ESTIMATION - DETECTION

TD 1 —Estimation — SOLUTIONS

Exercise 1

1. Maximum likelihood estimation

1.(a)

$$\begin{aligned}
f(x_1,\cdots,x_n;\beta,\theta) &= \frac{\prod_{i=1}^n x_i^{\beta-1}}{\Gamma(\beta)^n\theta^{n\beta}}\exp\left(-\frac{\sum_{i=1}^n x_i}{\theta}\right)\\
\ln f(x_1,\cdots,x_n;\beta,\theta) &= -n\beta\ln\theta - n\ln\Gamma(\beta) + (\beta-1)\sum_{i=1}^n \ln x_i - \frac{\sum_{i=1}^n x_i}{\theta}\\
\frac{\partial \ln f(x_1,\cdots,x_n;\beta,\theta)}{\partial\theta} &= -\frac{n\beta}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} \overset{!}{=} 0 \qquad \Longrightarrow \theta = \frac{\sum_{i=1}^n x_i}{n\beta} = \frac{\bar{x}}{\beta}\\
\frac{\partial^2 \ln f(x_1,\cdots,x_n;\beta,\theta)}{\partial\theta^2} &= \frac{n\beta}{\theta^2} - 2\frac{\sum_{i=1}^n x_i}{\theta^3}\\
\left.\frac{\partial^2 \ln f(x_1,\cdots,x_n;\beta,\theta)}{\partial\theta^2}\right|_{\theta=\bar{x}/\beta} &= \frac{n\beta}{\bar{x}^2/\beta^2} - 2\frac{n\bar{x}}{\bar{x}^3/\beta^3} = -\frac{n\beta^3}{\bar{x}^2} < 0\\
&\Longrightarrow \hat{\theta}_{ML} = \frac{\sum_{i=1}^n x_i}{n\beta} = \frac{\bar{x}}{\beta}
\end{aligned}$$

1.(b)

$$\begin{aligned}
\mathbb{E}\left[\hat{\theta}_{ML}\right] &= \mathbb{E}\left[\frac{\sum_{i=1}^n x_i}{n\beta}\right] = \frac{\sum_{i=1}^n \mathbb{E}[x_i]}{n\beta} = \frac{n\beta\theta}{n\beta} = \theta \qquad \Longrightarrow \text{unbiased}\\
\text{Var}\left[\hat{\theta}_{ML}\right] &= \text{Var}\left[\frac{\sum_{i=1}^n x_i}{n\beta}\right] = \frac{\sum_{i=1}^n \text{Var}[x_i]}{n^2\beta^2} = \frac{n\beta\theta^2}{n^2\beta^2} = \frac{\theta^2}{n\beta} \qquad \Longrightarrow \text{convergent}\\
I(\theta) &= \mathbb{E}\left[-\frac{\partial^2 \ln f(x_1,\cdots,x_n;\beta,\theta)}{\partial\theta^2}\right] = -\frac{n\beta}{\theta^2} + 2\frac{\sum_{i=1}^n \mathbb{E}[x_i]}{\theta^3} = \frac{n\beta}{\theta^2}\\
CRB(\theta) &= \frac{\theta^2}{n\beta} \qquad = \text{Var}\left[\hat{\theta}_{ML}\right] \Longrightarrow \hat{\theta}_{ML} \text{ is the efficient estimator for } \theta
\end{aligned}$$

1.(c)

$$\begin{aligned}
\ln f(x_1,\cdots,x_n;\beta,\theta) &= -n\beta\ln\theta - n\ln\Gamma(\beta) + (\beta-1)\sum_{i=1}^n \ln x_i - \frac{\sum_{i=1}^n x_i}{\theta}\\
\frac{\partial \ln f(x_1,\cdots,x_n;\beta,\theta)}{\partial\beta} &= -n\ln\theta - n\frac{\Gamma'(\beta)}{\Gamma(\beta)} + \sum_{i=1}^n \ln x_i \overset{!}{=} 0
\end{aligned}$$

This is a nonlinear equation in β that does not have a closed-form solution (it is, however, numerically very well behaved and numerical solutions are easy to obtain.)

2. Method of moments

$$\begin{aligned}
m_1 &= \mathbb{E}[X] = \beta\theta\\
m_2 &= \mathbb{E}[X^2] = \text{Var}[X] + \mathbb{E}[X]^2 = \beta\theta^2 + \beta^2\theta^2 = \theta^2\beta(1+\beta)
\end{aligned}$$

Substituting $\beta = \frac{m_1}{\theta}$ into the expression for m_2 we obtain

$$m_2 = \theta^2\frac{m_1}{\theta}\left(\frac{m_1}{\theta}+1\right) = m_1^2 + m_1\theta \longrightarrow \theta = \frac{m_2 - m_1^2}{m_1}$$

and

$$\beta = \frac{m_1^2}{m_2 - m_1^2}.$$

Replacing m_1 and m_2 with sample moments, we obtain the estimators

$$\hat{\theta} = \frac{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}{\sum_{i=1}^n x_i}$$
$$\hat{\beta} = \frac{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}$$

We conclude that it is easy to determine estimators for both parameters with the method of moments, while maximum likelihood estimates can only be obtained numerically.

3. Bayesian estimation with inverse Gamma prior $\mathcal{IG}(k,\tau)$ for θ: $\theta \sim \mathcal{IG}(k,\tau)$

3.(a)

$$\begin{aligned}
f(x_1,\cdots,x_n;\beta,\theta) &= \frac{\prod_{i=1}^n x_i^{\beta-1}}{\Gamma(\beta)^n\theta^{n\beta}}\exp\left(-\frac{\sum_{i=1}^n x_i}{\theta}\right)\\
p_\theta(\theta;k,\tau) &= \frac{\tau^k}{\Gamma(k)}\theta^{-k-1}\exp\left(-\tau/\theta\right)\\
f(\theta;x,\beta,k,\tau) &= \frac{\prod_{i=1}^n x_i^{\beta-1}}{\Gamma(\beta)^n\theta^{n\beta}}\exp\left(-\frac{\sum_{i=1}^n x_i}{\theta}\right)\frac{\tau^k}{\Gamma(k)}\theta^{-k-1}\exp\left(-\tau/\theta\right)\\
&= \underbrace{\frac{\tau^k\prod_{i=1}^n x_i^{\beta-1}}{\Gamma(k)\Gamma(\beta)^n}}_{C}\theta^{-(n\beta+k)-1}\exp\left(-\frac{\sum_{i=1}^n x_i+\tau}{\theta}\right)\\
&= \mathcal{IG}\left(n\beta+k,\sum_{i=1}^n x_i+\tau\right)
\end{aligned}$$

3.(b)

$$\begin{aligned}
\ln f(\theta;x,\beta,k,\tau) &= \ln C - (n\beta+k+1)\ln\theta - \frac{\sum_{i=1}^n x_i+\tau}{\theta}\\
\frac{\partial \ln f(\theta;x,\beta,k,\tau)}{\partial\theta} &= -\frac{n\beta+k+1}{\theta} + \frac{\sum_{i=1}^n x_i+\tau}{\theta^2} \quad \stackrel{!}{=} 0\\
\hat{\theta}_{MAP} &= \frac{\sum_{i=1}^n x_i+\tau}{n\beta+k+1}
\end{aligned}$$

(One verifies that $\left.\frac{\partial^2 \ln f(\theta;x,\beta,k,\tau)}{\partial\theta^2}\right|_{\theta=\frac{\sum_{i=1}^n x_i+\tau}{n\beta+k+1}} < 0$.)

