SOLUTIONS - EXERCISE 1

(a) The joint density of $X_1, \ldots, X_5$ is:

$$p(x_1, \ldots, x_5|\theta) = \prod_{i=1}^{5} p(x_i|\theta) = \prod_{i=1}^{5} \theta^{x_i} \cdot (1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{5} x_i} \cdot (1-\theta)^{5-\sum_{i=1}^{5} x_i} = \theta^4 \cdot (1-\theta)^1$$

For the density of the posterior we have:

$$p(\theta|x_1, \ldots, x_5) \propto p(x_1, \ldots, x_5|\theta) \cdot p(\theta)$$

$$= \theta^4 \cdot (1-\theta)^1 \cdot \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}$$

$$\propto \theta^4 \cdot (1-\theta)^1 \cdot \theta^{a-1} \cdot (1-\theta)^{b-1}$$

$$= \theta^{4+a-1} \cdot (1-\theta)^{1+b-1}$$

As a function of θ, this expression is proportional to the density of a Beta distribution with parameters: $\tilde{a} = 4 + a$ and $\tilde{b} = 1 + b$. Thus, the posterior distribution is a Beta distribution with $\tilde{a} = 4 + a = 6$ and $\tilde{b} = 1 + b = 2$:

$$\theta|(X_1 = x_1, \ldots, X_5 = x_5) \sim \text{Beta}(6, 2)$$

(b) For the predictive probability we have:

$$P(\tilde{X} = 1|X_1 = x_1, \ldots, X_5 = x_5) = \int p(\tilde{x}|\theta) \cdot p(\theta|x_1, \ldots, x_5) d\theta$$

where for $\tilde{x} = 1$:

$$p(\tilde{x}|\theta) = \theta^{\tilde{x}} \cdot (1-\theta)^{1-\tilde{x}} = \theta$$

and

$$p(\theta|x_1, \ldots, x_5) = \frac{\Gamma(8)}{\Gamma(6)\Gamma(2)} \cdot \theta^5 \cdot (1-\theta)^1$$

is the density of the Beta(6,2) posterior distribution of θ, computed in part (a).

Hence, we obtain:

$$P(\tilde{X} = 1|X_1 = x_1, \ldots, X_5 = x_5) = \int p(\tilde{x}|\theta) \cdot p(\theta|x_1, \ldots, x_5) d\theta$$

$$= \int \theta \cdot \frac{\Gamma(8)}{\Gamma(6)\Gamma(2)} \cdot \theta^5 \cdot (1-\theta)^1 d\theta$$

$$= \frac{\Gamma(8)}{\Gamma(6)\Gamma(2)} \cdot \int \theta^6 \cdot (1-\theta)^1 d\theta$$

$$= \frac{\Gamma(8)}{\Gamma(6)\Gamma(2)} \cdot \left(\frac{\Gamma(9)}{\Gamma(7)\Gamma(2)}\right)^{-1} \cdot \int \frac{\Gamma(9)}{\Gamma(7)\Gamma(2)} \cdot \theta^6 \cdot (1-\theta)^1 d\theta$$

$$= \frac{\Gamma(8)}{\Gamma(6)\Gamma(2)} \cdot \frac{\Gamma(7)\Gamma(2)}{\Gamma(9)} \cdot 1$$

$$= \frac{7!}{5! \cdot 1!} \cdot \frac{6! \cdot 1!}{8!} = \frac{6}{8} = 0.75$$

(c) Here we have for the posterior density:

$$p(\theta|x_1,\ldots,x_5) = \frac{p(x_1,\ldots,x_5|\theta)\cdot p(\theta)}{\sum_{\theta\in\{0,0.5,1\}} p(x_1,\ldots,x_5|\theta)\cdot p(\theta)}$$

as θ can only take on three different values ($\theta \in \{0, 0.5, 1\}$).

