Solution to Exam 20220412 Zwanzig

March 13, 2025

Task 1

There is typo in the task. The probability $\Pr(x=0 \mid \theta=0)$ should be 0.1 instead of 0.01.

1. he posterior probability can be updated by $\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta)$ as

$$\begin{aligned}
\pi(\theta=2 \mid x) &\propto (0.6)^{n_1}(0.1)^{n_2}0^{n_3}(0.3)^{n_4}\times 0.2,\\
\pi(\theta=1 \mid x) &\propto (0.1)^{n_1}(0.2)^{n_2}(0.1)^{n_3}(0.6)^{n_4}\times 0.6,\\
\pi(\theta=0 \mid x) &\propto 0^{n_1}(0.1)^{n_2}(0.8)^{n_3}(0.1)^{n_4}\times 0.2,
\end{aligned}$$

Let

$$\begin{aligned}
m(x) &= (0.6)^{n_1}(0.1)^{n_2}0^{n_3}(0.3)^{n_4}\times 0.2+(0.1)^{n_1}(0.2)^{n_2}(0.1)^{n_3}(0.6)^{n_4}\times 0.6\\
&\quad +0^{n_1}(0.1)^{n_2}(0.8)^{n_3}(0.1)^{n_4}\times 0.2.
\end{aligned}$$

Then the posterior probability is,

$$\begin{aligned}
\pi(\theta=2 \mid x) &= \frac{(0.6)^{n_1}(0.1)^{n_2}0^{n_3}(0.3)^{n_4}\times 0.2}{m(x)},\\
\pi(\theta=1 \mid x) &= \frac{(0.1)^{n_1}(0.2)^{n_2}(0.1)^{n_3}(0.6)^{n_4}\times 0.6}{m(x)},\\
\pi(\theta=0 \mid x) &= \frac{0^{n_1}(0.1)^{n_2}(0.8)^{n_3}(0.1)^{n_4}\times 0.2}{m(x)}.
\end{aligned}$$

2. The MAP estimator is the estimator that maximizes the above posterior.

3. The likelihood is

$$\begin{aligned}
L(\theta=2) &= \frac{(n_1+n_2+n_3+n_4)!}{n_1!n_2!n_3!n_4!}(0.6)^{n_1}(0.1)^{n_2}0^{n_3}(0.3)^{n_4}\\
L(\theta=1) &= \frac{(n_1+n_2+n_3+n_4)!}{n_1!n_2!n_3!n_4!}(0.1)^{n_1}(0.2)^{n_2}(0.1)^{n_3}(0.6)^{n_4}\\
L(\theta=0) &= \frac{(n_1+n_2+n_3+n_4)!}{n_1!n_2!n_3!n_4!}0^{n_1}(0.1)^{n_2}(0.8)^{n_3}(0.1)^{n_4}.
\end{aligned}$$

The MLE maximizes the likelihood. The MAP adjusts the likelihood by the prior.

Task 2

There is a typo in the task. The probability $\Pr(x=1 \mid \theta=0)$ should be 0.7, instead of 0.

1. The posterior probability can be updated by $\pi(\theta \mid x) \propto f(x \mid \theta)\pi(\theta)$. The likelihood is

$$\begin{aligned}
f(n_-, n_0, n_1 \mid \theta=0) &\propto (0.1)^{n_{-1}}(0.2)^{n_0}(0.7)^{n_1},\\
f(n_-, n_0, n_1 \mid \theta=1) &\propto (0.8)^{n_{-1}}(0.2)^{n_0}0^{n_1}.
\end{aligned}$$

Then,

$$\begin{aligned}
\pi(\theta=0 \mid n_-, n_0, n_1) &\propto (0.1)^{n_{-1}} (0.2)^{n_0} (0.7)^{n_1} \times (1-p), \\
\pi(\theta=1 \mid n_-, n_0, n_1) &\propto (0.8)^{n_{-1}} (0.2)^{n_0} \times p.
\end{aligned}$$