3.(c) For $X \sim \mathcal{IG}(k,\tau)$ we have

$$\begin{aligned}
\mathbb{E}[X] &= \int x\frac{\tau^k}{\Gamma(k)}x^{-k-1}\exp\left(-\tau/x\right)dx = \int \frac{\tau^k}{\Gamma(k)}x^{-(k-1)-1}\exp\left(-\tau/x\right)dx\\
&= \tau\frac{\Gamma(k-1)}{\Gamma(k)}\underbrace{\int\frac{\tau^{k-1}}{\Gamma(k-1)}x^{-(k-1)-1}\exp\left(-\tau/x\right)dx}_{=\int\mathcal{IG}(k-1,\tau)\,=\,1} = \tau\frac{\Gamma(k-1)}{\Gamma(k)} = \frac{\tau}{k-1}.
\end{aligned}$$

Since the posterior law of θ is $\mathcal{IG}\left(n\beta+k,\sum_{i=1}^n x_i+\tau\right)$, we immediately have

$$\hat{\theta}_{MMSE} = \frac{\sum_{i=1}^n x_i+\tau}{n\beta+k+1}.$$

Exercise 2

1.(a) We have that

$$y_i \sim f_i(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)^2}{2\sigma^2}\right)$$

and the likelihood and log-likelihood, respectively, are therefore given by

$$f(y_1, \cdots, y_n; \boldsymbol{\beta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)^2}{2\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)^2}{2\sigma^2}\right)$$

$$\ln f(x_1, \cdots, x_n; \boldsymbol{\beta}) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)^2}{2\sigma^2},$$

and the first and second order partial derivatives of the log-likelihood read

$$\frac{\partial \ln f(x_1, \cdots, x_n; \boldsymbol{\beta})}{\partial \beta_k} = -\frac{1}{\sigma^2}\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)(-x_{ik}) = -\frac{1}{\sigma^2}\left(\sum_{i=1}^n\sum_{j=1}^J x_{ik}x_{ij}\beta_j - \sum_{i=1}^n x_{ik}y_i\right)$$

$$\frac{\partial^2 \ln f(x_1, \cdots, x_n; \boldsymbol{\beta})}{\partial \beta_k^2} = -\frac{1}{\sigma^2}\sum_{i=1}^n x_{ik}^2 < 0.$$

By setting the first order partial derivative to zero and rearranging, we obtain the condition

$$\sum_{i=1}^n\sum_{j=1}^J x_{ik}x_{ij}\beta_j = \sum_{i=1}^n x_{ik}y_i,$$

which, in the LS context, are called the *normal equations*. In matrix form, they read

$$(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{y},$$

from which the maximum likelihood estimate follows as

$$\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{y}).$$

Note: Already from the expression of the (log-)likelihood, it is clear that the maximum likelihood and least squares estimate are identical since maximizing the (log-)likelihood with respect to $\boldsymbol{\theta}$ implies minimizing the sulm of squared residuals $\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)^2$.

1.(b) The first and second order partial derivative of the log-likelihood reads

$$\frac{\partial \ln f(x_1, \cdots, x_n; \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j\right)^2$$

$$\frac{\partial^2 \ln f(x_1, \cdots, x_n; \sigma^2)}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j\right)^2$$

Setting the first derivative to zero, we obtain

$$\hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j\right)^2$$

and plugging this expression in the second derivative, we find

$$\left.\frac{\partial^2 \ln f(x_1, \cdots, x_n; \sigma^2)}{\partial (\sigma^2)^2}\right|_{\sigma^2 = \hat{\sigma}^2_{ML}} \propto \frac{n}{2}\hat{\sigma}^2_{ML} - \sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j\right)^2 =$$

$$= \frac{n}{2}\frac{1}{n}\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j\right)^2 - \sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j\right)^2 = -\frac{1}{2}\sum_{i=1}^n \left(y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j\right)^2 < 0,$$

so that $\hat{\sigma}^2_{ML}$ is indeed the maximum likelihood estimator for σ^2. Defining the residual $r_i = y_i - \sum_{j=1}^J x_{ij}\hat{\beta}_j$, we see that this is the maximum likelihood estimator for σ^2 of $r_i \sim \mathcal{N}(0, \sigma^2)$.

1.(c) The solution does not change in any other way than the design variables x_{ij} being replaced by the new design variables ϕ_{ij}, and the model remains linear in the parameters $\boldsymbol{\beta}$. Note that the linear regression model in variables ϕ_{ij} can therefore represent a *nonlinear* regression model in x_{ik}.

2.(a) We have that

$$y_i \sim f_i(y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)^2}{2\sigma_i^2}\right)$$

and the likelihood and log-likelihood, respectively, are therefore given by

$$f(y_1, \cdots, y_n; \boldsymbol{\beta}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{\left(y_i - \sum_{j=1}^J x_{ij}\beta_j\right)^2}{2\sigma_i^2}\right) = \frac{(2\pi)^{-n/2}}{\sqrt{\prod_{i=1}^n \sigma_i^2}} \exp\left(-\sum_{i=1}^n \frac{1}{2\sigma_i^2}\big(y_i - \sum_{j=1}^J x_{ij}\beta_j\big)^2\right)$$

$$\ln f(x_1, \cdots, x_n; \boldsymbol{\beta}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\big(\prod_{i=1}^n \sigma_i^2\big) - \sum_{i=1}^n \frac{1}{2\sigma_i^2}\big(y_i - \sum_{j=1}^J x_{ij}\beta_j\big)^2,$$

and the first and second order partial derivatives of the log-likelihood read

$$\frac{\partial \ln f(x_1, \cdots, x_n; \boldsymbol{\beta})}{\partial \beta_k} = -\sum_{i=1}^n \frac{1}{\sigma_i^2}\big(y_i - \sum_{j=1}^J x_{ij}\beta_j\big)(-x_{ik}) = \Big(\sum_{i=1}^n\sum_{j=1}^J \frac{1}{\sigma_i^2} x_{ik}x_{ij}\beta_j - \sum_{i=1}^n \frac{1}{\sigma_i^2} x_{ik}y_i\Big)$$

$$\frac{\partial^2 \ln f(x_1, \cdots, x_n; \boldsymbol{\beta}^2)}{\partial \beta_k} = .$$

Setting first order partial derivative to zero and rearranging, we obtain the condition

$$\sum_{i=1}^n\sum_{j=1}^J \frac{1}{\sigma_i^2} x_{ik}x_{ij}\beta_j = \sum_{i=1}^n \frac{1}{\sigma_i^2} x_{ik}y_i,$$

which, in matrix form, read

$$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y},$$

where $\boldsymbol{W}$ is diagonal with entries $w_{ii} = \frac{1}{\sigma_i^2}$. The maximum likelihood estimate follows as

$$\hat{\boldsymbol{\beta}}_{ML} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{y}).$$

2.(b) The likelihood and log-likelihood now take the form

$$f(y_1, \cdots, y_n; \boldsymbol{\beta}) = \frac{1}{(2\pi\det(\boldsymbol{\Sigma}))^{n/2}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right)$$

$$\ln f(y_1, \cdots, y_n; \boldsymbol{\beta}) \propto (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \propto (\boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{\beta}) - (\boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{\beta})$$
$$\propto (\boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}\boldsymbol{\beta}) - 2(\boldsymbol{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$

and setting the partial derivatives with respect to $\boldsymbol{\beta}$ to zero yields the GLS normal equations

$$\frac{\partial \ln f(y_1, \cdots, y_n; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} + 2\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{\beta} \overset{!}{=} 0$$

with formal solution

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}).$$

One can check that the Hessian of $\ln f(y_1, \cdots, y_n; \boldsymbol{\beta})$ is indeed negative.

