**Exam Project Statistical Reasoning**

Date: Wednesday, April 3, 2024
Time: 18.15-20.15h
Place: Martiniplaza
Progress code: WBMA038-05

**Rules to follow:**

- This is a two hours long closed book exam. Do not forget to write your name and student number onto each paper sheet and onto the envelope.
- There are 5 exercises and the number of points per exercise are indicated within boxes. You can reach 90 points.
- If you have to compute something, then include the relevant equations.
- **We wish you success with the completion of the exam!**

START OF FINAL EXAM

1. **Marginal Likelihood.** **15**

   Consider a random variable $Y$ whose distribution has density

   $$p(y|\theta) = \frac{\theta^y \cdot e^{-\theta}}{y!} \qquad (\theta > 0 \text{ and } y = 0, 1, 2, \ldots)$$

   and assume that $\theta$ has a distribution with density

   $$p(\theta) = \lambda \cdot e^{-\lambda\theta} \qquad (\lambda > 0 \text{ and } \theta > 0)$$

   **EXERCISE:** Show that the marginal distribution of $Y$ has density:

   $$p(y) = \left(\frac{1}{1+\lambda}\right)^y \cdot \frac{\lambda}{1+\lambda} \qquad (y = 0, 1, 2, \ldots)$$

   **HINT:** A Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ has density

   $$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x} \qquad (x > 0)$$

2. **Full conditional distributions.** $\boxed{20}$
Consider a random sample $Y_1, \ldots, Y_n$ from a Gaussian distribution with mean $\mu = 0$ and unknown variance $\sigma^2 > 0$. We impose an inverse Gamma distribution with parameters $a > 0$ and $b > 0$ on $\sigma^2$, where $a$ is known and $b$ is unknown. A Gamma hyperprior with known parameters $\alpha > 0$ and $\beta > 0$ is imposed on $b$. We have:

$$\begin{aligned} Y_1, \ldots, Y_n|\sigma^2 &\sim \mathcal{N}(0, \sigma^2) \\ \sigma^{-2}|b &\sim \text{GAM}(a, b) \\ b &\sim \text{GAM}(\alpha, \beta) \end{aligned}$$

where $a$, $\alpha$ and $\beta$ are fixed and known.

(a) $\boxed{10}$ Compute the full conditional distribution of $\sigma^{-2}$.

(b) $\boxed{10}$ Compute the full conditional distribution of $b$.

<u>**HINT**</u>: The density of the Gamma and multivariate Gaussian distribution are provided in the hints of Exercises 1 and 5.

3. **Predictive Distribution of Multinomial-Dirichlet Model.** $\boxed{25}$

Let the random vector $(N_1, \ldots, N_K)^T$ be multinomial distributed with parameters $(\theta_1, \ldots, \theta_K)$ and $n \in \mathbb{N}$, so that the density is

$$p(n_1, \ldots, n_K|\theta_1, \ldots, \theta_K) = \frac{n!}{n_1! \cdot \ldots \cdot n_K!} \cdot \prod_{k=1}^{K} \theta_k^{n_k}$$

where $\theta_k > 0$ for all $k$, $\theta_1 + \ldots + \theta_K = 1$, $n_k \in \{0, \ldots, n\}$ for all $k$ and $n_1 + \ldots + n_K = n$. Assume that $(\theta_1, \ldots, \theta_K)^T$ is Dirichlet distributed with density:

$$p(\theta_1, ..., \theta_K) = \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \cdot \prod_{k=1}^{K} \theta_k^{\alpha_k - 1}$$

where $\alpha_1, \ldots, \alpha_K > 0$ are fixed hyperparameters.

<u>**EXERCISES:**</u>

(a) $\boxed{10}$ Compute the posterior distribution of $(\theta_1, \ldots, \theta_K)$.

(b) $\boxed{15}$ Show that for $n = 1$ the predictive probability for a new realisation

$$(\tilde{n}_1, \ldots, \tilde{n}_{j-1}, \tilde{n}_j, \tilde{n}_{j+1}, \ldots, \tilde{n}_K) = (0, \ldots, 0, 1, 0 \ldots, 0)$$

is:

$$p(\tilde{n}_1, \ldots, \tilde{n}_K|n_1, \ldots, n_K) = \frac{\alpha_j + n_j}{\alpha + n}$$

where $\alpha := \alpha_1 + \ldots + \alpha_K$.

