Solution to Exam 20230526 Zwanzig

Task 1

a) The posterior probability can be updated by $\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta)$. Then, $f(x \mid \theta)\pi(\theta)$ is calculated at

	x			
θ	0	1	2	3
0	0.16	0.02	0.02	0
1	0.3	0.18	0.12	0
2	0	0.04	0.1	0.06

Hence, the posterior distribution is

	x			
θ	0	1	2	3
0	$8/23$	$1/12$	$1/12$	0
1	$15/23$	$3/4$	$1/2$	0
2	0	$1/6$	$5/12$	1

b) If we observe $x = 0$, the MLE is 0.8 and the MAP is 1. If we observe $x = 1$, the MLE is 0.3 and the MAP is 1. If we observe $x = 2$, the MLE is 2 and the MAP is 1. If we observe $x = 2$, the MLE is 2 and the MAP is 2.

c) (Don't know exactly what Silvelyn means) MAP is penalized MLE, where penalization is done from the prior. The prior strongly favors $\theta = 1$. Thus, with one observation, the MAP is often $\theta = 1$, except $x = 3$ where we observe unlimited strike.

Task 2

a) The posterior probability can be updated by $\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta)$. Then, $f(x \mid \theta)\pi(\theta)$ is calculated at

	x		
θ	0	1	2
0	$0.8(1-p)$	$0.2(1-p)$	0
1	$0.2p$	$0.7p$	$0.1p$

Hence, the posterior distribution is

	x		
θ	0	1	2
0	$\frac{0.8(1-p)}{0.8-0.6p}$	$\frac{0.2(1-p)}{0.2+0.5p}$	0
1	$\frac{0.2p}{0.8-0.6p}$	$\frac{0.7p}{0.2+0.5p}$	1

b) Given the loss, the posterior expected loss is

$$
\begin{aligned}
\mathrm{E}\left[L(\theta,0)\mid x=0\right] &= 1\times\frac{0.2p}{0.8-0.6p}\\
\mathrm{E}\left[L(\theta,1)\mid x=0\right] &= 0.5\times\frac{0.8(1-p)}{0.8-0.6p}=\frac{0.4(1-p)}{0.8-0.6p}\\
\mathrm{E}\left[L(\theta,0)\mid x=1\right] &= 1\times\frac{0.7p}{0.2+0.5p}\\
\mathrm{E}\left[L(\theta,1)\mid x=1\right] &= 0.5\times\frac{0.2(1-p)}{0.2+0.5p}=\frac{0.1(1-p)}{0.2+0.5p}\\
\mathrm{E}\left[L(\theta,0)\mid x=2\right] &= 1\times 1\\
\mathrm{E}\left[L(\theta,1)\mid x=2\right] &= 0.5\times 0.
\end{aligned}
$$

c) The Bayes estimator minimizes the posterior risk. Hence, the Bayes estimator is

$$
\begin{aligned}
\delta(0) &= \begin{cases}1 & \text{if } p\geq 2/3\\ 0 & \text{if } p<2/3\end{cases}\\
\delta(1) &= \begin{cases}1 & \text{if } p\geq 1/8\\ 0 & \text{if } p<1/8\end{cases}\\
\delta(2) &= 1.
\end{aligned}
$$

d) When $\theta=0$, the frequentist risk is

$$
\begin{aligned}
\mathrm{E}\left[L(\theta,\delta(X))\mid\theta\right] &= L(0,\delta(0))\Pr(X=0\mid\theta=0)+L(0,\delta(1))\Pr(X=1\mid\theta=0)\\
&\quad +L(0,\delta(2))\Pr(X=2\mid\theta=0)\\
&= L(0,\delta(0))\times\frac{4}{5}+L(0,\delta(1))\times\frac{1}{5}.
\end{aligned}
$$

When $\theta=1$, the frequentist risk is

$$
\begin{aligned}
\mathrm{E}\left[L(\theta,\delta(X))\mid\theta\right] &= L(1,\delta(0))\Pr(X=0\mid\theta=1)+L(1,\delta(1))\Pr(X=1\mid\theta=1)\\
&\quad +L(1,\delta(2))\Pr(X=2\mid\theta=1)\\
&= L(1,\delta(0))\times\frac{1}{5}+L(1,\delta(1))\times\frac{7}{10}+L(1,\delta(2))\times\frac{1}{10}.
\end{aligned}
$$