3.(a) With the likelihood given in 1.(a), the posterior and log-posterior distributions read

$$p(\boldsymbol{\beta}|y_1, \cdots, y_n) \propto \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n \left(y_i - x_i\beta\right)^2}{2\sigma^2}\right) \frac{1}{2\pi\nu^2} \exp\left(-\frac{(\beta - \tilde{\beta})^2}{2\nu^2}\right)$$

$$\ln p(\boldsymbol{\beta}|y_1, \cdots, y_n) \propto -\sum_{i=1}^n \left(y_i - x_i\beta\right)^2 - \frac{\sigma^2}{\nu^2}(\beta - \tilde{\beta})^2$$

Since the likelihood and the prior are Gaussian, the posterior is Gaussian, the MMSE and MAP estimate coincide. We could therefore either determine the mean of the posterior (by completion of squares, cf. cours) or determine the maximum of the posterior using the first and second partial derivatives of the log-posterior:

$$\frac{\partial \ln p(\boldsymbol{\beta}|y_1,\cdots,y_n)}{\partial \beta} \propto \sum_{i=1}^{n} \big(y_i x_i - \beta x_i^2\big) - \frac{\sigma^2}{\nu^2}(\beta - \tilde{\beta}) \overset{!}{=} 0$$

$$\frac{\partial^2 \ln p(\boldsymbol{\beta}|y_1,\cdots,y_n)}{\partial \beta^2} \propto -\sum_{i=1}^{n} x_i^2 - \frac{\sigma^2}{\nu^2} < 0$$

from which we conclude

$$\hat{\beta}_{ML} = \hat{\beta}_{MAP} = \frac{\sum_{i=1}^{n} x_i y_i + \frac{\sigma^2}{\nu^2}\tilde{\beta}}{\sum_{i=1}^{n} x_i^2 + \frac{\sigma^2}{\nu^2}}$$

3.(b) With the likelihood given in 2.(a), the posterior and log-posterior distributions read

$$p(\boldsymbol{\beta}|y_1,\cdots,y_n) \propto \exp\left(-\sum_{i=1}^{n} \frac{1}{2\sigma_i^2}\big(y_i - x_i\beta\big)^2\right) \exp\left(-\frac{(\beta-\tilde{\beta})^2}{2\nu^2}\right)$$

$$\ln p(\boldsymbol{\beta}|y_1,\cdots,y_n) \propto -\sum_{i=1}^{n} \frac{1}{\sigma_i^2}\big(y_i - x_i\beta\big)^2 - \frac{1}{\nu^2}(\beta-\tilde{\beta})^2$$

The posterior is again Gaussian, and the MMSE and MAP estimate coincide. We determine the maximum of the posterior using the first and second partial derivatives of the log-posterior:

$$\frac{\partial \ln p(\boldsymbol{\beta}|y_1,\cdots,y_n)}{\partial \beta} \propto \sum_{i=1}^{n} \frac{1}{\sigma_i^2}\big(y_i x_i - \beta x_i^2\big) - \frac{1}{\nu^2}(\beta - \tilde{\beta}) \overset{!}{=} 0$$

$$\frac{\partial^2 \ln p(\boldsymbol{\beta}|y_1,\cdots,y_n)}{\partial \beta^2} \propto -\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i^2 - \frac{1}{\nu^2} < 0$$

from which we conclude

$$\hat{\beta}_{ML} = \hat{\beta}_{MAP} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i y_i + \frac{1}{\nu^2}\tilde{\beta}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i^2 + \frac{1}{\nu^2}}$$

3.(c) We can give an interpretation by examining the log-posteriors $\ln p(\boldsymbol{\beta}|y_1,\cdots,y_n)$: while the least squares estimates minimize the (weighted) sum of the squared residuals, the Bayesian estimators jointly minimize the (weighted) sum of square of residuals and a *regularization term* that penalizing the squared deviation from the prior mean $\hat{\beta}$ (the degree of deviation being controlled by the variance ν^2 of the prior.)
For completeness, if we had $J > 1$, we would obtain the matrix equations:

$$\big(\boldsymbol{X}^T\boldsymbol{X} + \frac{\sigma^2}{\nu^2}\boldsymbol{I}\big)\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{y} + \frac{\sigma^2}{\nu^2}\tilde{\boldsymbol{\beta}}$$

$$\hat{\boldsymbol{\beta}} = \big(\boldsymbol{X}^T\boldsymbol{X} + \frac{\sigma^2}{\nu^2}\boldsymbol{I}\big)^{-1}\big(\boldsymbol{X}^T\boldsymbol{y} + \frac{\sigma^2}{\nu^2}\tilde{\boldsymbol{\beta}}\big)$$

4.(a) *(Note: The data in 4 have been generated with $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0,\sigma_i^2)$ and $\beta = 1$.)*

$$\hat{\beta}_{ML} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{1.3 + 2\cdot 0.5 + 3\cdot 2.4 + 4\cdot 6.5 + 5\cdot 5.1}{1+4+9+16+25} = \frac{1.3+1+7.2+26+25.5}{55} = \frac{61.1}{55} \approx 1.111$$

$$\hat{\beta}_{MAP} = \frac{\sum_{i=1}^{n} x_i y_i + \frac{\sigma^2}{\nu^2}\tilde{\beta}}{\sum_{i=1}^{n} x_i^2 + \frac{\sigma^2}{\nu^2}} = \frac{61.1+30}{55+30} \approx 1.072$$

4.(b)

$$\hat{\beta}_{ML} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i y_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i^2} = \frac{1.3+0.5+2.4+6.5+5.1}{1+2+3+4+5} = \frac{15.8}{15} \approx 1.053$$

$$\hat{\beta}_{MAP} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i y_i + \frac{1}{\nu^2}\tilde{\beta}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} x_i^2 + \frac{1}{\nu^2}} = \frac{15.8+10}{15+10} = 1.032$$

Exercise 3

1.

$$f(x_1,\dots,x_n;\eta) = \frac{(2\pi\sigma^2)^{-n/2}}{\prod_{i=1}^n x_i} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i - \eta)^2\right)$$

The maximum likelihood estimate is given by the η_m that minimizes $\sum_{i=1}^n (\ln x_i - \eta)^2$,

$$\hat{\eta}_{ML} = \arg \min_{\eta\in\{\eta_1,\dots,\eta_M\}} \sum_{i=1}^n (\ln x_i - \eta)^2$$

i.e., the value for which the point $\eta_m(1,1,\dots,1)^T \in \mathbb{R}^n$ is closest to the vector of observations $(\ln x_1, \ln x_2, \dots, \ln x_n)^T$. Note that if the parameter η was not constrained to belong to the discrete set $\eta \in \{\eta_1,\dots,\eta_M\}$ but $\eta \in \mathbb{R}$ instead, the maximum likelihood estimator would be of the form

$$\tilde{\hat{\eta}}_{ML} = \frac{1}{n}\sum_{i=1}^n \ln x_i,$$

(since $\ln f(x_1,\dots,x_n;\eta) \propto \sum_{i=1}^n (\ln x_i - \eta)^2$ and $\frac{\partial \ln f(x_1,\dots,x_n;\eta)}{\partial \eta} \propto \sum_{i=1}^n \ln x_i - n\eta$). If we set $y_i = \ln x_i$, the above expressions equal the maximum likelihood estimators in the Gaussian case. Indeed, for $x_i \sim \ln\mathcal{N}(\eta,\nu^2)$, $y_i = \ln x_i \sim \mathcal{N}(\eta,\nu^2)$, which illustrates another property of maximum likelihood estimators: if the data are transformed by a one-to-one function that does not depend on the parameters to be estimated (in our case, the logarithm), we find the same expressions for the maximum likelihood estimators with transformed variables.