4. **Uniform distribution with Pareto prior** $\boxed{15}$
Consider the following sampling model. The random variables $Y_1, \ldots, Y_n$ are i.i.d. and continuous uniformly distributed on the interval $[0, b]$, symbolically:

$$Y_1, \ldots, Y_n | b \sim \text{UNI}[0, b]$$

where the upper bound $b > 0$ is unknown. Impose a Pareto prior with parameters $k > 0$ and $m > 0$ on $b$:

$$b \sim \text{Pareto}(k, m)$$

   (a) $\boxed{10}$ Compute the posterior distribution of $b$.

   (b) $\boxed{5}$ Give an interpretation of $k$ and $m$ in terms of pseudo observations.

**HINT**: The densities of the uniform distribution on $[0, b]$ and the Pareto distribution with parameters $k > 0$ and $m > 0$ are

$$p(x|b) = \begin{cases} \frac{1}{b}, & x \in [0, b] \\ 0, & \text{else} \end{cases}$$

$$p(x|k, m) = \begin{cases} \left(\frac{m}{x}\right)^k, & m \leq x \\ 0, & \text{else} \end{cases}$$

5. **Bayesian Linear Regression** $\boxed{15}$
Consider a regression problem where $Y$ is the response variable that depends on covariates $X_1, \ldots, X_k$. Let $\mathbf{y} \in \mathbb{R}^n$ be the response vector and let $\mathbf{X} \in \mathbb{R}^{n,k+1}$ denote the design matrix, where the first column is a column of 1's for the intercept. We have the likelihood:

$$\mathbf{Y} | \boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_1)$$

where $\boldsymbol{\beta} \in \mathbb{R}^{k+1}$ is the regression coefficient vector, $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{n,n}$ is a known positive definite covariance matrix, and $\mathbf{Y} = \mathbf{y}$ has been observed. On $\boldsymbol{\beta}$ we impose a Gaussian prior with:

$$\boldsymbol{\beta} \sim \mathcal{N}_{k+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$$

where $\boldsymbol{\mu} \in \mathbb{R}^{k+1}$, and $\boldsymbol{\Sigma}_2 \in \mathbb{R}^{k+1,k+1}$ is a known positive definite covariance matrix.

**EXERCISE:**
Compute the power posterior distribution of $\boldsymbol{\beta}$ given any temperature $\tau \in [0, 1]$.

**HINT**:
A Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ has density:

$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2} \cdot \det(\boldsymbol{\Sigma})^{-1/2} \cdot \exp\{-\frac{1}{2} \cdot (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\} \qquad (\mathbf{x} \in \mathbb{R}^n)$$

Recall that a random vector whose density fulfills: $p(\mathbf{x}) \propto \exp\{-\frac{1}{2} \cdot \mathbf{x}^T \mathbf{B}\, \mathbf{x} + \mathbf{x}^T \mathbf{a}\}$ must have a $\mathcal{N}(\mathbf{B}^{-1}\mathbf{a}, \mathbf{B}^{-1})$ distribution.

**END OF EXAM**

**Sample Solutions for Exercise 1**

For the density of the marginal distribution we have:

$$p(y) = \int p(y,\theta)d\theta = \int p(y|\theta)p(\theta)d\theta$$

$$= \int \frac{\theta^y \cdot e^{-\theta}}{y!} \cdot \lambda \cdot e^{-\lambda\theta} d\theta = \frac{\lambda}{y!} \cdot \int \theta^y \cdot e^{-(1+\lambda)\theta} d\theta$$

$$= \frac{\lambda}{y!} \cdot \frac{\Gamma(y+1)}{(1+\lambda)^{y+1}} \cdot \int \frac{(1+\lambda)^{y+1}}{\Gamma(y+1)} \cdot \theta^y \cdot e^{-(1+\lambda)\theta} d\theta$$

$$= \frac{\lambda}{y!} \cdot \frac{\Gamma(y+1)}{(1+\lambda)^{y+1}} \cdot 1 = \frac{\lambda}{(1+\lambda)^{y+1}} = \left(\frac{1}{1+\lambda}\right)^y \cdot \frac{\lambda}{1+\lambda}$$

Just a remark: This is the density of a geometric distribution with parameter $p := \frac{\lambda}{1+\lambda}$.