The integrated risk is

$$
\begin{aligned}
\mathrm{E}\left[L(\theta,\delta(X))\right] &= \int \mathrm{E}\left[L(\theta,\delta(X))\mid\theta\right]\pi(\theta)\,d\theta\\
&= \left[L(0,\delta(0))\times\frac{4}{5}+L(0,\delta(1))\times\frac{1}{5}\right](1-p)\\
&\quad +\left[L(1,\delta(0))\times\frac{1}{5}+L(1,\delta(1))\times\frac{7}{10}+L(1,\delta(2))\times\frac{1}{10}\right]p.
\end{aligned}
$$

Using the Bayes estimator, we obtain

(a) if $p<1/8$, then

$$
\begin{aligned}
\mathrm{E}\left[L(\theta,\delta(X))\right] &= \left[L(0,0)\times\frac{4}{5}+L(0,0)\times\frac{1}{5}\right](1-p)+\left[L(1,0)\times\frac{1}{5}+L(1,0)\times\frac{7}{10}+L(1,1)\times\frac{1}{10}\right]p\\
&= \left[0\times\frac{4}{5}+0\times\frac{1}{5}\right](1-p)+\left[1\times\frac{1}{5}+1\times\frac{7}{10}+0\times\frac{1}{10}\right]p\\
&= \frac{9}{10}p.
\end{aligned}
$$

(b) if $1/8 \leq p < 2/3$, then

$$\begin{aligned}
\mathrm{E}\left[L\left(\theta,\delta\left(X\right)\right)\right] &= \left[L\left(0,0\right)\times\frac{4}{5}+L\left(0,1\right)\times\frac{1}{5}\right]\left(1-p\right)+\left[L\left(1,0\right)\times\frac{1}{5}+L\left(1,1\right)\times\frac{7}{10}+L\left(1,1\right)\times\frac{1}{10}\right]p \\
&= \left[0\times\frac{4}{5}+\frac{1}{2}\times\frac{1}{5}\right]\left(1-p\right)+\left[1\times\frac{1}{5}+0\times\frac{7}{10}+0\times\frac{1}{10}\right]p \\
&= \frac{1-p}{10}.
\end{aligned}$$

(c) if $p \geq 2/3$, then

$$\begin{aligned}
\mathrm{E}\left[L\left(\theta,\delta\left(X\right)\right)\right] &= \left[L\left(0,1\right)\times\frac{4}{5}+L\left(0,1\right)\times\frac{1}{5}\right]\left(1-p\right)+\left[L\left(1,1\right)\times\frac{1}{5}+L\left(1,1\right)\times\frac{7}{10}+L\left(1,1\right)\times\frac{1}{10}\right]p \\
&= \left[\frac{1}{2}\times\frac{4}{5}+\frac{1}{2}\times\frac{1}{5}\right]\left(1-p\right)+\left[0\times\frac{1}{5}+0\times\frac{7}{10}+0\times\frac{1}{10}\right]p \\
&= \frac{1-p}{2}.
\end{aligned}$$

e) The least favorable prior maximizes the integrated risk. Hence, we let $p = 2/3$.

Task 3

a) We can assume $X_1 \sim \mathrm{Bin}\left(11037, p_1\right)$ and $X_2 \sim \mathrm{Bin}\left(11034, p_2\right)$. They are independent.

b) The likelihood is

$$\begin{aligned}
L\left(p_1,p_2\right) &= \binom{n_1}{x_1}p_1^{x_1}\left(1-p_1\right)^{n_1-x_1}\times\binom{n_2}{x_2}p_2^{x_2}\left(1-p_2\right)^{n_2-x_2} \\
&= \binom{n_1}{x_1}\binom{n_2}{x_2}\exp\left\{x_1\log p_1+\left(n_1-x_1\right)\log\left(1-p_1\right)+x_2\log p_2+\left(n_2-x_2\right)\log\left(1-p_2\right)\right\}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\frac{\partial\log L\left(p_1,p_2\right)}{\partial p_1} &= \frac{x_1}{p_1}-\frac{n_1-x_1}{1-p_1}=\frac{x_1-n_1p_1}{p_1\left(1-p_1\right)}, \\
\frac{\partial\log L\left(p_1,p_2\right)}{\partial p_2} &= \frac{x_2-n_2p_2}{p_2\left(1-p_2\right)}.
\end{aligned}$$