The posterior is

$$\begin{aligned}
\pi(\theta=0 \mid n_-, n_0, n_1) &= \frac{(0.1)^{n_{-1}} (0.2)^{n_0} (0.7)^{n_1} \times (1-p)}{(0.1)^{n_{-1}} (0.2)^{n_0} (0.7)^{n_1} \times (1-p) + (0.8)^{n_{-1}} (0.2)^{n_0} \times p}, \\
\pi(\theta=1 \mid n_-, n_0, n_1) &= \frac{(0.8)^{n_{-1}} (0.2)^{n_0} \times p}{(0.1)^{n_{-1}} (0.2)^{n_0} (0.7)^{n_1} \times (1-p) + (0.8)^{n_{-1}} (0.2)^{n_0} \times p}.
\end{aligned}$$

2. The posterior expected loss is

$$\begin{aligned}
\mathrm{E}[L(\theta,0) \mid n_-, n_0, n_1] &= \pi(\theta=1 \mid n_-, n_0, n_1), \\
\mathrm{E}[L(\theta,1) \mid n_-, n_0, n_1] &= \pi(\theta=0 \mid n_-, n_0, n_1).
\end{aligned}$$

3. The Bayes estimator minimizes the posterior expected loss above. That is,

$$\begin{aligned}
\delta_B &= \begin{cases} 1 & \text{if } \pi(\theta=0 \mid n_-, n_0, n_1) < \pi(\theta=1 \mid n_-, n_0, n_1), \\ 0 & \text{if } \pi(\theta=0 \mid n_-, n_0, n_1) > \pi(\theta=1 \mid n_-, n_0, n_1), \end{cases} \\
&= \begin{cases} 1 & \text{if } p > \frac{\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1}}{1+\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1}}, \\ 0 & \text{if } p < \frac{\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1}}{1+\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1}}. \end{cases}
\end{aligned}$$

4. The frequentist risk of the Bayes estimator is

$$\begin{aligned}
& \mathrm{E}[L(0, \delta(n_-, n_0, n_1)) \mid \theta=0] \\
=\ & \Pr(\pi(\theta=0 \mid n_-, n_0, n_1) < \pi(\theta=1 \mid n_-, n_0, n_1) \mid \theta=0) \\
=\ & \Pr((0.1)^{n_{-1}} (0.2)^{n_0} (0.7)^{n_1} \times (1-p) < (0.8)^{n_{-1}} (0.2)^{n_0} \times p \mid \theta=0) \\
=\ & \Pr\left(\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1} \times (1-p) < p \mid \theta=0\right) \\
=\ & \sum_{n_{-1}, n_0, n_1} \frac{(n_{-1}+n_0+n_1)!}{n_{-1}! n_0! n_1!} (0.1)^{n_{-1}} (0.2)^{n_0} (0.7)^{n_1} 1\left\{\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1} \times (1-p) < p\right\} \\
& \mathrm{E}[L(1, \delta(n_-, n_0, n_1)) \mid \theta=1] \\
=\ & \Pr(\pi(\theta=0 \mid n_-, n_0, n_1) > \pi(\theta=1 \mid n_-, n_0, n_1) \mid \theta=0) \\
=\ & \Pr\left(\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1} \times (1-p) > p \mid \theta=1\right) \\
=\ & \sum_{n_{-1}, n_0, n_1} \frac{(n_{-1}+n_0+n_1)!}{n_{-1}! n_0! n_1!} (0.8)^{n_{-1}} (0.2)^{n_0} 1\left\{\left(\frac{1}{8}\right)^{n_{-1}} (0.7)^{n_1} \times (1-p) > p\right\},
\end{aligned}$$

where $1\{\}$ is the indicator function. The integrated risk is

$$\begin{aligned}
\mathrm{E}[L(\theta, \delta(X))] &= \int \mathrm{E}[L(\theta, \delta(X)) \mid \theta] \pi(\theta) \, d\theta \\
&= \mathrm{E}[L(0, \delta(n_-, n_0, n_1)) \mid \theta=0](1-p) + \mathrm{E}[L(1, \delta(n_-, n_0, n_1)) \mid \theta=1] p,
\end{aligned}$$

which is the Bayes risk when plugging in the frequentist risks.

