Assignment 1
Bayesian Inference FTN0548

Usama Zafar, Li Ju

February 25, 2024

1 Problem 2.1

With the given $p(x|\theta)$ in the table and $p(\theta)$, the joint distribution and marginal distributions of x and θ are computed as follows and shown in Table 1.

$$p(x,\theta) = p(x|\theta)p(\theta)$$
$$p(x) = \sum_{\theta\in\{0,1,2\}} p(x,\theta)$$

	D	C	B	A	$p(\theta)$
$\theta = 0$	0.08	0.01	0.01	0	0.1
$\theta = 1$	0.12	0.3	0.12	0.06	0.6
$\theta = 2$	0	0.03	0.15	0.12	0.3
$p(x)$	0.2	0.34	0.28	0.18	

Table 1: Joint and marginal distributions of x and θ.

The posterior distribution of θ given x is computed as follows and shown in Table 2.

$$p(\theta|x) = \frac{p(x,\theta)}{p(x)}$$

	D	C	B	A
$\theta = 0$	0.4	0.03	0.04	0
$\theta = 1$	0.6	0.88	0.43	0.33
$\theta = 2$	0	0.09	0.53	0.67

Table 2: Posterior distribution of θ given x.

2 Problem 2.2

Proof. $T(X)$ is sufficient for underlying parameter θ if $p(X|T(X))$, does not depend on θ, formally as $p(X|T(X)) = p(X|T(X),\theta)$ [Wikipedia, 2024].

First, we prove that $\pi(\theta|X) = \pi(\theta|T(X))$ implies $p(X|T(X)) = p(X|T(X), \theta)$.

$$\begin{aligned}
\pi(\theta|X) &= \pi(\theta|T(X)) \\
\frac{p(X|\theta)\pi(\theta)}{p(X)} &= \frac{p(T(X)|\theta)\pi(\theta)}{p(T(X))} \\
p(T(X)|X)\frac{p(X)}{p(T(X))} &= \frac{p(X|\theta)}{p(T(X)|\theta)}p(T(X)|X) \\
p(X|T(X)) &= p(X|T(X), \theta)
\end{aligned}$$

All steps above are if-and-only-if implications, thus we have also proved that $p(X|T(X)) = p(X|T(X), \theta)$ implies $\pi(\theta|X) = \pi(\theta|T(X))$. $\square$

3 Problem 2.7

(a) Given $\pi(\mu) = \mathcal{N}(0, 1)$ and $\pi(\theta_i|\mu) = \mathcal{N}(\mu, 1)$ for all $i \in \{1, 2, \dots, n\}$, we have

$$\begin{aligned}
\pi(\theta_i) &= \int \pi(\theta_i, \mu)d\mu = \int \pi(\theta_i|\mu)\pi(\mu)d\mu \\
&= \int \mathcal{N}_{\theta_i}(\mu, 1)\mathcal{N}_\mu(0, 1)d\mu \\
&\propto \int \exp\left\{-\frac{1}{2}(\theta_i - \mu)^2\right\} \exp(-\frac{1}{2}\mu^2)d\mu
\end{aligned}$$

We recognize that the integration results in the convolution of two standard Gaussian distribution $\mathcal{N}(0, 1)$. Recall that the convolution of $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$ results $\mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$, we have $\pi(\theta_i) = \mathcal{N}(0, 2)$.

Given that θ_i for all $i \in \{1, 2, \dots, n\}$ are independent, we have the prior distribution

$$\pi(\boldsymbol{\theta}) = \mathcal{N}(\mathbf{0}, 2\boldsymbol{I})$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots \theta_n)^\top$.

(b) Let $\boldsymbol{X}_i$ denotes $(X_{i,1}, X_{i,2}, \dots, X_{i,k})^\top$ for all $i \in \{1, 2, \dots, n\}$. Given $\pi(\theta_i) = \mathcal{N}(0, 2)$ and $p(\boldsymbol{X}_i|\theta_i) = \prod_{j=1}^k \mathcal{N}(\theta_i, 1)$, we have

$$\begin{aligned}
\pi(\theta_i|\boldsymbol{x}) = \pi(\theta_i|\boldsymbol{x}_i) &= \frac{p(\boldsymbol{x}_i|\theta_i)\pi(\theta_i)}{p(\boldsymbol{x}_i)} \\
&\propto \prod_{j=1}^k \exp\left\{-\frac{1}{2}(\theta_i - x_{i,j})^2\right\} \exp\left\{-\frac{1}{2}\left(\frac{\theta_i}{2}\right)^2\right\} \\
&\propto \exp\left\{-\frac{(k+1/4)\theta_i^2 - 2\theta_i\sum_{j=1}^k x_{i,j}}{2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{\theta_i - \sum_{j=1}^k \frac{x_{i,j}}{k+1/4}}{\frac{1}{\sqrt{k+1/4}}}\right)^2\right\} \\
\pi(\theta_i|\boldsymbol{x}) &= \mathcal{N}\left(\sum_{j=1}^k \frac{4x_{i,j}}{4k+1}, \frac{4}{4k+1}\right)
\end{aligned}$$