The Fisher information matrix is

$$\mathcal{I}\left(p_1,p_2\right) = \begin{bmatrix}\frac{n_1}{p_1\left(1-p_1\right)} & 0 \\ 0 & \frac{n_2}{p_2\left(1-p_2\right)}\end{bmatrix}.$$

The Jeffreys prior is

$$\begin{aligned}
\pi\left(p_1,p_2\right) &\propto \sqrt{\det\left(\mathcal{I}\left(p_1,p_2\right)\right)}=\sqrt{\frac{n_1}{p_1\left(1-p_1\right)}\times\frac{n_2}{p_2\left(1-p_2\right)}} \\
&\propto p_1^{-1/2}\left(1-p_1\right)^{-1/2}p_2^{-1/2}\left(1-p_2\right)^{-1/2}.
\end{aligned}$$

c) To find the least favorable prior, we start with the Bayes decision rule and its frequentist risk. Consider the prior $p_1 \sim \mathrm{Beta}\left(a_1, b_1\right)$ and $p_2 \sim \mathrm{Beta}\left(a_2, b_2\right)$. The posterior is

$$\begin{aligned}
\pi\left(p_1,p_2 \mid x_1,x_2\right) &\propto L\left(p_1,p_2\right)\pi\left(p_1\right)\pi\left(p_2\right) \\
&\propto p_1^{x_1+a_1-1}\left(1-p_1\right)^{n_1-x_1+b_1-1}\times p_2^{x_2+a_2-1}\left(1-p_2\right)^{n_2-x_2+b_2-1} \\
&\sim \mathrm{Beta}\left(x_1+a_1,n_1-x_1+b_1\right)\times\mathrm{Beta}\left(x_2+a_2,n_2-x_2+b_2\right).
\end{aligned}$$

The Bayes decision rule under the L_2 loss is $\mathrm{E}[p_1 \mid x] = \frac{x_1+a_1}{n_1+a_1+b_1}$ and $\mathrm{E}[p_2 \mid x] = \frac{x_2+a_2}{a_2+n_2+b_2}$. The frequentist risk is

$$
\begin{aligned}
R(\theta, \delta_B) &= \mathrm{E}\left[\left(\frac{x_1+a_1}{n_1+a_1+b_1} - p_1\right)^2 + \left(\frac{x_2+a_2}{n_2+a_2+b_2} - p_2\right)^2 \mid p_1, p_2\right] \\
&= \frac{\left[(a_1+b_1)^2 - n_1\right] p_1^2 + \left[n_1 - 2a_1(a_1+b_1)\right] p_1 + a_1^2}{(n_1+a_1+b_1)^2} \\
&\quad + \frac{\left[(a_2+b_2)^2 - n_2\right] p_2^2 + \left[n_2 - 2a_2(a_2+b_2)\right] p_2 + a_2^2}{(n_2+a_2+b_2)^2}.
\end{aligned}
$$

The numerator is a polynomial in θ. It is a constant if $(a_i+b_i)^2 = n_i$ and $n_i = 2a_i(a_i+b_i)$ for $i = 1, 2$. In such a case,

$$
R(\theta, \delta_B) = \frac{a_1^2}{(n_1+a_1+b_1)^2} + \frac{a_2^2}{(n_2+a_2+b_2)^2} \text{ is a constant.}
$$

Hence, the Bayes decision rule is minimax. The solutions of a_i and b_i are $a_i = \sqrt{n_i}/2$ and $b_i = \sqrt{n_i}/2$. These are the hyperparameters for the least favorable prior.

d) We have derived above the posterior for p_1 and p_2 as

$$
\begin{aligned}
\pi(p_1, p_2 \mid x_1, x_2) &\propto L(p_1, p_2)\, \pi(p_1)\, \pi(p_2) \\
&\propto p_1^{x_1+a_1-1} (1-p_1)^{n_1-x_1+b_1-1} \times p_2^{x_2+a_2-1} (1-p_2)^{n_2-x_2+b_2-1} \\
&\sim \mathrm{Beta}(x_1+a_1, n_1-x_1+b_1) \times \mathrm{Beta}(x_2+a_2, n_2-x_2+b_2).
\end{aligned}
$$

The least favourable prior corresponds to $a_i = \sqrt{n_i}/2$ and $b_i = \sqrt{n_i}/2$. The Jeffreys prior corresponds to $a_i = b_i = 1/2$.

e) We can draw a posterior sample from $\pi(p_1, p_2 \mid x_1, x_2)$, then compute the ratio $\theta = p_1/p_2$. Repeat the procedure N times. Then we obtain a posterior sample for θ.

f) To test the hypothesis H_0: $\theta \geq 1$, we can compute the posterior probability $\Pr(H_0 \mid x)$. It seems $\Pr(H_0 \mid x) > 0.5$ for both priors. If we choose the $0-1$ loss, then we cannot reject H_0.

