac{4}{3}\right]$.

Exercise 77

a)

FALSE: The significance level represents the probability of rejecting the null hypothesis when it is true.

b)

FALSE: If we lower the level, then we reject the null less often. This means that we might accept the null more often when it is not true, effectively increasing β. Since the power is $1 - \beta$, and β increases, then the power also decreases.

c)

FALSE: α does not measure the probability of the null hypothesis being true or not, just how probable we are to reject it assuming it is true.

d)

FALSE: The power is the probability of correctly rejecting the null.

e)

FALSE: The statistic might fall in the rejection region and correctly reject the null hypothesis, incurring in no Type 1 error whatsoever.

f)

FALSE: There is no way to measure whether accepting something that is wrong or rejecting something that is right is more serious than the other. The seriousness of the error might depend on the context, so generalizing this is not good.

g)

FALSE: The power is just the probability of correctly rejecting the null assuming the alternative is true. This means it rather depends on the statistic conditional on its distribution conditional on the alternative hypothesis.

h)

TRUE: It depends on observations of random variables, and as such it is not deterministic. Its value changes from sample to sample, and that's why we incur in the two types of errors.

Exercise 78

a)

Consider the coin tossing scenario where the coin is tossed until the first head is observed and let X be the total number of tosses. This represents a geometric distribution with success probability p. The probability of observing x tosses is given by:

$$P(X = x|p) = (1-p)^{x-1}p$$

Given the prior probabilities $P(H_0)$ and $P(H_1)$, the posterior odds in favor of H_0 against H_1 are given by:

$$\frac{P(H_0|X)}{P(H_1|X)} = \frac{P(H_0) \cdot P(X|H_0)}{P(H_1) \cdot P(X|H_1)}$$

We accept H_0 if $P(H_0|X) > P(H_1|X)$, which is equivalent to:

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

Now, before assuming equal probabilities for H_0, H_1, let's look at the general case:

$$\begin{aligned}
\frac{P(H_0|X)}{P(H_1|X)} > 1 &\iff \frac{P(X|H_0)}{P(X|H_1)} > \frac{P(H_1)}{P(H_0)} \\
&\iff \frac{(1-p_0)^{x-1}p_0}{(1-p_1)^{x-1}p_1} > \frac{P(H_1)}{P(H_0)} \\
&\iff \left(\frac{1-p_0}{1-p_1}\right)^{x-1} > \frac{p_1}{p_0}\frac{P(H_1)}{P(H_0)} \\
&\iff (x-1)ln\left(\frac{1-p_0}{1-p_1}\right) > ln\left(\frac{p_1}{p_0}\frac{P(H_1)}{P(H_0)}\right) \\
&\iff x > \frac{ln\left(\frac{p_1}{p_0}\frac{P(H_1)}{P(H_0)}\right)}{ln\left(\frac{1-p_0}{1-p_1}\right)} + 1
\end{aligned}$$

Setting $p_0 = 0.5$ and $p_1 = 0.7$ and assuming $P(H_0) = P(H_1)$, we find the critical value to accept H_0 with our LRT:

```r
p0 <- 0.5
p1 <- 0.7
log(p1/p0)/log((1-p0)/(1-p1))+1
```

```
## [1] 1.658683
```

Therefore we accept the null if we observe a value of X larger than 1.658683.

b)

It's enough to substitute $\frac{P(H_1)}{P(H_0)} = \frac{1}{10}$ in the inequality from part a) to obtain that:

```r
log(p1/p0 * 1/10)/log((1-p0)/(1-p1))+1
```

```
## [1] -2.848892
```

We accept the null hypothesis for any value of X as these are all non-negative.

c)

The significance level of the test that rejects H_0 if $X \geq 8$ is the probability of rejecting H_0 when it is true, which is:

$$\begin{aligned}
P(X \geq 8|H_0) &= 1 - P(X < 8|p = 0.5) \\
&= 1 - \sum_{i=1}^{7} P(X = i|p = 0.5) \\
&= 1 - \sum_{i=1}^{7} \left(\frac{1}{2}\right)^i \\
&= 1 - \frac{127}{128} = \frac{1}{128}
\end{aligned}$$

d)