Given that $\theta_1, \dots, \theta_n$ are independent, we have

$$\begin{aligned}\pi(\boldsymbol{\theta}|\boldsymbol{x}) &= \prod_{i=1}^{n} \pi(\theta_i|\boldsymbol{x}) \\ &= \mathcal{N}\left((\mu_1^p, \dots, \mu_n^p)^\top, \frac{4}{1+4k}\boldsymbol{I}\right)\end{aligned}$$

where $\mu_i^p = \sum_{j=1}^{k} \frac{4x_{i,j}}{4k+1}, \forall i \in \{1, \dots, n\}$.

(c) Given that $\pi(\theta_i|\boldsymbol{x}) = \mathcal{N}\left(\sum_{j=1}^{k} \frac{4x_{i,j}}{4k+1}, \frac{4}{4k+1}\right)$ for all $i \in \{1, \dots, n\}$, we have

$$\pi\left(\frac{\theta_i}{n}|\boldsymbol{x}\right) = \mathcal{N}\left(\sum_{j=1}^{k} \frac{4x_{i,j}}{n(4k+1)}, \frac{4}{(4k+1)n^2}\right)$$

Then for $\hat{\theta} = \sum_{i=1}^{n} \frac{\theta_i}{n}$, we have

$$\pi\left(\hat{\theta}|\boldsymbol{x}\right) = \mathcal{N}\left(\sum_{i=1}^{n}\sum_{j=1}^{k} \frac{4x_{i,j}}{n(4k+1)}, \frac{4}{(4k+1)n}\right)$$

4 Problem 3.2

(a) Given the prior information that the parameter μ lies symmetrically around 3, the location parameter of the Cauchy prior can be set as $m = 3$.

(b)

$$\begin{aligned}\Pr(\mu > 10) > 0.3 \Longrightarrow \mathrm{CDF}(\mu = 10) &\leq \frac{7}{10} \\ \frac{1}{\pi}\arctan\left(\frac{10-3}{\gamma}\right) + \frac{1}{2} &\leq \frac{7}{10} \\ \gamma &\geq \frac{7}{\tan\frac{\pi}{5}} \approx 9.63\end{aligned}$$

For simplicity, we can set $\gamma = 10$ and thus, we have prior for μ as $\pi(\mu) = \mathcal{C}(3, 10)$.

(c) Notice that the log-normal likelihood function belongs to the exponential family while the Cauchy prior does not, $\pi(\mu) = \mathcal{C}(3, 10)$ is not a conjugate prior.

(d) Given $p(x|\mu) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(\ln x - \mu)^2\}$, the score function $V(\mu|x)$ and the Fisher information $I(\mu)$ are derived as follows:

$$\begin{aligned}V(\mu|x) &= \frac{\partial \ln p(x|\mu)}{\partial \mu} = \frac{1}{p(x|\mu)}\frac{\partial p(x|\mu)}{\partial \mu} \\ &= \frac{\partial - \frac{1}{2\sigma^2}(\ln x - \mu)^2}{\partial \mu} = \frac{\ln x - \mu}{\sigma^2} \\ I(\mu) &= -\mathbb{E}_{p(x|\mu)}\left[\frac{\partial V(\mu|x)}{\partial \mu}\right] \quad \text{(Holds when } \ln p(x|\mu) \text{ is twice differentiable w.r.t } \mu.) \\ &= \frac{1}{\sigma^2}\end{aligned}$$

Then Jeffreys prior is given by $\pi_{\mathrm{Jeff}}(\mu) \propto \sqrt{\det(I(\mu))} = \frac{1}{\sigma}$.