5. The least favorable prior maximizes the integrated risk. I don't think there exists a closed form expression here.

Task 3 Skip the sufficient statistics in Q3(a) and Prop 3.3.13 in Q3(c)

The Poisson probability mass function is wrong.

1. Note that

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

   Hence, it belongs to exponential family with $A(\theta) = \exp(-n\theta)$, $\eta(\theta) = \log\theta$, $T(x) = \sum_{i=1}^n x_i$, and $h(x) = \frac{1}{\prod_{i=1}^n x_i!}$. The sufficient statistic is T and the natural parameter is η.

2. Note that

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

   Hence, the Fisher information is $\mathcal{I}(\theta) = \frac{n}{\theta}$.

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

$$\pi(\theta) \propto \theta^{a-1}\exp\{-b\theta\}.$$

4. The conjugate family is a Gamma density.

5. The posterior under the conjugate prior is

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

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

6. The Jeffreys prior is

$$\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)} \propto \theta^{-1/2}.$$

7. The Jeffreys prior is an improper prior, but we can obtain it by letting $a = \frac{1}{2}$ and $b = 0$.

Task 4

1. The prior for θ is

$$\begin{aligned}\pi(\theta) &= \int \pi(\theta \mid \mu)\pi(\mu)\,d\mu = \int\left[\prod_{i=1}^n \pi(\theta_i \mid \mu)\right]\pi(\mu)\,d\mu \\ &= \int\left[\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{\sum_{i=1}^n \theta_i^2 - 2\mu\sum_{i=1}^n\theta_i + n\mu^2}{2}\right\}\right] \times \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\mu^2}{2}\right\}d\mu \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n+1}\exp\left\{-\frac{\sum_{i=1}^n\theta_i^2}{2}\right\}\int\left[\exp\left\{-\frac{(n+1)\mu^2 - 2\mu\sum_{i=1}^n\theta_i}{2}\right\}\right]d\mu \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^{n+1}\exp\left\{-\frac{\sum_{i=1}^n\theta_i^2}{2} + \frac{\left(\sum_{i=1}^n\theta_i\right)^2}{2(n+1)}\right\}\sqrt{2\pi(n+1)^{-1}} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{n+1}}\exp\left\{-\frac{\sum_{i=1}^n\theta_i^2}{2} + \frac{\left(\sum_{i=1}^n\theta_i\right)^2}{2(n+1)}\right\}.\end{aligned}$$

2. The posterior satisfies

$$\begin{aligned}
\pi(\theta \mid x) &\propto f(x \mid \theta)\,\pi(\theta) = \left[\prod_{i=1}^{n}\prod_{j=1}^{k} f(x_{ij} \mid \theta_i)\right]\pi(\theta) \\
&= \left[\prod_{i=1}^{n}\prod_{j=1}^{k} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x_{ij}^2 - 2x_{ij}\theta_i + \theta_i^2}{2}\right\}\right] \times \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{n+1}} \exp\left\{-\frac{\sum_{i=1}^n \theta_i^2}{2} + \frac{\left(\sum_{i=1}^n \theta_i\right)^2}{2(n+1)}\right\} \\
&\propto \exp\left\{-\frac{k\sum_{i=1}^n \theta_i^2 - 2\sum_{i=1}^n \sum_{j=1}^k x_{ij}\theta_i}{2}\right\} \exp\left\{-\frac{\sum_{i=1}^n \theta_i^2}{2} + \frac{\left(\sum_{i=1}^n \theta_i\right)^2}{2(n+1)}\right\} \\
&\propto \exp\left\{-\frac{(k+1)\sum_{i=1}^n \theta_i^2 - 2\sum_{i=1}^n \sum_{j=1}^k x_{ij}\theta_i}{2} + \frac{\left(\sum_{i=1}^n \theta_i\right)^2}{2(n+1)}\right\}.
\end{aligned}$$

3. The posterior distribution of $\bar{\theta}$ can be obtained by change of variables as

$$\begin{bmatrix} \bar{\theta} \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} = \begin{bmatrix} n^{-1} & n^{-1} & \cdots & n^{-1} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}.$$

4. To test the hypothesis H_0: $-0.2 \le \bar{\theta} \le 0.2$, we can compute the posterior probability $\Pr\left(-0.2 \le \bar{\theta} \le 0.2 \mid \boldsymbol{X}\right)$. We reject H_0 if such posterior probability is below 0.5. I don't think you can derive the closed form expression.