Task 4

a) The posterior is

$$
\begin{aligned}
\pi(\theta \mid x) &\propto \left[\prod_{i=1}^{n} \frac{\theta^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} \exp\{-\theta x_i\}\right] \times \theta^{\alpha_0-1} \exp(-\beta_0 \theta) \\
&\propto \theta^{n\alpha+\alpha_0-1} \exp\left\{-\theta\left(\sum_{i=1}^{n} x_i + \beta_0\right)\right\} \\
&\sim \mathrm{Gamma}\left(n\alpha+\alpha_0, \sum_{i=1}^{n} x_i + \beta_0\right).
\end{aligned}
$$

b) The Bayes factor is

$$
B_{01}(x) = \frac{\displaystyle\int_{\Theta_0} f_0(x \mid \theta)\, \pi_0(\theta)\, d\theta}{\displaystyle\int_{\Theta_1} f_1(x \mid \theta)\, \pi_1(\theta)\, d\theta},
$$

where

$$
\begin{aligned}
\int_{\Theta_0} f_0(x \mid \theta)\,\pi_0(\theta)\,d\theta &= f_0(x \mid \theta = 5) = \left[\prod_{i=1}^{n} \frac{5^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} \exp\{-5x_i\}\right] \text{ Dirac measure,} \\
\int_{\Theta_1} f_1(x \mid \theta)\,\pi_1(\theta)\,d\theta &= \int_0^{\infty} \left[\prod_{i=1}^{n} \frac{\theta^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} \exp\{-\theta x_i\}\right] \times \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0-1} \exp(-\beta_0\theta)\,d\theta \\
&= \frac{\beta_0^{\alpha_0}}{\Gamma^n(\alpha)\,\Gamma(\alpha_0)} \left[\prod_{i=1}^{n} x_i^{\alpha-1}\right] \int_0^{\infty} \theta^{n\alpha+\alpha_0-1} \exp\left\{-\theta\left(\sum_{i=1}^{n} x_i + \beta_0\right)\right\} d\theta \\
&= \frac{\Gamma(n\alpha+\alpha_0)\,\beta_0^{\alpha_0}}{\Gamma^n(\alpha)\,\Gamma(\alpha_0)\left(\sum_{i=1}^{n} x_i + \beta_0\right)^{n\alpha+\alpha_0}} \left[\prod_{i=1}^{n} x_i^{\alpha-1}\right].
\end{aligned}
$$

Hence,

$$
B_{01}(x) = \frac{\frac{5^{n\alpha}}{\Gamma^n(\alpha)}\left[\prod_{i=1}^{n} x_i^{\alpha-1}\right] \exp\left\{-5\sum_{i=1}^{n} x_i\right\}}{\frac{\Gamma(n\alpha+\alpha_0)\beta_0^{\alpha_0}}{\Gamma^n(\alpha)\Gamma(\alpha_0)\left(\sum_{i=1}^{n} x_i+\beta_0\right)^{n\alpha+\alpha_0}}\left[\prod_{i=1}^{n} x_i^{\alpha-1}\right]} = \frac{5^{n\alpha}\Gamma(\alpha_0)\left(\sum_{i=1}^{n} x_i + \beta_0\right)^{n\alpha+\alpha_0} \exp\left\{-5\sum_{i=1}^{n} x_i\right\}}{\Gamma(n\alpha+\alpha_0)\,\beta_0^{\alpha_0}}.
$$

c) If $B_{10} = 200$, then we have strong evidence against H_0 if we use the rule-of-thumb.