5 Problem 3.4

(a) Let $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$. Given $X_1, \ldots, X_n$ are i.i.d and $p(X_i) = \text{Geo}(\theta)$, we have

$$\begin{aligned} p(\boldsymbol{X}|\theta) &= \prod_{i=1}^{n} p(X_i|\theta) = \prod_{i_1}^{n} \text{Geo}_{X_i}(\theta) \\ &= (1-\theta)^{\sum_{i=1}^{n} X_i} \theta^n \\ &= \exp\left\{ \ln(1-\theta) \sum_{i=1}^{n} X_i + n \ln\theta \right\} \end{aligned}$$

Therefore, the sample distribution $p(\boldsymbol{X}|\theta)$ belongs to exponential family, with natural parameter $\zeta(\theta) = \ln(1-\theta)$ and sufficient statistics $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$.

(b) We denote the natural parameter with $\zeta = \ln(1-\theta)$, and we can rewrite the probability function as follows

$$p(\boldsymbol{X}|\zeta) = \exp\left\{ \zeta \sum_{i=1}^{n} X_i - (-n\ln(1-\exp\zeta)) \right\}$$

Using Theorem 3.3, we have $h(\boldsymbol{X} = 1$, $\Phi(\zeta) = -n\ln(1-\exp\zeta)$. With $\mu, \lambda \in \mathbb{R}$ and $\lambda > 0$, the the conjugate family over ζ is given by

$$\mathcal{F} = \{p(\zeta|\mu,\lambda)\} \propto \exp\{\zeta\mu - \lambda\Phi(\zeta)\}$$

The equivalent conjugate prior over θ is given by

$$\begin{aligned} \mathcal{F} = \{p(\theta|\mu,\lambda)\} &\propto \exp\{\mu\ln(1-\theta) + \lambda n \ln\theta\} \\ &= \theta^{\lambda n}(1-\theta)^{\mu} \end{aligned}$$

(c) The conjugate family for θ is the beta distribution family.

(d) With the prior over ζ as $\pi(\zeta|\mu,\lambda)$, the conjugate posterior $\pi(\zeta|x_i, \ldots, x_n)$ is given by

$$\begin{aligned} \pi(\zeta|x_i, \ldots, x_n) &\propto \pi(\zeta) p(\boldsymbol{x}|\zeta) \\ &= \exp\{\zeta\mu + \lambda n \ln(1-\exp\zeta)\} \cdot \exp\left\{ \zeta \sum_{i=1}^{n} x_i + (n\ln(1-\exp\zeta)) \right\} \\ &= \exp\left\{ \zeta \left(\mu + \sum_{i=1}^{n} x_i \right) + (\lambda+1) n \ln(1-\exp\zeta) \right\} \end{aligned}$$

The equivalent posterior over θ $\pi(\theta|x_1, \ldots, x_n)$ is given by

$$\pi(\theta|x_1, \ldots, x_n) \propto \theta^{\lambda n + n}(1-\theta)^{\mu + T(\boldsymbol{x})}$$

where $T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$. By normalizing the distribution, $\pi(\theta|x_1, \ldots, x_n) = \text{Beta}(\lambda n + n + 1, \mu + T(\boldsymbol{x}) + 1)$.

(e) The Fisher information $I(\theta)$ is derived as follows:

$$\begin{aligned}
V(\theta|\boldsymbol{x}) &= \frac{\partial p(\boldsymbol{x}|\theta)}{p(\boldsymbol{x}|\theta)\partial\theta} \\
&= \frac{T(\boldsymbol{x})}{\theta-1} + \frac{n}{\theta} \\
I(\theta) &= -\mathbb{E}_{p(\boldsymbol{X}|\theta)}\left[\frac{\partial V(\theta|\boldsymbol{x})}{\partial\theta}\right] \\
&= \mathbb{E}_{p(\boldsymbol{X}|\theta)}\left[\frac{T(\boldsymbol{x})}{(1-\theta)^2} + \frac{n}{\theta^2}\right] \\
&= \frac{\mathbb{E}[T(\boldsymbol{x})]}{(1-\theta)^2} + \frac{n}{\theta^2}
\end{aligned}$$

We know that $\mathbb{E}[T(\boldsymbol{x})] = \sum_{i=1}^{n} \mathbb{E}[x_i] = n(1-\theta)/\theta$. Thus we have

$$T(\theta) = \frac{n}{\theta(1-\theta)} + \frac{n}{\theta^2} = \frac{n}{\theta^2(1-\theta)}$$

(f) With the Fisher information $I(\theta) = \frac{n}{\theta^2(1-\theta)}$, Jeffreys prior is

$$\pi_{\text{Jeff}}(\theta) \propto \sqrt{\det(I(\theta))} = \frac{1}{\theta}\sqrt{\frac{n}{1-\theta}}$$

(g) Jeffreys prior can be reformed as follows

$$\begin{aligned}
\pi_{\text{Jeff}}(\theta) &\propto \frac{1}{\theta}\sqrt{\frac{n}{1-\theta}} \\
&\propto \theta^{-1}(1-\theta)^{-1/2}
\end{aligned}$$

in a similar form of Beta$(0, \frac{1}{2})$. Recall that Beta(a, b) is undefined for $a = 0$, Jeffreys prior does not belong to Beta distribution family, implying that it is not a conjugate prior.