As $p(x_1,\ldots,x_5|\theta = 0) = 0$ and $p(x_1,\ldots,x_5|\theta = 1) = 0$, it follows

$$p(\theta = 0|x_1,\ldots,x_5) = 0$$
$$p(\theta = 1|x_1,\ldots,x_5) = 0$$

So the only possible parameter is $\theta = 0.5$, and henceforth it follows:

$$p(\theta = 0.5|x_1,\ldots,x_5) = 1$$

SOLUTIONS EXERCISE 2:
For the posterior density we have:

$$\begin{aligned}
p(\mu|x) &\propto p(x|\mu)\cdot p(\mu)\\
&\propto \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{1}\cdot\exp\{-\frac{1}{2}\cdot\frac{(x-\mu)^2}{1}\}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{1}\cdot\exp\{-\frac{1}{2}\cdot\frac{(\mu-0)^2}{1}\}\\
&\propto \exp\{-\frac{1}{2}\cdot(x-\mu)^2\}\cdot\exp\{-\frac{1}{2}\cdot\mu^2\}\\
&\propto \exp\{-\frac{1}{2}\cdot(x^2-2x\mu+\mu^2)-\frac{1}{2}\cdot\mu^2\}\\
&\propto \exp\{-\frac{1}{2}x^2+x\mu-\frac{1}{2}\cdot\mu^2-\frac{1}{2}\cdot\mu^2\}\\
&\propto \exp\{-\mu^2+x\mu\}\\
&\propto \exp\{-\frac{1}{2}\cdot 2\cdot\mu^2+x\cdot\mu\}
\end{aligned}$$

It follows from HINT (2):

$$\mu|(X = x) \sim N(\frac{x}{2},\frac{1}{2})$$

(b) For the marginal likelihood we have:

$$\begin{aligned}
p(x) &= \int p(x|\mu)\cdot p(\mu)\,d\mu \\
&= \int \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{1}\cdot\exp\{-\frac{1}{2}\cdot\frac{(x-\mu)^2}{1}\}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{1}\cdot\exp\{-\frac{1}{2}\cdot\frac{(\mu-0)^2}{1}\}\,d\mu \\
&= \int \frac{1}{\sqrt{2\pi}}\cdot\exp\{-\frac{1}{2}\cdot(x-\mu)^2\}\cdot\frac{1}{\sqrt{2\pi}}\cdot\exp\{-\frac{1}{2}\cdot\mu^2\}\,d\mu \\
&= \frac{1}{\sqrt{2\pi}}\int \frac{1}{\sqrt{2\pi}}\cdot\exp\{-\frac{1}{2}x^2+x\mu-\frac{1}{2}\mu^2-\frac{1}{2}\cdot\mu^2\}\,d\mu \\
&= \frac{1}{\sqrt{2\pi}}\int \frac{1}{\sqrt{2\pi}}\cdot\exp\{-\frac{1}{2}x^2+x\mu-\mu^2\}\,d\mu \\
&= \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}x^2\}\int \frac{1}{\sqrt{2\pi}}\cdot\exp\{x\mu-\mu^2\}\,d\mu
\end{aligned}$$

The function in the integral is proportional to the density of a Gaussian $N(\frac{x}{2},\frac{1}{2})$ distribution. We used that already above when computing the posterior.
The density of a $N(\frac{x}{2},\frac{1}{2})$ distribution is:

$$\begin{aligned}
p(\mu) &= \frac{1}{\sqrt{2\pi}}\cdot\sqrt{2}\cdot\exp\{-\frac{1}{2}\cdot 2(\mu-x/2)^2\} \\
&= \frac{1}{\sqrt{2\pi}}\cdot\sqrt{2}\cdot\exp\{-\mu^2+x\mu-\frac{x^2}{4}\}
\end{aligned}$$

We continue:

$$\begin{aligned}
p(x) &= \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}x^2\}\int \frac{1}{\sqrt{2\pi}}\cdot\exp\{x\mu-\mu^2\}\,d\mu \\
&= \frac{1}{\sqrt{2\pi}}\cdot\exp\{-\frac{1}{4}x^2\}\cdot\frac{1}{\sqrt{2}}\cdot\int \exp\{-\frac{1}{4}x^2\}\cdot\sqrt{2}\cdot\frac{1}{\sqrt{2\pi}}\cdot\exp\{x\mu-\mu^2\}\,d\mu \\
&= \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{2}}\exp\{-\frac{1}{4}x^2\}\int \frac{1}{\sqrt{2\pi}}\cdot\sqrt{2}\cdot\exp\{-\mu^2+x\mu-\frac{x^2}{4}\}\,d\mu \\
&= \frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{2}}\cdot\exp\{-\frac{1}{4}x^2\}
\end{aligned}$$

The latter expression the density of a $N(0,2)$ distribution.