The power of the test, which is the probability of correctly rejecting H_0 when H_1 is true, is given by:

$$\begin{aligned} P(X \geq 8|H_1) &= 1 - P(X < 8|p = 0.7) \\ &= 1 - \sum_{i=1}^{7} P(X = i|p = 0.7) \\ &= 1 - \sum_{i=1}^{7} (0.3)^{i-1} \cdot (0.7) \\ &= 1 - (0.7) \cdot \sum_{i=1}^{7} (0.3)^{i-1} \\ &\approx 0.000218 \end{aligned}$$

```
1 - 0.7*sum(0.3^(0:6))
```

```
## [1] 0.0002187
```

Exercise 79

Let $X_1, \ldots, X_n$ be an i.i.d. sample from a Poisson distribution. We want to test the null hypothesis $H_0 : \lambda = \lambda_0$ against the alternative hypothesis $H_A : \lambda = \lambda_A$, where $\lambda_A > \lambda_0$.

The likelihood ratio is given by:

$$\begin{aligned} \frac{L(X_1, \ldots, X_n|\lambda_0)}{L(X_1, \ldots, X_n|\lambda_A)} &= \frac{\prod_{i=1}^{n} \frac{e^{-\lambda_0} \lambda_0^{X_i}}{X_i!}}{\prod_{i=1}^{n} \frac{e^{-\lambda_A} \lambda_A^{X_i}}{X_i!}} \\ &= \left(\frac{\lambda_0}{\lambda_A}\right)^{\sum X_i} e^{-n(\lambda_A - \lambda_0)} \end{aligned}$$

For a test level α, we want to find a constant c such that:

$$\begin{aligned} \alpha &= P(\text{Reject } H_0|H_0) \\ &= P\left(\left(\frac{\lambda_0}{\lambda_A}\right)^{\sum X_i} e^{-n(\lambda_A - \lambda_0)} < c|H_0\right) \\ &= P\left(\sum_{i=1}^{n} X_i > \frac{ln(c) - n(\lambda_A - \lambda_0)}{ln(\lambda_0) - ln(\lambda_A)}|H_0\right) \\ &:= P\left(\sum_{i=1}^{n} X_i > c_\alpha|H_0\right) \end{aligned}$$

Now, $\sum_{i=1}^{n} X_i$ has a Poisson distribution with parameter $n\lambda_0$ (since we are conditioning on the null hypothesis being true). Then:

$$
\begin{aligned}
\alpha = P(\text{Reject } H_0|H_0) &\iff P\left(\sum_{i=1}^{n} X_i > c_\alpha|H_0\right) = \alpha \\
&\iff 1 - P\left(\sum_{i=1}^{n} X_i \leq c_\alpha|H_0\right) = \alpha \\
&\iff P\left(\sum_{i=1}^{n} X_i \leq c_\alpha|H_0\right) = 1-\alpha \\
&\iff c_\alpha = Q_{Poisson(n\lambda_0)}(1-\alpha)
\end{aligned}
$$

Using the properties of the Poisson distribution, we can calculate c_α and thus establish the rejection region.

Notice that c_α depends only on n, λ_0, and α and not on λ_A. This implies that the test is the same for any choice of λ_A as long as $\lambda_A > \lambda_0$, which establishes that the test is Uniformly Most Powerful (UMP) for testing $H_0 : \lambda = \lambda_0$ against the alternative $H_A : \lambda > \lambda_0$.

Exercise 80

a)

The likelihood ratio is given by:

$$
\begin{aligned}
LR &= \left(\prod_{i=1}^{n} \frac{\theta_0 \gamma^{\theta_0}}{X_i^{\theta_0+1}}\right)\left(\prod_{i=1}^{n} \frac{\theta_A \gamma^{\theta_A}}{X_i^{\theta_A+1}}\right)^{-1} \\
&= \left(\frac{\theta_0}{\theta_A}\right)^n \gamma^{n(\theta_0-\theta_A)} \prod_{i=1}^{n} X_i^{\theta_A-\theta_0} \\
&= \left(\frac{\theta_0}{\theta_A}\right)^n \prod_{i=1}^{n} \left(\frac{X_i}{\gamma}\right)^{\theta_A-\theta_0}
\end{aligned}
$$