Task 5 Skip Q5(d)

1. The MLE in Model 1 is $\hat{p} = X/1000$ and the MLE in Model 2 is also $\hat{\lambda} = X/1000$.

2. Model 1 is the beta-binomial model. The posterior is $\text{Beta}(1+x, 20+n-x)$. The Bayes estimator under the L_2 loss is $\frac{1+x}{21+n}$, where $x = 15$ and $n = 1000$.

3. Model 2 is the Poisson-gamma model. The posterior is $\text{Gamma}(x+20, n+1)$. The Bayes estimator under the L_2 loss is $\frac{20+x}{1+n}$, where $x = 15$ and $n = 1000$.

4. Skip Q5(d)

5. For the beta-binomial model,

$$\begin{aligned}
\int_{\Theta_1} f_1(x \mid \theta_1)\,\pi_1(\theta_1)\,d\theta_1 &= \int_0^1 \binom{n}{x} p^x (1-p)^{n-x} \frac{1}{B(a_0,b_0)} p^{a_0-1}(1-p)^{b_0-1}\,dp \\
&= \binom{n}{x} \frac{B(a_0+x, b_0+n-x)}{B(a_0,b_0)},
\end{aligned}$$

where $a_0 = 1$ and $b_0 = 20$. For the Gamma-poisson model,

$$\begin{aligned}
\int_{\Theta_2} f_2(x \mid \theta_2)\,\pi_2(\theta_2)\,d\theta_2 &= \int_0^\infty \frac{\lambda^x}{x!}\exp(-\lambda) \cdot \frac{b_1^{a_1}}{\Gamma(a_1)} \lambda^{a_1-1} \exp(-b_1\lambda)\,d\lambda \\
&= \frac{1}{x!} \frac{b_1^{a_1}\Gamma(a_1+x)}{(b_1+1)^{a_1+x}\,\Gamma(a_1)},
\end{aligned}$$

where $a_1 = 20$ and $b_1 = 1$. The Bayes factor is

$$B_{12} = \frac{\int_{\Theta_1} f_1(x \mid \theta_1)\pi_1(\theta_1)\, d\theta_1}{\int_{\Theta_2} f_2(x \mid \theta_2)\pi_2(\theta_2)\, d\theta_2} = \frac{\binom{n}{x}\frac{B(a_0+x,b_0+n-x)}{B(a_0,b_0)}}{\frac{1}{x!}\frac{b_1^{a_1}\Gamma(a_1+x)}{(b_1+1)^{a_1+x}\Gamma(a_1)}}.$$

6. Using the rule of thumb, we have positive evidence against Model 1.

Task 6 Skip Method 3 in Q6(b) and Q6(f)

1. The integral to be calculated is

$$\int_\Theta f(x \mid p)\pi(p)\, dp = \int_0^1 \frac{1}{B(1,20)} p^{1-1}(1-p)^{20-1} \times \binom{1000}{15} p^{15}(1-p)^{1000-15}\, dp$$

   which is the marginal likelihood of Model 1.

2. Method 1 is importance sampling with uniform distribution as the importance distribution. Method 2 is independent Monte Carlo, sampling directly from the beta distribution. Method 3 is quadrature approximation. Method 4 is Metropolis algorithm.

3. In Method 2, we draw R independent samples from $\pi(p)$ and the independent Monte Carlo approximation is

$$\frac{1}{R}\sum_{r=1}^{R} f\left(x \mid \theta^{(r)}\right).$$

4. In Method 4, For each iteration t, we Sample a candidate θ^* from a proposal distribution $T\left(\theta^{(t)}, \theta \mid x\right)$, calculate the ratio $R\left(\theta^{(t)}, \theta^*\right) = \frac{\pi(\theta^* \mid x)}{\pi\left(\theta^{(t)} \mid x\right)}$, draw $U \sim \mathrm{U}[0,1]$, and update $\theta^{(t+1)}$ by

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

   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 $R^{-1}\sum_{r=1}^{R}\theta^{(r)}$.