2.

$$f(\eta|x_1,\dots,x_n) \propto \prod_{m=1}^M p_m \mathbf{1}_{\eta=\eta_m} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i - \eta)^2\right)$$

hence $\hat{\eta}_{MAP}$ is given by the $\eta_m \in \{\eta_1,\dots,\eta_M\}$ with m determined by the expression

$$\arg \max_{m\in\{1,\dots,M\}} p_m \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (\ln x_i - \eta_m)^2\right)$$

or, equivalently,

$$\arg \min_{m\in\{1,\dots,M\}} -\ln p_m + \sum_{i=1}^n (\ln x_i - \eta_m)^2.$$

If $p = \frac{1}{M}$, this gives the ML estimator above.

3. The MMSE estimator does not take discrete values $\{\eta_1,\dots,\eta_M\}$ and is hence not adequate for this problem.

Exercise 4

1.

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{x_m}^{\infty} x\alpha x_m^{\alpha} x^{-(\alpha+1)}dx = \alpha x_m^{\alpha}\int_{x_m}^{\infty} x^{-\alpha}dx = \alpha x_m^{\alpha}\left.\frac{x^{-\alpha+1}}{1-\alpha}\right|_{x_m}^{\infty} \\
&= \alpha x_m^{\alpha}\left(\frac{x_m^{-\alpha+1}}{\alpha-1} + \frac{1}{1-\alpha}\lim_{n\to\infty} x^{1-\alpha}\right) \\
&= \begin{cases} \frac{\alpha x_m}{\alpha-1} & \text{if } \alpha > 1 \\ \infty & \text{if } \alpha \le 1 \end{cases} \\
\mathbb{E}[X^2] &= \int_{x_m}^{\infty} x^2\alpha x_m^{\alpha} x^{-(\alpha+1)}dx = \alpha x_m^{\alpha}\int_{x_m}^{\infty} x^{-(\alpha-1)}dx = \alpha x_m^{\alpha}\left.\frac{x^{-\alpha+2}}{2-\alpha}\right|_{x_m}^{\infty} \\
&= \alpha x_m^{\alpha}\left(\frac{x_m^{-\alpha+2}}{\alpha-2} + \frac{1}{2-\alpha}\lim_{n\to\infty} x^{2-\alpha}\right) \\
&= \begin{cases} \frac{\alpha x_m^2}{\alpha-2} & \text{if } \alpha > 2 \\ \infty & \text{if } \alpha \le 2. \end{cases}
\end{aligned}
$$

For $\alpha > 2$,

$$
\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{\alpha x_m^2}{\alpha-2} - \frac{\alpha^2 x_m^2}{(\alpha-1)^2} = x_m^2\frac{\alpha(\alpha-1)^2 - \alpha^2(\alpha-2)}{(\alpha-1)^2(\alpha-2)} = \frac{\alpha x_m^2}{(\alpha-1)^2(\alpha-2)}
$$

i.e.,

$$
\text{Var}[X] = \begin{cases} \frac{\alpha x_m^2}{(\alpha-1)^2(\alpha-2)} & \text{if } \alpha > 2 \\ \infty & \text{if } \alpha \le 2. \end{cases}
$$

2.

$$
f(x_1,\cdots,x_n;\alpha,x_m) = \alpha^n x_m^{n\alpha}\prod_{i=1}^{n} x^{-(\alpha+1)}
$$

$$
\ln f(x_1,\cdots,x_n;\alpha,x_m) = n\ln\alpha + n\alpha\ln x_m - (\alpha+1)\sum_{i=1}^{n}\ln x_i
$$

$$
\frac{\partial \ln f(x_1,\cdots,x_n;\alpha,x_m)}{\partial x_m} = \frac{n\alpha}{x_m} > 0
$$

Since $\ln x_m$ is a monotonously increasing function of x_m and $x_i \ge x_m$ we have

$$
\hat{x}_{m,ML} = \min_i x_i.
$$

$$
\begin{aligned}
\frac{\partial \ln f(x_1,\cdots,x_n;\alpha,x_m)}{\partial \alpha} &= \frac{n}{\alpha} + n\ln x_m - \sum_{i=1}^{n}\ln x_i \stackrel{!}{=} 0 \\
\hat{\alpha}_{ML} &= \frac{n}{\sum_{i=1}^{n}\ln x_i - n\ln\hat{x}_{m,ML}} = \frac{n}{\sum_{i=1}^{n}\ln x_i - n\ln\min_i x_i}.
\end{aligned}
$$

We verify that $\frac{\partial^2 \ln f(x_1,\cdots,x_n;\alpha,x_m)}{\partial\alpha^2} < 0$ and $\frac{\partial^2 \ln f(x_1,\cdots,x_n;\alpha,x_m)}{\partial x_m^2} < 0$ (cf. 3.).

3.

$$
\begin{aligned}
\frac{\partial^2 \ln f(x_1,\cdots,x_n;\alpha,x_m)}{\partial\alpha^2} &= -\frac{n}{\alpha} \\
\frac{\partial^2 \ln f(x_1,\cdots,x_n;\alpha,x_m)}{\partial x_m^2} &= -\frac{n\alpha}{x_m^2} \\
\frac{\partial^2 \ln f(x_1,\cdots,x_n;\alpha,x_m)}{\partial\alpha\partial x_m} &= \frac{n}{x_m}
\end{aligned}
$$

Since the off-diagonal terms $\frac{n}{x_m}$ are non-zero the ML estimates $\hat{\alpha}_{ML}$ and $\hat{x}_{m,ML}$ are not independent and the variance of $\hat{\alpha}_{ML}$ depends on whether x_m is known or not.

4.

$$\mathbb{E}[X] = \frac{\alpha x_m}{\alpha - 1} \implies \alpha = \frac{\mathbb{E}[X]}{\mathbb{E}[X] - x_m} \implies \hat{\alpha}_M = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i - n x_m} = \frac{1}{1 - \frac{x_m}{\overline{x}}}$$

5.(a) From 3., we immediately have: $p(\alpha) \propto \sqrt{\frac{n}{\alpha^2}} \propto \frac{1}{\alpha}$

5.(b)

$$f(\alpha|x_1, \cdots, x_n; x_m) \propto \alpha^n x_m^{n\alpha} \prod_{i=1}^n x^{-(\alpha+1)} \alpha^{-1} \propto \alpha^{n-1} x_m^{n\alpha} \prod_{i=1}^n x^{-\alpha}$$

$$\ln f(\alpha|x_1, \cdots, x_n; x_m) \propto (n-1)\alpha + n\alpha \ln x_m - \alpha \sum_{i=1}^n \ln x_i$$

$$\frac{\partial \ln f(\alpha|x_1, \cdots, x_n; x_m)}{\partial \alpha} = \frac{n-1}{\alpha} + n \ln x_m - \sum_{i=1}^n \ln x_i \overset{!}{=} 0 \implies \alpha = \frac{n-1}{\sum_{i=1}^n \ln x_i - n \ln x_m}$$

$$\frac{\partial^2 \ln f(\alpha|x_1, \cdots, x_n; x_m)}{\partial \alpha^2} = -\frac{n-1}{\alpha^2} < 0$$

$$\implies \hat{\alpha}_{MAP} = \frac{n-1}{\sum_{i=1}^n \ln x_i - n \ln x_m}$$

6.