We want to find c such that:

$$
\begin{aligned}
\alpha &= P\left(\left(\frac{\theta_0}{\theta_A}\right)^n \prod_{i=1}^{n} \left(\frac{X_i}{\gamma}\right)^{\theta_A-\theta_0} < c|\theta = \theta_0\right) \\
&= P\left(\sum_{i=1}^{n} ln\left(\frac{X_i}{\gamma}\right) < \frac{ln(c) + nln(\theta_A/\theta_0)}{\theta_A - \theta_0}|\theta = \theta_0\right) \\
&:= P\left(\sum_{i=1}^{n} ln\left(\frac{X_i}{\gamma}\right) < c_\alpha|\theta = \theta_0\right)
\end{aligned}
$$

Now, notice that:

$$
\begin{aligned}
P\left(ln\left(\frac{X_i}{\gamma}\right) < y\right) &= P\left(X_i < \gamma e^y\right) \\
&= 1 - \left(\frac{\gamma}{\gamma e^y}\right)^\theta \\
&= 1 - e^{-\theta y}
\end{aligned}
$$

Which is the CDF of an exponential distribution with parameter θ. Then $\sum_{i=1}^{n} ln\left(\frac{X_i}{\gamma}\right)$ has a $Gamma(n, \theta)$ distribution, which implies that:

$$P\left(\sum_{i=1}^{n} ln\left(\frac{X_i}{\gamma}\right) < c_\alpha | \theta = \theta_0\right) = \alpha \iff c_\alpha = Q_{Gamma(n,\theta_0)}(\alpha)$$

b)

Yes, notice the previously found c_α depends only on n, θ_0, α but not on θ_A, which means it is the same for any choice of θ_A. In other words, the LRT is UMP. Given a level α, we have that the power is:

$$\begin{aligned}
P(\text{Reject } H_0|H_1) &= P\left(\sum_{i=1}^{n} ln\left(\frac{X_i}{\gamma}\right) < c_\alpha | \theta = \theta_A\right) \\
&= F_{Gamma(n,\theta_A)}(c_\alpha) \\
&= F_{Gamma(n,\theta_A)}\left(Q_{Gamma(n,\theta_0)}(\alpha)\right)
\end{aligned}$$

Exercise 81

Let $X_1, ..., X_n \sim f(x|\theta)$ such that T is a sufficient statistic for θ. The likelihood ratio, by decomposition theorem, is given by:

$$\begin{aligned}
LR &= \frac{f(X|\theta_0)}{f(X|\theta_A)} \\
&= \frac{g(T, \theta_0) \cdot h(X)}{g(T, \theta_A) \cdot h(X)} \\
&= \frac{g(T, \theta_0)}{g(T, \theta_A)} \\
&:= s(T)
\end{aligned}$$

Which is clearly a function of T, which we call s. Now, if we know the distribution of T under H_0, and given a level α, then we want to find c such that:

$$\alpha = P(s(T) < c)$$

In the case where s is invertible, then it suffices to note that

$$\alpha = P(s(T) < c) = P(T < s^{-1}(c)) \iff s^{-1}(c) = F_T^{-1}(\alpha) \iff c = s\left(F_T^{-1}(\alpha)\right)$$

In the case where s is not invertible but the distribution of $s(T)$ is known, then it might be useful to find a pivot, after which the respective quantiles can be computed and solved for. Note that neither of the two approaches are trivial.

Exercise 82

a)

The likelihood function for a normal distribution with mean 0 and variance σ^2 for a single observation x is given by:

$$L(\sigma|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

The likelihood ratio, is the likelihood under H_0 divided by the likelihood under H_A:

$$\begin{aligned}
\Lambda(x) &= \frac{L(\sigma_0|x)}{L(\sigma_A|x)} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{x^2}{2\sigma_0^2}}}{\frac{1}{\sqrt{2\pi\sigma_A^2}} e^{-\frac{x^2}{2\sigma_A^2}}} \\
&= \frac{\sigma_A}{\sigma_0} e^{\frac{x^2}{2}\left(\frac{1}{\sigma_A^2} - \frac{1}{\sigma_0^2}\right)}
\end{aligned}$$

Note that large values of $\Lambda(x)$ favor H_0. Therefore, since $\sigma_A > \sigma_0$, $\frac{1}{\sigma_A} - \frac{1}{\sigma_0} < 0$, which makes the exponent negative. Hence, values of x close to 0 make $\Lambda(x)$ large, and will favor H_0.