6 Problem 3.7

(a)

$$\begin{aligned}
p(X|\theta) &\propto \theta^X (1-\theta)^{n-X} \\
&= \exp\left\{X \ln\theta + (n-X)\ln(1-\theta)\right\} \\
&= \exp\left\{X \ln\left(\frac{\theta}{1-\theta}\right) + n\ln(1-\theta)\right\}
\end{aligned}$$

Thus, $p(X|\theta)$ belongs to the exponential family with the natural parameter $\zeta = \ln\left(\frac{\theta}{1-\theta}\right)$.

(b)

$$\begin{aligned}
&p(X|\zeta) \propto \exp\left\{\zeta X - \Phi(\zeta)\right\} \\
&\text{where } \Phi(\zeta) = n\ln(\exp(\zeta)+1)
\end{aligned}$$

With $\mu, \lambda \in \mathbb{R}$ and $\lambda > 0$, the conjugate family over ζ is given by

$$\mathcal{F} = \{p(\zeta|\mu,\lambda)\} \propto \exp\left\{\zeta\mu - \lambda\Phi(\zeta)\right\}$$

(c) With the prior $\pi(\zeta|\mu,\lambda)$ over ζ, the corresponding posterior over ζ is

$$\begin{aligned}\pi(\zeta|x) &\propto p(x|\zeta)\pi(\zeta)\\ &\propto \exp\{\zeta x - \Phi(\zeta)\}\cdot \exp\{\zeta\mu - \lambda\Phi(\zeta)\}\\ &= \exp\{\zeta(\mu+x) - (\lambda+1)n\ln(\exp(\zeta)+1)\}\end{aligned}$$

The equivalent posterior over θ is

$$\pi(\theta|x) \propto \exp\left\{\ln\left(\frac{\theta}{1-\theta}\right)(\mu+x) + (\lambda+1)n\ln(1-\theta)\right\}$$

(d) The conjugate prior $\pi(\theta|\mu,\lambda)$ can be reformed as

$$\begin{aligned}\pi(\theta|\mu,\lambda) &\propto \exp\left\{\ln\left(\frac{\theta}{1-\theta}\right)\mu + \lambda n\ln(1-\theta)\right\}\\ &= \left(\frac{\theta}{1-\theta}\right)^{\mu}\cdot(1-\theta)^{\lambda n}\\ &= \theta^{\mu}(1-\theta)^{\lambda n-\mu}\end{aligned}$$

It is recognized that the conjugate prior belongs to the Beta distribution family.

7 Problem 3.10

(a) Given $p(X|\theta) = \mathcal{N}(\theta, 1)$, we have

$$\begin{aligned}V(\theta|x) &= \frac{\partial p(x|\theta)}{p(x|\theta)\partial\theta}\\ &= \frac{\partial - \frac{1}{2}(x-\theta)^2}{\partial\theta}\\ &= x-\theta\\ I(\theta) &= -\mathbb{E}_{p(X|\theta)}\left[\frac{\partial V(\theta|x)}{\partial\theta}\right]\\ &= 1\end{aligned}$$

Thus, Jeffreys prior is given by $\pi_{\text{Jeff}}(\theta) \propto 1$. Then the corresponding posterior is derived as follows

$$\begin{aligned}\pi_{\text{Jeff}}(\theta|x) &\propto \pi_{\text{Jeff}}(\theta)\cdot p(x|\theta)\\ &\propto \exp\left\{-\frac{1}{2}(x-\theta)^2\right\}\end{aligned}$$

It is recognized that the posterior is a Gaussian $\pi_{\text{Jeff}}(\theta|x) = \mathcal{N}(x, 1)$.

(b)

8 Problem 3.12

(a)

References

Wikipedia. Sufficient statistic — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Sufficient%20statistic&oldid=1196916794, 2024. [Online; accessed 23-February-2024].