$$f(\alpha|x_1, \cdots, x_n; x_m) \propto \alpha^n x_m^{n\alpha} \prod_{i=1}^n x_i^{-(\alpha+1)} \alpha^{k-1} \exp\left(-\frac{\alpha}{\theta}\right)$$

$$\propto \alpha^{n+k-1} x_m^{n\alpha} \prod_{i=1}^n x_i^{-\alpha} \exp\left(-\frac{\alpha}{\theta}\right)$$

$$\propto \alpha^{n+k-1} \exp\left(-\alpha \left(\frac{1}{\theta} + \sum_{i=1}^n \ln x_i - n \ln x_m\right)\right)$$

$$\propto \mathcal{G}\left(n+k,\ \frac{1}{\frac{1}{\theta} + \sum_{i=1}^n \ln x_i - n \ln x_m}\right)$$

$$\ln f(\alpha|x_1, \cdots, x_n; x_m) \propto (n+k-1)\ln\alpha + n\alpha \ln x_m - \alpha \sum_{i=1}^n \ln x_i - \frac{\alpha}{\theta}$$

$$\frac{\partial \ln f(\alpha|x_1, \cdots, x_n; x_m)}{\partial \alpha} = \frac{n+k-1}{\alpha} + n \ln x_m - \sum_{i=1}^n x_i - \frac{1}{\theta} \propto \overset{!}{=} 0$$

$$\implies \alpha = \frac{n+k-1}{\frac{1}{\theta} + \sum_{i=1}^n \ln x_i - n \ln x_m}$$

$$\frac{\partial^2 \ln f(\alpha|x_1, \cdots, x_n; x_m)}{\partial \alpha^2} = -\frac{n+k-1}{\alpha^2} < 0$$

$$\implies \hat{\alpha}_{MAP} = \frac{n+k-1}{\frac{1}{\theta} + \sum_{i=1}^n \ln x_i - n \ln x_m}$$

CONTROL: The mode of $\mathcal{G}(a,b)$ is given by $m = (a-1)b \implies$ $\hat{\alpha}_{MAP} = \frac{n+k-1}{\frac{1}{\theta} + \sum_{i=1}^n \ln x_i - n \ln x_m}$.

The mean of $\mathcal{G}(a,b)$ is given by $m = ab$ *(CONTROL: cf. Exercise 2.2)*

$$\implies \hat{\alpha}_{MMSE} = \frac{n+k}{\frac{1}{\theta} + \sum_{i=1}^n \ln x_i - n \ln x_m}$$

Exercise 5

1. The likelihood and log-likelihood, respectively, are given by

$$f(x_1, \cdots, x_n; \sigma^2) = \prod_{i=1}^{n} \frac{x_i}{\sigma^2} \exp\left(-\frac{x_i^2}{2\sigma^2}\right) = \frac{1}{(\sigma^2)^n} \prod_{i=1}^{n} x_i \exp\left(-\frac{x_i^2}{2\sigma^2}\right)$$

$$\ln f(x_1, \cdots, x_n; \sigma^2) = -n \ln \sigma^2 + \sum_{i=1}^{n} \ln x_i - \frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2,$$

and the first and second order partial derivatives of the log-likelihood read

$$\frac{\partial \ln f(x_1, \cdots, x_n; \sigma^2)}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} x_i^2$$

$$\frac{\partial^2 \ln f(x_1, \cdots, x_n; \sigma^2)}{\partial (\sigma^2)^2} = \frac{n}{(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^{n} x_i^2 = \frac{1}{(\sigma^2)^2} \left(n - \frac{\sum_{i=1}^{n} x_i^2}{\sigma^2}\right).$$

We search for the maximum of $\ln f(x_1, \cdots, x_n; \sigma^2)$

$$\frac{\partial \ln f(x_1, \cdots, x_n; \sigma^2)}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} x_i^2 \overset{!}{=} 0$$

$$-n\sigma^2 + \frac{1}{2} + \sum_{i=1}^{n} x_i^2 \overset{!}{=} 0 \implies \sigma^2 = \frac{1}{2n} \sum_{i=1}^{n} x_i^2$$

$$\left.\frac{\partial^2 \ln f(x_1, \cdots, x_n; \sigma^2)}{\partial (\sigma^2)^2}\right|_{\sigma^2 = \frac{1}{2n}\sum_{i=1}^{n} x_i^2} = \frac{1}{(\sigma^2)^2} (n - 2n) < 0 \implies \text{maximum}$$

$$\text{and therefore} \qquad \hat{\sigma}_{ML}^2 = \frac{1}{2n} \sum_{i=1}^{n} x_i^2.$$

2. From $\hat{\sigma}_{ML}^2 \sim \mathcal{G}\left(n, \frac{\sigma^2}{n}\right)$ (cp. hint) follows immediately

$$\mathbb{E}\left[\hat{\sigma}_{ML}^2\right] = n\frac{\sigma^2}{n} = \sigma^2$$

$$\text{Var}\left[\hat{\sigma}_{ML}^2\right] = n\left(\frac{\sigma^2}{n}\right)^2 = \frac{\sigma^4}{n}$$

and hence $\hat{\sigma}_{ML}^2$ is an unbiased and convergent estimator of σ^2.
The Fisher information is

$$-\mathbb{E}\left[\frac{\partial^2 \ln f(X_1, \cdots, X_n; \sigma^2)}{\partial (\sigma^2)^2}\right] = -\frac{1}{\sigma^4}\left(n - \frac{\sum_{i=1}^{n} \mathbb{E}[x_i^2]}{\sigma^2}\right) = -\frac{1}{\sigma^4}\left(n - \frac{n}{\sigma^2}\left(\text{Var}[x_i] + (\mathbb{E}[x_i])^2\right)\right) =$$

$$= -\frac{1}{\sigma^4}\left(n - \frac{n}{\sigma^2}\left(\frac{4-\pi}{2}\sigma^2 + \frac{\pi}{2}\sigma^2\right)\right) = -\frac{n}{\sigma^4} + \frac{2n}{\sigma^4} = \frac{n}{\sigma^4},$$

and hence the Cramér-Rao bound reads

$$\text{Var}[\hat{\sigma}_n^2] \geq CRB(\sigma^2) = \frac{\sigma^4}{n}.$$

Since $\text{Var}[\hat{\sigma}_{ML}^2] = CRB(\sigma^2)$ and $\hat{\sigma}_{ML}^2$ is unbiased, $\hat{\sigma}_{ML}^2$ is the efficient estimator of σ^2.

3.(a) The second moment of $X \sim \mathcal{R}(\sigma^2)$ is given by

$$\mathbb{E}[X^2] = \text{Var}[X] + (\mathbb{E}[X])^2 = \frac{4-\pi}{2}\sigma^2 + \frac{\pi}{2}\sigma^2 = 2\sigma^2$$

and therefore

$$\sigma^2 = \frac{1}{2}\mathbb{E}[X^2].$$

Replacing $\mathbb{E}[X^2]$ with the second order sample moment $\hat{m}_2 = \frac{1}{n}\sum_{i=1}^n x_i^2$ yields the estimator

$$\hat{\sigma}^2_{m2} = \frac{1}{2n}\sum_{i=1}^n x_i^2.$$

Since $\hat{\sigma}^2_{m2} \equiv \hat{\sigma}^2_{ML}$, it is the efficient estimator of σ^2.