Since $X\ N(0, \sigma^2)$,

The rejection region of a level α test can be determined using the distribution of X under H_0. Since $X\ N(0, \sigma^2)$, x^2/σ_0^2 follows a chi-squared distribution with 1 degree of freedom. The rejection region is then:

$$P\left(\frac{X^2}{\sigma_0^2} > \chi^2_{1,\alpha}\right) = \alpha$$

where $\chi^2_{1,\alpha}$ is the critical value of the chi-squared distribution with 1 degree of freedom that cuts off the upper α tail of the distribution.

b)

The likelihood for each observation X_i given a variance σ^2 is:

$$L(\sigma|X_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{X_i^2}{2\sigma^2}}.$$

Thus, the joint likelihood $L(\sigma|X_1, ..., X_n)$ is the product of the individual likelihoods:

$$L(\sigma|X_1, ..., X_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n X_i^2}.$$

Now, we form the likelihood ratio $\Lambda(X_1, ..., X_n)$ for the sample:

$$\Lambda(X_1, ..., X_n) = \frac{L(\sigma_0|X_1, ..., X_n)}{L(\sigma_A|X_1, ..., X_n)} = \left(\frac{\sigma_A}{\sigma_0}\right)^n e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_A^2}\right)\sum_{i=1}^n X_i^2}.$$

As in part (a), we are looking for values of $\sum_{i=1}^n X_i^2$ that are close to 0 to favor H_0 because this will make $\Lambda(X_1, ..., X_n)$ large.

To find the rejection region for a level α test, we need to determine when $\Lambda(X_1, ..., X_n)$ is sufficiently small to reject H_0. This occurs for values X_i that make $\sum_{i=1}^n X_i^2/\sigma_0^2$ move away from 0. The test statistic for the likelihood ratio test is then $\sum_{i=1}^n X_i^2/\sigma_0^2$, which, under H_0, follows a chi-squared distribution with n degrees of freedom. Therefore, the rejection region R is:

$$R = \left\{ (X_1, ..., X_n) : \frac{\sum_{i=1}^n X_i^2}{\sigma_0^2} > \chi^2_{n,\alpha} \right\},$$

where $\chi^2_{n,\alpha}$ is the critical value from the chi-squared distribution with n degrees of freedom that cuts off the upper α tail of the distribution. This means we reject H_0 if the observed value of $\sum_{i=1}^n X_i^2/\sigma_0^2$ exceeds $\chi^2_{n,\alpha}$.

c)

The power of a test is the probability that it correctly rejects the null hypothesis when the alternative hypothesis is true. It is calculated as:

$$\text{Power} = P(\text{Reject } H_0|\sigma = \sigma_A) = P\left(\frac{(n-1)S^2}{\sigma_0^2} > \chi^2_{n-1,\alpha}|\sigma = \sigma_A\right),$$

where $\chi^2_{n-1,\alpha}$ is the critical value from the chi-squared distribution for $n-1$ degrees of freedom that corresponds to the chosen significance level α, and $)S^2$ is the sample variance.

To be the UMP test, the test statistic used must be the most powerful for all values of σ greater than σ_0 when compared to any other test statistic that could be used at the same significance level.

For the case of testing for the variance of a normal distribution, the test based on the sample variance is indeed the UMP test. This is because of the Neyman-Pearson lemma, which states that the likelihood ratio test is the most powerful test for simple hypotheses. In this case, while the alternative hypothesis H_A is composite (since it involves all values of $\sigma > \sigma_0$), it can be shown that the test based on the sample variance is the UMP test for all values greater than σ_0 due to the monotone likelihood ratio property of the normal distribution with respect to the variance. This property ensures that the likelihood ratio test remains the most powerful test even as the variance under the alternative hypothesis increases beyond σ_0.

Exercise 83

a)

For a test to have a Type I error rate of $\alpha = 0$, we must design a test where the probability of incorrectly rejecting a true null hypothesis is zero. We can use the event where X exceeds 1 as our criterion for this continuous distribution. Under the null hypothesis ($H_0 : \theta = 1$), X is uniformly distributed over the interval $[0, 1]$. Therefore, the probability of a Type I error is:

$$\alpha = P(X > 1|H_0) = 1 - P(X \leq 1|H_0) = 1 - 1 = 0$$

For the power of the test, which is the likelihood of correctly rejecting H_0 when H_A is true, we note that under $H_A : \theta = 2$, X is uniformly distributed over $[0, 2]$. Hence, the probability that X will be more than 1 is:

$$1 - \beta = P(X > 1|H_A) = 1 - P(X \leq 1|H_A) = 1 - \frac{1}{2} = \frac{1}{2}$$

b)