**Sample Solutions for Exercise 2**

**(a)** Full conditional of $\sigma^{-2}$:

$$\begin{aligned}
p(\sigma^{-2}|y_1,\ldots,y_n,b) &\propto p(y_1,\ldots,y_n|\sigma^{-2}) \cdot p(\sigma^{-2}|b) \\
&\propto \left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} \cdot \exp\{-\frac{1}{2}\frac{y_i^2}{\sigma^2}\}\right) \cdot \frac{b^a}{\Gamma(a)} \cdot (\sigma^{-2})^{a-1} \cdot \exp\{-\sigma^{-2} \cdot b\} \\
&\propto \frac{1}{\sigma^n} \cdot \exp\{-\frac{1}{2}\sum_{i=1}^{n} \frac{y_i^2}{\sigma^2}\} \cdot (\sigma^{-2})^{a-1} \cdot \exp\{-\sigma^{-2} \cdot b\} \\
&\propto (\sigma^{-2})^{\frac{n}{2}} \cdot \exp\{-\frac{1}{2}\sigma^{-2}\sum_{i=1}^{n} y_i^2\} \cdot (\sigma^{-2})^{a-1} \cdot \exp\{-\sigma^{-2} \cdot b\} \\
&\propto (\sigma^{-2})^{\frac{n}{2}+a-1} \exp\{-\sigma^{-2} \cdot (\frac{1}{2} \cdot \sum_{i=1}^{n} y_i^2 + b)\}
\end{aligned}$$

From the shape it follows that the full conditional distribution of $\sigma^{-2}$ must be a Gamma distribution with parameters $\tilde{a} := \frac{n}{2} + a$ and $\tilde{b} := \frac{1}{2} \cdot \sum_{i=1}^{n} y_i^2 + b$.

**(b)** Full conditional of $b$:

$$\begin{aligned}
p(b|y_1,\ldots,y_n,\sigma^{-2}) &\propto p(\sigma^{-2}|b) \cdot p(b) \\
&\propto \frac{b^a}{\Gamma(a)} \cdot (\sigma^{-2})^{a-1} \cdot \exp\{-\sigma^{-2} \cdot b\} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot b^{\alpha-1} \cdot \exp\{-b \cdot \beta\} \\
&\propto b^a \cdot \exp\{-\sigma^{-2} \cdot b\} \cdot b^{\alpha-1} \cdot \exp\{-b \cdot \beta\} \\
&\propto b^{a+\alpha-1} \cdot \exp\{-b(\sigma^{-2} + \beta)\}
\end{aligned}$$

From the shape it follows that of the full conditional distribution of $b$ must be a Gamma distribution with parameters $\tilde{\alpha} := \alpha + a$ and $\tilde{\beta} := \beta + \sigma^{-2}$.

**Sample Solutions for Exercise 3**

**(a)** For the posterior density we have

$$p(\theta_1,\dots,\theta_K|n_1,\dots,n_K) \propto p(n_1,\dots,n_K|\theta_1,\dots,\theta_K)\cdot p(\theta_1,\dots,\theta_K)$$

$$\propto \frac{n!}{n_1!\cdot\ldots\cdot n_K!}\cdot\prod_{k=1}^K \theta_k^{n_k}\cdot\frac{\Gamma(\sum_{k=1}^K\alpha_k)}{\prod_{k=1}^K\Gamma(\alpha_k)}\cdot\prod_{k=1}^K\theta_k^{\alpha_k-1}\propto\prod_{k=1}^K\theta_k^{n_k+\alpha_k-1}$$

The density of the posterior is proportional to the density of a Dirichlet distribution with parameters $\alpha_1+n_1,\dots,\alpha_K+n_K$. We conclude

$$(\theta_1,\dots,\theta_K)|(N_1=n_1,\dots,N_K=n_K)\sim \mathrm{DIR}(\alpha_1+n_1,\dots,\alpha_K+n_K)$$

**(b)** For the new observation the density is:

$$p(\tilde{n}_1,\dots,\tilde{n}_K|\theta_1,\dots,\theta_K)=\frac{1!}{0!\cdot\ldots 0!\cdot 1!\cdot 0!\cdot\ldots\cdot 0!}\cdot\prod_{k=1}^K\theta_k^{\tilde{n}_k}=\prod_{k=1}^K\theta_k^{\tilde{n}_k}$$