3.(b) From the first moment of $X \sim \mathcal{R}(\sigma^2)$ we obtain

$$\mathbb{E}[X] = \sigma^2\sqrt{\frac{\pi}{2}} \quad \Longrightarrow \quad \sigma^2 = \frac{2}{\pi}\left(\mathbb{E}[X]\right)^2.$$

Replacing $\mathbb{E}[X]$ with the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ yields the estimator

$$\hat{\sigma}^2_{m1} = \frac{2}{\pi}\left(\overline{x}\right)^2 = \frac{2}{n^2\pi}\left(\sum_{i=1}^n x_i\right)^2 = \frac{2}{n^2\pi}\sum_{i=1}^n\sum_{j=1}^n x_i x_j.$$

The expected value of $\hat{\sigma}^2_{m1}$ is

$$\mathbb{E}[\hat{\sigma}^2_{m1}] = \frac{2}{n^2\pi}\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}[X_i X_j]$$

where

$$\mathbb{E}[X_i X_j] = \begin{cases} \mathbb{E}[X_i^2] = \text{Var}[X_i] + (\mathbb{E}[X_i])^2 = 2\sigma^2 & \text{if } i = j \\ \mathbb{E}[X_i]\mathbb{E}[X_j] = \sigma^2\frac{\pi}{2} & \text{if } i \neq j \text{ (i.i.d. r.v)} \end{cases}$$

and hence

$$\mathbb{E}[\hat{\sigma}^2_{m1}] = \frac{2}{n^2\pi}\left(2n\sigma^2 + n(n-1)\sigma^2\frac{\pi}{2}\right) = \sigma^2\left(\frac{4}{n\pi} + \frac{n-1}{n}\right) = \sigma^2\left(1 + \frac{4-\pi}{n\pi}\right).$$

The estimator $\hat{\sigma}^2_{m1}$ is therefore biased with bias

$$b_n = \sigma^2\frac{4-\pi}{n\pi}.$$

CONTROL: Alternatively, we obtain this result through calculation of the variance of the sample mean,

$$\mathbb{E}[\hat{\sigma}^2_{m1}] = \mathbb{E}\left[\frac{2}{\pi}\left(\overline{x}\right)^2\right] = \frac{2}{\pi}\mathbb{E}\left[\left(\overline{x}\right)^2\right] = \frac{2}{\pi}\left(Var[\overline{x}] + (\mathbb{E}[\overline{x}])^2\right) = \frac{2}{\pi}\left(Var[\overline{x}] + \sigma^2\frac{\pi}{2}\right) = \sigma^2 + \frac{2}{\pi}Var[\overline{x}].$$

The bias is hence given by

$$b_n = \frac{2}{\pi}Var[\overline{x}] = \frac{2}{\pi}Var\left[\frac{1}{n}\sum_{i=1}^n x_i\right] = \frac{2}{n^2\pi}\sum_{i=1}^n Var[x_i] = \frac{2}{n^2\pi}n\frac{4-\pi}{2}\sigma^2 = \sigma^2\frac{4-\pi}{n\pi}.$$

3.(c) Since

$$\mathbb{E}[\hat{\sigma}^2_{m1}] = \sigma^2\left(1 + \frac{4-\pi}{n\pi}\right)$$

an unbiased estimator $\tilde{\sigma}^2_{m1}$ can be obtained as

$$\tilde{\sigma}^2_{m1} = \frac{\hat{\sigma}^2_{m1}}{1 + \frac{4-\pi}{n\pi}} = \frac{n\pi}{(n-1)\pi + 4}\hat{\sigma}^2_{m1} = \frac{2n}{(n-1)\pi + 4}(\overline{x})^2.$$

The ratio of $\tilde{\sigma}^2_{m1} = \frac{2n}{(n-1)\pi+4}(\overline{x})^2$ and $\hat{\sigma}^2_{m1} = \frac{2}{\pi}(\overline{x})^2$ is

$$\frac{2n}{(n-1)\pi + 4}\frac{\pi}{2} = \frac{n}{(n-1) + 4/\pi} \leq \frac{n}{(n-1) + 1} = 1.$$

Therefore, $\text{Var}[\tilde{\sigma}^2_{m1}] \leq \text{Var}[\hat{\sigma}^2_{m1}]$ and since $\tilde{\sigma}^2_{m1}$ is also unbiased, it also has smaller mean squared error and is overall preferable.

4.(a) The likelihood is (cf. 1.)

$$f(x_1, \cdots, x_n; \sigma^2) = \prod_{i=1}^{n} \frac{x_i}{\sigma^2} \exp\left(-\frac{x_i^2}{2\sigma^2}\right)$$

and the prior is given by

$$p(\sigma^2) = \mathcal{IG}(\alpha, \beta/2) = \frac{\left(\frac{\beta}{2}\right)^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{\beta}{2\sigma^2}\right) \propto \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left(-\frac{\beta}{2\sigma^2}\right)$$

and the posterior therefore reads

$$\begin{aligned} f(\sigma^2|x_1, \cdots, x_n, \alpha, \beta) &\propto \left(\frac{1}{\sigma^2}\right)^{n+\alpha+1} \exp\left(-\frac{\beta}{2\sigma^2}\right) \exp\left(-\sum_{i=1}^{n} \frac{x_i^2}{2\sigma^2}\right) \prod_{i=1}^{n} x_i \\ &\propto \left(\frac{1}{\sigma^2}\right)^{n+\alpha+1} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)\right) \prod_{i=1}^{n} x_i \\ &\propto \left(\frac{1}{\sigma^2}\right)^{n+\alpha+1} \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)\right) \\ &\propto \mathcal{IG}\left(n+\alpha,\ \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)\right). \end{aligned}$$

The log-posterior for σ^2 is

$$\ln f(\sigma^2|x_1, \cdots, x_n, \alpha, \beta) \ \propto \ -(n+\alpha+1)\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right).$$

4.(b) We are searching for the maximum of the log-posterior. Taking partial derivatives w.r.t. σ^2

$$\begin{aligned} \frac{\partial \ln f(\sigma^2|x_1, \cdots, x_n, \alpha, \beta)}{\partial \sigma^2} &= -\frac{n+\alpha+1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right) \\ \frac{\partial^2 \ln f(\sigma^2|x_1, \cdots, x_n, \alpha, \beta)}{\partial \left(\sigma^2\right)^2} &= \frac{n+\alpha+1}{(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\left(\sum_{i=1}^{n} x_i^2 + \beta\right) = \frac{1}{(\sigma^2)^2}\left((n+\alpha+1) - \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)\right) \end{aligned}$$

we find

$$\begin{aligned} \frac{\partial \ln f(\sigma^2|x_1, \cdots, x_n, \alpha, \beta)}{\partial \sigma^2} = -\frac{n+\alpha+1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right) &\stackrel{!}{=} 0 \\ (n+\alpha+1)\sigma^2 &\stackrel{!}{=} \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right) \\ \sigma^2 &\stackrel{!}{=} \frac{1}{2(n+\alpha+1)}\left(\sum_{i=1}^{n} x_i^2 + \beta\right) \end{aligned}$$

and upon substitution of the expression for σ^2 into $\frac{\partial^2 \ln f(\sigma^2|x_1, \cdots, x_n, \alpha, \beta)}{\partial(\sigma^2)^2}$

$$\left.\frac{\partial^2 \ln f(\sigma^2|x_1, \cdots, x_n, \alpha, \beta)}{\partial(\sigma^2)^2}\right|_{\sigma^2 = \frac{1}{2(n+\alpha+1)}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)} \propto \frac{1}{(\sigma^2)^2}\left((n+\alpha+1) - 2(n+\alpha+1)\right) < 0$$

and therefore

$$\hat{\sigma}^2_{MAP} = \frac{\sum_{i=1}^{n} x_i^2 + \beta}{2(n+\alpha+1)}.$$

CONTROL: The mode of $\mathcal{IG}(a, b)$ is given by $\frac{b}{a+1}$. Here, $a = n + \alpha$ and $b = \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)$ and therefore the mode is given by $\frac{1}{2(n+\alpha+1)}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)$.