Considering a test that rejects H_0 if $X \leq c$, the significance level becomes:

$$\alpha = P(X \leq c|H_0) = \frac{c - 0}{1 - 0} = c$$

Conversely, the test's power, which is the probability of rejecting H_0 when H_A is actually true, is given by:

$$1 - \beta = 1 - P(X > c|H_A) = 1 - P(X \leq c|H_A) = \frac{c-0}{2-0} = \frac{c}{2}$$

c)

The significance level if the test rejects H_0 for $1 - c \leq X \leq 1$, is:

$$\alpha = P(1 - c \leq X \leq 1|H_0) = c$$

The power of the test, or the chance that the test will reject H_0 when H_A is true, is:

$$1 - \beta = 1 - P(1 - c \leq X \leq 1|H_A) = \frac{c}{2}$$

d)

Once the cutoff point c is specified, the likelihood ratio test (LRT) establishes a specific rejection region for a particular sample and given significance level.

e)

If the roles of the null and alternative hypotheses were reversed, the tests' significance and power would also switch places, reflecting their new assumptions. Similarly, the likelihood ratio test's rejection region would adapt to the new hypothesis definitions.

Exercise 84

a)

The density of $U(0, 1)$ is $f_U(x) = 1$ on $[0, 1]$. For the Beta distribution with parameters θ and 1, we have:

$$f(x|\theta) = \theta x^{\theta-1}$$

To prove that the uniform distribution is a special case of the Beta distribution, we need to find a value of θ such that $f(x|\theta)$ is equal to 1 for all x in the interval $[0, 1]$.

By inspection, we can see that if $\theta = 1$, then $x^{\theta-1} = x^0 = 1$, because any number to the power of 0 is 1. Hence, for $\theta = 1$, we have:

$$f(x|1) = 1 \cdot x^{1-1} = x^0 = 1 \; \forall x$$

So, the probability density function of the Beta distribution when $\theta = 1$ is constant and equal to 1 for all x in $[0, 1]$, which is exactly the definition of the uniform distribution on $[0, 1]$.

b)

The likelihood of a Beta distribution given an i.i.d. sample $X_1, X_2, ..., X_n$ is given by

$$L(\theta|X_1, X_2, ..., X_n) = \prod_{i=1}^{n} f(X_i|\theta) = \prod_{i=1}^{n} \theta X_i^{\theta-1}$$

while the likelihood for a uniform distribution has likelihood

$$L_0 = \prod_{i=1}^{n} f_U(x) = \prod_{i=1}^{n} 1 = 1.$$

Then, the generalized likelihood ratio Λ is

$$\Lambda = \frac{L_0}{\sup_\theta L(\theta|X_1, X_2, ..., X_n)} = \frac{1}{\sup_\theta \prod_{i=1}^{n} \theta X_i^{\theta-1}}$$

To find the supremum of L, we take its log-likelihood, differentiate it with respect to θ and set it to 0, which gives us the MLE of θ.

$$LL(\theta|X_1, X_2, ..., X_n) = n \cdot \ln(\theta) + (\theta - 1) \sum_{i=1}^{n} \ln(X_i) \frac{d}{d\theta} LL(\theta|X_1, X_2, ..., X_n) = \frac{n}{\hat{\theta}} + \sum_{i=1}^{n} \ln(X_i) = 0 \implies \hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \ln(X_i)}$$

Substituting $\hat{\theta}$ we get

$$\begin{aligned} LL(\hat{\theta}|X_1, X_2, ..., X_n) &= n \ln\left(-\frac{n}{\sum_{i=1}^{n} \ln(X_i)}\right) + \left(-\frac{n}{\sum_{i=1}^{n} \ln(X_i)} - 1\right) \sum_{i=1}^{n} \ln(X_i) \\ &= n \ln\left(-\frac{n}{\sum_{i=1}^{n} \ln(X_i)}\right) - n - \sum_{i=1}^{n} \ln(X_i) \end{aligned}$$

Taking the log of Λ and substituting $LL(\hat{\theta})$

$$\begin{aligned} L(\Lambda) &= LL_0 - LL(\hat{\theta}|X_1, X_2, ..., X_n) \\ &= -n \ln\left(-\frac{n}{\sum_{i=1}^{n} \ln(X_i)}\right) + n + \sum_{i=1}^{n} \ln(X_i) \end{aligned}$$