5. The results look different mainly because of uncertainty in sampling. I prefer Method 1 and 2 because of its simplicity in this case. We haven't studied comparison of importance sampling and independent Monte Carlo in this course.

Task 7

1. The least squares estimator minimizes

$$\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2.$$

   The minimizer is

$$\begin{aligned} \hat\beta &= \frac{\sum_{i=1}^{n}(x_i - \bar x)(y_i - \bar y)}{\sum_{i=1}^{n}(x_i - \bar x)^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}, \\ \hat\alpha &= \bar y - \hat\beta \bar x = \bar y. \end{aligned}$$

2. The prior is $(\alpha, \beta) \sim N\left(\mu_0,\ \Lambda_0^{-1}\right)$ where

$$
\begin{aligned}
\mu_0 &= \begin{bmatrix} a \\ b \end{bmatrix}, \\
\Lambda_0^{-1} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\
\Sigma &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\end{aligned}
$$

Using the result in the slides, the posterior is

$$
\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \mid \text{data} \quad \sim \quad N\left(\begin{bmatrix} 7 & 0 \\ 0 & 5 \end{bmatrix}^{-1}\left(\mu_0 + X^T y\right),\ \begin{bmatrix} 7 & 0 \\ 0 & 5 \end{bmatrix}^{-1}\right),
$$

where

$$
X = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}.
$$

3. The MAP estimator is $\begin{bmatrix} 7 & 0 \\ 0 & 5 \end{bmatrix}^{-1}\left(\mu_0 + X^T y\right)$.

4. Let $\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, $z = \begin{bmatrix} 1 \\ x \end{bmatrix}$, $\mu_0 = \begin{bmatrix} a \\ b \end{bmatrix}$, and $\mu_n = \left(Z^T Z + I\right)^{-1}\left(\mu_0 + Z^T y\right)$. The marginal distribution of Y is given by

$$
\begin{aligned}
f(y) &= \int\int f(y \mid \theta)\, \pi(\theta)\, d\theta \\
&= \int \frac{1}{(2\pi)^3} \exp\left\{-\frac{1}{2}(y - Z\theta)^T (y - Z\theta)\right\} \times \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(\theta - \mu_0)^T (\theta - \mu_0)\right\} d\theta \\
&= \frac{1}{(2\pi)^4} \exp\left\{-\frac{1}{2}\left(y^T y + \mu_0^T \mu_0\right)\right\} \int \exp\left\{-\frac{1}{2}\left[\theta^T\left(Z^T Z + I\right)\theta - 2\left(\mu_0 + Z^T y\right)^T \theta\right]\right\} d\theta \\
&= \frac{1}{(2\pi)^4} \exp\left\{-\frac{1}{2}\left(y^T y + \mu_0^T \mu_0 - \mu_n^T\left(Z^T Z + I\right)\mu_n\right)\right\} \int \exp\left\{-\frac{1}{2}(\theta - \mu_n)^T\left(Z^T Z + I\right)(\theta - \mu_n)\right\} d\theta \\
&= \frac{1}{(2\pi)^4} \exp\left\{-\frac{1}{2}\left(y^T y + \mu_0^T \mu_0 - \mu_n^T\left(Z^T Z + I\right)\mu_n\right)\right\} 2\pi \sqrt{\det\left\{\left(Z^T Z + I\right)^{-1}\right\}} \\
&= \frac{\sqrt{\det\left\{\left(Z^T Z + I\right)^{-1}\right\}}}{(2\pi)^3} \exp\left\{-\frac{1}{2}\left(y^T y + \mu_0^T \mu_0 - \left(\mu_0 + Z^T y\right)^T\left(Z^T Z + I\right)^{-1}\left(\mu_0 + Z^T y\right)\right)\right\}.
\end{aligned}
$$

Task 8 Ignore.