4.(c) The MMSE estimator for a parameter θ is given by

$$\hat{\theta}_{MMSE} = \mathbb{E}[\theta|x_1, \cdots, x_n; \alpha, \beta] = \int \theta f(\theta|x_1, \cdots, x_n; \alpha, \beta) d\theta$$

and hence

$$\hat{\sigma}^2_{MMSE} = \mathbb{E}[\sigma^2|x1, \cdots, x_n, \alpha, \beta] = \int \sigma^2 \mathcal{IG}\left(\underbrace{n+\alpha}_{\alpha'}, \underbrace{\frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)}_{\beta'}\right) d\sigma^2 = \int \sigma^2 \mathcal{IG}(\alpha', \beta') d\sigma^2$$

$$= \int \sigma^2 \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \left(\frac{1}{\sigma^2}\right)^{\alpha'+1} \exp\left(-\frac{\beta'}{\sigma^2}\right) d\sigma^2 = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \int \left(\frac{1}{\sigma^2}\right)^{\alpha'} \exp\left(-\frac{\beta'}{\sigma^2}\right) d\sigma^2$$

$$= \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \int \frac{\Gamma(\alpha'-1)}{(\beta')^{\alpha'-1}} \mathcal{IG}(\alpha'-1, \beta') d\sigma^2 = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')} \frac{\Gamma(\alpha'-1)}{(\beta')^{\alpha'-1}} \underbrace{\int \mathcal{IG}(\alpha'-1, \beta') d\sigma^2}_{=1}$$

$$= \beta' \frac{\Gamma(\alpha'-1)}{\Gamma(\alpha')} = \frac{\beta'}{\alpha'-1} = \frac{\sum_{i=1}^{n} x_i^2 + \beta}{2(n+\alpha-1)}$$

CONTROL: The mean of $\mathcal{IG}(a, b)$ is given by $\frac{b}{a-1}$. Here, $a = n + \alpha$ and $b = \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)$ and therefore the mode is $\frac{1}{2(n+\alpha-1)}\left(\sum_{i=1}^{n} x_i^2 + \beta\right)$.

4.(d) Since $2(n + \alpha - 1) < 2(n + \alpha + 1)$, we know that Var$[\hat{\sigma}^2_{MMSE}] >$ Var$[\hat{\sigma}^2_{MAP}]$ and since the MMSE estimator minimizes the mean squared error, we have that Bias$[\hat{\sigma}^2_{MMSE}] <$ Bias$[\hat{\sigma}^2_{MAP}]$.

5. As the sample size increases, $n \to \infty$, the ML, MMSE and MAP estimators become equivalent.

Exercise 6

1.

$$f(x_1,\cdots,x_n;\lambda)=\prod_{i=1}^{n}\frac{\lambda_i^x}{x_i!}\exp(-\lambda)=\exp(-n\lambda)\frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\propto\exp(-n\lambda)\lambda^{\sum_{i=1}^n x_i}$$

$$\ln f(x_1,\cdots,x_n;\lambda)\propto -n\lambda+\ln\lambda\sum_{i=1}^n x_i$$

$$\frac{\partial\ln f(x_1,\cdots,x_n;\lambda)}{\partial\lambda}=-n+\frac{\sum_{i=1}^n x_i}{\lambda}\overset{!}{=}0\quad\Longrightarrow\quad\lambda=\frac{1}{n}\sum_{i=1}^n x_i$$

$$\frac{\partial^2\ln f(x_1,\cdots,x_n;\lambda)}{\partial\lambda^2}=-\frac{\sum_{i=1}^n x_i}{\lambda^2}<0$$

$$\Longrightarrow\hat\lambda_{ML}=\frac{1}{n}\sum_{i=1}^n x_i=\overline{x}$$

$\mathbb{E}[\hat\lambda_{ML}]=\lambda$ and Var$[\hat\lambda_{ML}]=\frac{\lambda}{n}\longrightarrow$ unbiased and convergent. Since

$$CRB(\lambda)=\left(\mathbb{E}\left[-\frac{\partial^2\ln f(x_1,\cdots,x_n;\lambda)}{\partial\lambda^2}\right]\right)^{-1}=\left(\mathbb{E}\left[\frac{\sum_{i=1}^n x_i}{\lambda^2}\right]\right)^{-1}=\frac{\lambda^2}{n\lambda}=\frac{\lambda}{n}$$

$\hat\lambda_{ML}$ is also the efficient estimator for λ.

2. The prior law has density

$$p(\lambda)=\frac{\lambda^{\alpha-1}\exp\left(-\frac{\lambda}{\beta}\right)}{\Gamma(\alpha)\beta^\alpha}$$

with mean $\mu=\alpha\beta$ and variance $\nu^2=\alpha\beta^2$. There, the posterior law is

$$\begin{aligned}f(\lambda|x_1,\cdots,x_n;\alpha,\beta)&\propto\exp(-n\lambda)\lambda^{\sum_{i=1}^n x_i}\lambda^{\alpha-1}\exp\left(-\frac{\lambda}{\beta}\right)\\&\propto\exp\left(-\lambda(n+\frac{1}{\beta})\right)\lambda^{\alpha-1+\sum_{i=1}^n x_i}\\&\propto\mathcal{G}\left(\alpha+\sum_{i=1}^n x_i,\ \frac{1}{n+\frac{1}{\beta}}\right).\end{aligned}$$

The mean of $z\sim\mathcal{G}(\alpha,\beta)$ is $\mu=\alpha\beta$, giving the MMSE estimator for λ

$$\hat\lambda_{MMSE}=\frac{\alpha+\sum_{i=1}^n x_i}{n+\frac{1}{\beta}}$$

CONTROL:

$$\hat\lambda_{MMSE}=\mathbb{E}\left[\mathcal{G}\left(\alpha'=\alpha+\sum_{i=1}^n x_i,\ \beta'=\frac{1}{n+\frac{1}{\beta}}\right)\right]=\int\lambda\mathcal{G}\left(\alpha',\ \beta'\right)d\lambda=\int\frac{\lambda^{\alpha'}}{\Gamma(\alpha')(\beta')^{\alpha'}}\exp(-\lambda/\beta')d\lambda$$

$$=\int\frac{\alpha'\beta'\lambda^{\alpha'}}{\Gamma(\alpha'+1)(\beta')^{\alpha'+1}}\exp(-\lambda/\beta')d\lambda=\alpha'\beta'\underbrace{\int\mathcal{G}\left(\alpha'+1,\ \beta'\right)d\lambda}_{=1}=\alpha'\beta'.$$