SOLUTIONS EXERCISE 3:

(a) As a function of $\boldsymbol{\beta}$ we have for the likelihood:

$$\begin{aligned}
p(\mathbf{y}|\boldsymbol{\beta}) &\propto \exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Sigma}_1^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\} \\
&\propto \exp\{-\frac{1}{2}(\mathbf{y}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}-\mathbf{y}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}+\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\boldsymbol{\beta})\} \\
&\propto \exp\{-\frac{1}{2}(-2\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}+\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\boldsymbol{\beta})\} \\
&\propto \exp\{\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}-\frac{1}{2}\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\boldsymbol{\beta}\}
\end{aligned}$$

And for the prior of $\boldsymbol{\beta}$ we have:

$$\begin{aligned}
p(\boldsymbol{\beta}) &\propto \exp\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\} \\
&\propto \exp\{-\frac{1}{2}(\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\beta}-\boldsymbol{\mu}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\beta}-\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}\} \\
&\propto \exp\{-\frac{1}{2}(\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\beta}-2\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}\} \\
&\propto \exp\{-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\beta}+\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}\}
\end{aligned}$$

For the posterior of $\boldsymbol{\beta}$ we get:

$$\begin{aligned}
p(\boldsymbol{\beta}|\mathbf{y}) &\propto p(\mathbf{y}|\boldsymbol{\beta})\cdot p(\boldsymbol{\beta}) \\
&\propto \exp\{\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}-\frac{1}{2}\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\boldsymbol{\beta}\}\cdot\exp\{-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\beta}+\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}\} \\
&\propto \exp\{\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}-\frac{1}{2}\boldsymbol{\beta}^T\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}\boldsymbol{\beta}-\frac{1}{2}\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\beta}+\boldsymbol{\beta}^T\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}\} \\
&\propto \exp\{\boldsymbol{\beta}^T[\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}+\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}]-\frac{1}{2}\boldsymbol{\beta}^T[\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}+\boldsymbol{\Sigma}_2^{-1}]\boldsymbol{\beta}\}
\end{aligned}$$

From the shape it follows for the full conditional distribution of $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}|\mathbf{y} \sim \mathcal{N}([\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}+\boldsymbol{\Sigma}_2^{-1}]^{-1}[\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{y}+\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}]\ ,\ [\mathbf{X}^T\boldsymbol{\Sigma}_1^{-1}\mathbf{X}+\boldsymbol{\Sigma}_2^{-1}]^{-1})$$

(b) For the special case we have:

$$\begin{aligned}
\boldsymbol{\beta}|\mathbf{y} &\sim \mathcal{N}([\mathbf{X}^T\mathbf{I}^{-1}\mathbf{X}+(\epsilon\cdot\mathbf{I})^{-1}]^{-1}[\mathbf{X}^T\mathbf{I}^{-1}\mathbf{y}+(\epsilon\cdot\mathbf{I})^{-1}\mathbf{0}]\ ,\ [\mathbf{X}^T\mathbf{I}^{-1}\mathbf{X}+(\epsilon\cdot\mathbf{I})^{-1}]^{-1}) \\
&\sim \mathcal{N}([\mathbf{X}^T\mathbf{X}+\epsilon^{-1}\cdot\mathbf{I})]^{-1}[\mathbf{X}^T\mathbf{y}]\ ,\ [\mathbf{X}^T\mathbf{X}+\epsilon^{-1}\cdot\mathbf{I})]^{-1})
\end{aligned}$$

(c) For very large ϵ values, $\epsilon^{-1}\to 0$, and we have approximately:

$$\boldsymbol{\beta}|\mathbf{y} \sim \mathcal{N}([\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{y}\ ,\ [\mathbf{X}^T\mathbf{X}]^{-1})$$

(cf. Least Squares Estimator).