Task 5 Skip the sufficient statistics in Q5(b)

a) Note that

$$
f(x_1, ..., x_n \mid \theta) = \lambda^{nk} k^n \exp\left\{-\lambda^k \sum_{i=1}^{n} x_i^k\right\} \prod_{i=1}^{n} x_i^{k-1},
$$

where we cannot separate k from x^k. Hence, the Weibull distribution with $\theta = (\lambda, k)$ does not belong to exponential family.

b) If $k = 1$, then

$$
f(x_1, ..., x_n \mid \lambda) = \lambda^n \exp\left\{-\lambda \sum_{i=1}^{n} x_i\right\}.
$$

The sufficient statistic is $T = \sum_{i=1}^{n} x_i$ and the natural parameter is λ. ($-\lambda$ can also be used as natural parameter.)

c) Note that

$$
\begin{aligned}
\frac{d \log f(x_1, ..., x_n \mid \theta)}{d\lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i, \\
\frac{d^2 \log f(x_1, ..., x_n \mid \theta)}{d\lambda^2} &= -\frac{n}{\lambda^2}.
\end{aligned}
$$

Hence, the Fisher information is $\mathcal{I}(\lambda) = n\lambda^{-2}$.

d) The conjugate prior is determined by the form of the likelihood. Hence, we must have

$$
\pi(\lambda) \propto \lambda^{a-1} \exp\{-b\lambda\},
$$

which is a Gamma density.

e) The posterior under the conjugate prior is

$$\pi(\lambda \mid x_1, ..., x_n) \quad \propto \quad \lambda^{n+a-1} \exp\left\{-\lambda\left(\sum_{i=1}^{n} x_i + b\right)\right\},$$

which is a Gamma $(n+a, \sum_{i=1}^{n} x_i + b)$.

f) The Jeffreys prior is

$$\pi(\lambda) \quad \propto \quad \sqrt{\mathcal{I}(\lambda)} \propto \lambda^{-1}.$$

g) The Jeffreys prior is an improper prior, but we can obtain it by letting $a = 0$ and $b = 0$.

Task 6

a) When $\lambda = 1$, we have $\mathrm{E}[X] = \Gamma\left(1 + k^{-1}\right)$ and median $(X) = [\log 2]^{1/k}$. Given that the chance for high wind speed is high, we expect a high mean and median. Hence, we want the prior of k to be high for relatively large k, and low for small k.

b) The Bayes estimator of θ under the L_2 loss is the posterior mean $\mathrm{E}[\theta \mid x_1, ..., x_n]$. The main steps of Metropolis-Hastings is given as follows. Choose an initial state $\theta^{(0)}$. For each iteration t, we Sample a candidate θ^* from a proposal distribution $T\left(\theta^{(t)}, \theta \mid x\right)$, calculate the ratio $R\left(\theta^{(t)}, \theta^*\right) = \frac{\pi(\theta^* \mid x) T\left(\theta^*, \theta^{(t)}\right)}{\pi\left(\theta^{(t)} \mid x\right) T\left(\theta^{(t)}, \theta^*\right)}$, draw $U \sim \mathrm{U}[0, 1]$, and update $\theta^{(t+1)}$ by

$$\theta^{(t+1)} \quad = \quad \begin{cases} \theta^*, & \text{if } U \leq R\left(\theta^{(t)}, \theta^*\right), \\ \theta^{(t)}, & \text{otherwise.} \end{cases}$$

We drop a burn-in period and obtain a Markov chain of R iterations. After obtaining a posterior sample from MCMC, we approximate the posterior mean by $R^{-1} \sum_{r=1}^{R} \theta^{(r)}$.

Task 7

a) The model is

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & x_1 & z_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}.$$

b) Under the conjugate prior for $\beta \sim N\left(\mu_0, \Lambda_0^{-1}\right)$, using the results from the lectures, the posterior is $\beta \mid y \sim N\left(\mu_n, \Lambda_n^{-1}\right)$, where

$$\begin{aligned} \Lambda_n &= \Lambda_0 + X^T \Sigma^{-1} X, \\ \mu_n &= \Lambda_n^{-1}\left(\Lambda_0 \mu_0 + X^T \Sigma^{-1} y\right), \\ \mu_0 &= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ \Lambda_0^{-1} &= \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \Sigma &= 0.25\boldsymbol{I}. \end{aligned}$$

c) The MAP estimator maximizes the posterior of β, that is, the mode of $\beta \mid y \sim N\left(\mu_n, \Lambda_n^{-1}\right)$.

d) The MAP estimator is $\beta = \mu_n$.

e) If μ_0 were 0, the MAP corresponds to a ridge regression estimator. When μ_0 is not zero, the penalization term is

$$(\beta - \mu_0)^T \Lambda_0 (\beta - \mu_0).$$

Task 8 Skip.

It is rejection-acceptance sampling from posterior distribution, which is not included in this course.