We find the MAP estimator

$$\ln f(\lambda|x_1,\cdots,x_n;\alpha,\beta)\propto-\lambda\left(n+\frac{1}{\beta}\right)+\ln\lambda\left(\alpha-1+\sum_{i=1}^n x_i\right)$$

$$\frac{\partial f(\lambda|x_1,\cdots,x_n;\alpha,\beta)}{\partial\lambda}=-\left(n+\frac{1}{\beta}\right)+\frac{1}{\lambda}\left(\alpha-1+\sum_{i=1}^n x_i\right)\overset{!}{=}0$$

$$\Longrightarrow\lambda=\frac{\alpha-1+\sum_{i=1}^n x_i}{n+\frac{1}{\beta}}$$

$$\left.\frac{\partial^2 f(\lambda|x_1,\cdots,x_n;\alpha,\beta)}{\partial\lambda^2}\right|_{\lambda=\frac{1+\alpha+\sum_{i=1}^n x_i}{n+\frac{1}{\beta}}}=-\frac{n+\frac{1}{\beta}}{\alpha-1+\sum_{i=1}^n x_i}$$

which is negative[1] as long as at least one of the observations are informative (i.e. $\exists i: \; x_i \neq 0$) or if $\alpha > 1$ and hence

$$\hat{\lambda}_{MAP} = \frac{\alpha - 1 + \sum_{i=1}^{n} x_i}{n + \frac{1}{\beta}}$$

CONTROL: the mode of $\mathcal{G}(\alpha, \beta)$ is given by $m = (\alpha - 1)\beta$ as long as $\alpha > 1$. In our problem, this gives $\lambda_{mode} = \frac{\alpha - 1 + \sum_{i=1}^{n} x_i}{n + \frac{1}{\beta}}$ as long as we have one non-zero observation x_i or $\alpha > 1$.

3.

$$\mathbb{E}\left[\hat{\lambda}_{MMSE}|\alpha, \beta\right] = \frac{\alpha + \sum_{i=1}^{n} \mathbb{E}[x_i]}{n + \frac{1}{\beta}} = \frac{\alpha + n\lambda}{n + \frac{1}{\beta}} = \lambda + \underbrace{\frac{\alpha - \frac{\lambda}{\beta}}{n + \frac{1}{\beta}}}_{b(\lambda)}$$

$$\text{Var}\left[\hat{\lambda}_{MMSE}|\alpha, \beta\right] = \frac{\sum_{i=1}^{n} \text{Var}[x_i]}{\left(n + \frac{1}{\beta}\right)^2} = \frac{n\lambda}{\left(n + \frac{1}{\beta}\right)^2}$$

$$\mathbb{E}\left[\hat{\lambda}_{MAP}|\alpha, \beta\right] = \frac{\alpha - 1 + \sum_{i=1}^{n} \mathbb{E}[x_i]}{n + \frac{1}{\beta}} = \frac{\alpha - 1 + n\lambda}{n + \frac{1}{\beta}} = \lambda + \underbrace{\frac{\alpha - 1 - \frac{\lambda}{\beta}}{n + \frac{1}{\beta}}}_{b(\lambda)}$$

$$\text{Var}\left[\hat{\lambda}_{MAP}|\alpha, \beta\right] = \text{Var}\left[\hat{\lambda}_{MMSE}|\alpha, \beta\right]$$

$\hat{\lambda}_{ML}$ is unbiased, and $\hat{\lambda}_{MMSE}$ and $\hat{\lambda}_{MAP}$ are asymptotically unbiased (cf. 3.). The variance of $\hat{\lambda}_{MMSE}$ and $\hat{\lambda}_{MAP}$ tends to $\text{Var}[\hat{\lambda}_{ML}] = \frac{\lambda}{n}$ as $n \longrightarrow \infty$. The three estimators are hence equivalent for n "large".

4.

$$MSE\left[\hat{\lambda}_{MMSE}|\alpha, \beta\right] = \text{Var}\left[\hat{\lambda}_{MMSE}|\alpha, \beta\right] + \left(\mathbb{E}\left[\hat{\lambda}_{MMSE}|\alpha, \beta\right] - \lambda\right)^2 = \frac{n\lambda + \left(\alpha - \frac{\lambda}{\beta}\right)^2}{\left(n + \frac{1}{\beta}\right)^2}$$

$$MSE\left[\hat{\lambda}_{MAP}|\alpha, \beta\right] = \frac{n\lambda + \left(\alpha - 1 - \frac{\lambda}{\beta}\right)^2}{\left(n + \frac{1}{\beta}\right)^2}$$

From 3., we know that $\hat{\lambda}_{MMSE}$ and $\hat{\lambda}_{MAP}$ have the same variance. Therefore, $\hat{\lambda}_{MMSE}$ has smaller bias.

[1]Note that $\hat{\lambda}_{ML} \equiv 0$ when $\forall i: \; x_i \equiv 0$ although $\lambda > 0$.

Exercise 7

1.

$$f(z_1, z_2; \xi) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(z_1 - \xi)^2 + (z_2 - \xi)^2}{2\sigma^2}\right)$$

Hence,

$$\begin{aligned}\hat{\xi}_{ML} &= 0 \quad \text{if} \quad f(z_1, z_2; 0) \geq f(z_1, z_2; 1) \\ &= 0 \quad \text{if} \quad z_1^2 + z_2^2 \leq (z_1 - 1)^2 + (z_2 - 1)^2\end{aligned}$$

i.e., if (z_1, z_2) is closer to $(0, 0)$ than to $(1, 1)$.

2.

$$f(\xi | z_1, z_2) \propto (1 - p)^\xi p^{1-\xi} \exp\left(-\frac{(z_1 - \xi)^2 + (z_2 - \xi)^2}{2\sigma^2}\right)$$

hence

$$\hat{\xi}_{MAP} = 0 \quad \text{if} \quad f(0|z_1, z_2) \geq f(1|z_1, z_2)$$

i.e.,

$$\begin{aligned} p \exp\left(-\frac{z_1^2 + z_2^2}{2\sigma^2}\right) &\geq (1 - p) \exp\left(-\frac{(z_1 - 1)^2 + (z_2 - 1)^2}{2\sigma^2}\right) \\ \ln p - \frac{z_1^2 + z_2^2}{2\sigma^2} &\geq \ln(1 - p) - \frac{(z_1 - 1)^2 + (z_2 - 1)^2}{2\sigma^2} \\ z_1^2 + z_2^2 &\leq (z_1 - 1)^2 + (z_2 - 1)^2 - 2\sigma^2 \ln\left(\frac{1 - p}{p}\right)\end{aligned}$$

and hence

$$\hat{\xi}_{MAP} = 0 \quad \text{if} \quad z_1^2 + z_2^2 \leq (z_1 - 1)^2 + (z_2 - 1)^2 - 2\sigma^2 \ln\left(\frac{1 - p}{p}\right).$$

If $p = \frac{1}{2}$, this gives the ML estimator. If $p < \frac{1}{2}$ ($p > \frac{1}{2}$), we have higher (smaller) chances to have a bit 1 than a bit 0, and the region (z_1, z_2) in the plane for which $\hat{\xi}_{MAP} = 0$ is smaller (larger).

3. The MMSE estimator does not take values $\{0, 1\}$ and is hence not adequate for this problem.