The density of the predictive distribution is:

$$p(\tilde{n}_1,\dots,\tilde{n}_K|n_1,\dots,n_K)=\int p(\tilde{n}_1,\dots,\tilde{n}_K|\theta_1,\dots,\theta_K)\cdot p(\theta_1,\dots,\theta_K|n_1,\dots,n_K)\,d\vec{\theta}$$

where $\vec{\theta}=(\theta_1,\dots,\theta_K)^T$

$$=\int\left(\prod_{k=1}^K\theta_k^{\tilde{n}_k}\right)\cdot\frac{\Gamma(\sum\limits_{k=1}^K(\alpha_k+n_k))}{\prod\limits_{k=1}^K\Gamma(\alpha_k+n_k)}\cdot\prod_{k=1}^K\theta_k^{\alpha_k+n_k-1}\;d\vec{\theta}=\frac{\Gamma(\sum\limits_{k=1}^K(\alpha_k+n_k))}{\prod\limits_{k=1}^K\Gamma(\alpha_k+n_k)}\cdot\int\left(\prod_{k=1}^K\theta_k^{\tilde{n}_k}\right)\cdot\prod_{k=1}^K\theta_k^{\alpha_k+n_k-1}\;d\vec{\theta}$$

$$=\frac{\Gamma(\sum\limits_{k=1}^K(\alpha_k+n_k))}{\prod\limits_{k=1}^K\Gamma(\alpha_k+n_k)}\cdot\int\prod_{k=1}^K\theta_k^{\alpha_k+n_k+\tilde{n}_k-1}\;d\vec{\theta}=\frac{\Gamma(\sum\limits_{k=1}^K(\alpha_k+n_k))}{\prod\limits_{k=1}^K\Gamma(\alpha_k+n_k)}\cdot\frac{\prod\limits_{k=1}^K\Gamma(\alpha_k+n_k+\tilde{n}_k)}{\Gamma(\sum\limits_{k=1}^K(\alpha_k+n_k+\tilde{n}_k))}\cdot 1$$

Since $\tilde{n}_k=0$ for all $k\neq j$ and $\tilde{n}_j=1$:

$$=\frac{\Gamma(\sum\limits_{k=1}^K(\alpha_k+n_k))}{\Gamma(\alpha_j+n_j)}\cdot\frac{\Gamma(\alpha_j+n_j+1)}{\Gamma(1+\sum\limits_{k=1}^K(\alpha_k+n_k))}=\frac{\alpha_j+n_j}{\sum\limits_{k=1}^K(\alpha_k+n_k)}$$

With $\alpha:=\sum\limits_{k=1}^K\alpha_k$, and $n=\sum\limits_{k=1}^K n_k$ we have:

$$p(\tilde{n}_1,\dots,\tilde{n}_K|n_1,\dots,n_K)=\frac{\alpha_j+n_j}{\alpha+n}$$

**Sample Solutions for Exercise 4**

**(a)** Compute the joint density

$$p(y_1,\ldots,y_n|b) = \prod_{i=1}^{n} p(y_i|b) = \begin{cases} \left(\frac{1}{b}\right)^n, & y_1 \le b, \ldots, y_n \le b \\ 0, & \text{else} \end{cases}$$
$$= \begin{cases} \left(\frac{1}{b}\right)^n, & \max\{y_1,\ldots,y_n\} \le b \\ 0, & \text{else} \end{cases}$$

We have for the posterior density

$$p(b|y_1,\ldots,y_n) \propto p(y_1,\ldots,y_n|b) \cdot p(b)$$
$$= \begin{cases} \left(\frac{1}{b}\right)^n \cdot \left(\frac{m}{b}\right)^k, & \max\{y_1,\ldots,y_n\} \le b \text{ and } m \le b \\ 0, & \text{else} \end{cases}$$
$$\propto \begin{cases} \left(\frac{1}{b}\right)^{n+k}, & \max\{y_1,\ldots,y_n,m\} \le b \\ 0, & \text{else} \end{cases}$$

The posterior density is proportional to the density of a Pareto distribution with

$$\begin{aligned} \tilde{a} &= n+k \\ \tilde{b} &= \max\{y_1,\ldots,y_n,m\} \end{aligned}$$

**(b)** Interpretation: There are $k$ additional pseudo observations, and the maximum of the $k$ pseudo observations is equal to $m$.

**Sample Solutions for Exercise 5**

As a function of $\boldsymbol{\beta}$ we have:

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

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 power posterior of $\boldsymbol{\beta}$ we have for the density:

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

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

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