Homework 1 Solution

PubH 7440: Introduction to Bayesian Analysis

Spring 2020

1. (Carlin and Louis, Chapter 2, Problem 8)

- **a.**

  The Fisher information for θ being:

$$\mathbb{I}(\theta) = -\mathbb{E}[\frac{d^2}{d\theta^2}\log(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\theta)^2/2\sigma^2})] = \frac{1}{\sigma^2}$$

  Then $p(\theta) \propto \sqrt{\mathbb{I}(\theta)} = \sqrt{\frac{1}{\sigma^2}} \propto 1$

- **b.**

  The Fisher information for σ being:

$$\mathbb{I}(\sigma) = -\mathbb{E}[\frac{d^2}{d\sigma^2}\log(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-\theta)^2/2\sigma^2})] = -\mathbb{E}[\frac{1}{\sigma^2} - \frac{3(x-\theta)^2}{\sigma^4}] = \frac{2}{\sigma^2}$$

  Then $p(\sigma) \propto \sqrt{\mathbb{I}(\sigma)} = \sqrt{\frac{2}{\sigma^2}} \propto \frac{1}{\sigma}$

2. (Gelman, Chapter 2, Problem 5)

- **a.**

$$\begin{aligned}
\mathbf{Pr}(y = k) &= \int_0^1 \mathbf{Pr}(y = k|\theta)d\theta \\
&= \int_0^1 \mathbf{C}_n^k \theta^k (1-\theta)^{n-k} d\theta \\
&= \mathbf{C}_n^k \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \\
&= \frac{1}{n+1}
\end{aligned}$$

  Where the third equal sign uses the property of Beta distribution and the fourth equal sign comes from the fact that $\Gamma(x) = (x-1)!$

- **b.**

  The posterior distribution $\mathbf{Pr}(\theta|y) \propto \theta^y(1-\theta)^{(n-y)}\theta^{\alpha-1}(1-\theta)^{\beta-1} = \theta^{(y+\alpha-1)}(1-\theta)^{n+\beta-y-1}$, hence $\theta|y \sim \mathbf{Beta}(y+\alpha, n+\beta-y)$, and the posterior mean is $\frac{y+\alpha}{\alpha+n-\beta}$

Suppose there exists λ such that $\frac{y+\alpha}{\alpha+n-\beta} = \lambda\frac{\alpha}{\alpha+\beta} + (1-\lambda)\frac{y}{n}$, solve it for λ, $\lambda = \frac{\alpha+\beta}{\alpha+\beta+n}$. Since α, β, n are all positive, $\lambda \in (0,1)$, i.e. the posterior mean is a weighted average of the prior mean and the observed relative frequency of heads, this yields the required inequality.

- **c.**

Through the discussion in **a.**, when the prior is uniform, the posterior distribution is $\mathbf{Beta}(y+1, n-y+1)$, hence the posterior variance of θ is $\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}$, while the prior variance is $\frac{1}{12}$.

$$
\begin{aligned}
\frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} &\leq \frac{\frac{1}{4}(y+1+n-y+1)^2}{(n+2)^2(n+3)} \\
&= \frac{1}{4(n+3)} \\
&\leq \frac{1}{12}
\end{aligned}
$$

Where the first inequality comes from the fact: $(a+b)^2 \geq 4ab$, and the second inequality comes from the condition that $n \geq 1$

3. (Gelman, Chapter 2, Problem 8)

- **a.** From the lecture that for n i.i.d. observations $\mathbf{y} = (y_1, \ldots, y_n)$ from normal distribution $N(\theta, \sigma^2)$, and the prior $\pi(\theta) = N(\mu, \tau^2)$, we have:

$$p(\theta|\mathbf{y}) = N(\frac{\frac{\sigma^2}{n}\mu + \tau^2\bar{y}}{\frac{\sigma^2}{n} + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2})$$

Hence the posterior distribution for θ is:

$$N(\frac{\frac{20^2}{n}180 + 40^2 \times 150}{\frac{20^2}{n} + 40^2}, \frac{20^2 \times 40^2}{20^2 + n \times 40^2})$$

- **b.**

A normal likelihood with normal prior will result in a normal predictive distribution. The mean and variance can be found by iterated expection:

$$E[\tilde{y}|\mathbf{y}] = E[E[\tilde{y}|\theta, \mathbf{y}]|\mathbf{y}] = E[\theta|\mathbf{y}] = \frac{\frac{20^2}{n}180 + 40^2 \times 150}{\frac{20^2}{n} + 40^2}$$

$$Var[\tilde{y}|\mathbf{y}] = E[Var[\tilde{y}|\theta, \mathbf{y}]] + Var[E[\tilde{y}|\theta, \mathbf{y}]] = E[20^2|\mathbf{y}] + Var[\theta|\mathbf{y}] = 20^2 + \frac{20^2 \times 40^2}{20^2 + 40^2 n}$$

- **c.**

Example R code:

```r
  post_mean <- function(n){
  #function calculating posterio mean
  #same for posterior predictive mean
  (20^2 * 180/n + 40^2 * 150)/(20^2/n + 40^2)
}
```

```
post_var <- function(n){
  #calculating posterior variance
  (20^2 * 40^2 / n) / (20^2 / n + 40^2)
}

pred_var <- function(n){
  #calculating posterior predictive variance
  20^2 + (40^2 * 20^2 / n) / (20^2 / n + 40^2)
}
```

```
#### credible interval ####
round(qnorm(c(0.025, .975), mean = post_mean(10), sd = sqrt(post_var(10))),1)
```

```
## [1] 138.5 163.0
```

```
#### prediction interval ####
round(qnorm(c(0.025, .975), mean = post_mean(10), sd = sqrt(pred_var(10))),1)
```

```
## [1] 109.7 191.8
```

- **d.**

```
####credible interval###
round(qnorm(c(0.025, .975), mean = post_mean(100), sd = sqrt(post_var(100))),1)
```

```
## [1] 146.2 154.0
```

```
###prediction interval
round(qnorm(c(0.025, .975), mean = post_mean(100), sd = sqrt(pred_var(100))),1)
```

```
## [1] 110.7 189.5
```

4. (Gelman, Chapter 2, Problem 9)

- **a.**

  From the property of Beta distribution:

$$E(\theta) = \frac{\alpha}{\alpha+\beta} = 0.6$$
$$Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.3^2$$

  solve it for α, β: $\alpha = 1$ and $\beta = \frac{2}{3}$

- **b.**

$$p(\theta|\mathbf{y}) \propto \theta^{650}(1-\theta)^{350}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

  which means $\theta|\mathbf{y} \sim Beta(650+\alpha, 350+\beta)$. Hence $E(\theta|\mathbf{y}) = \frac{650+\alpha}{1000+\alpha+\beta} = 0.650$, $Var(\theta|\mathbf{y}) = \frac{(650+\alpha)(350+\beta)}{(1000+\alpha+\beta)^2(1000+\alpha+\beta+1)} = 0.000227$

  Example R code for plots:

  p = seq(0,1, length=1000)

  plot(p, dbeta(p, 651, 350.667), ylab="density", type ="l")

- **c.**

  Example R code for plots:

  ```
  p = seq(0.5,0.7, length=100)
  plot(p, dbeta(p, 650, 350), ylab="density", type ="l", col = 1)
  lines(p, dbeta(p, 651, 350.667), col = 2)
  lines(p, dbeta(p, 680, 370), col = 3)
  legend(0.5, 25, legend = c("Non-informative","Beta(1,2/3)","Beta(30,20)"),lty=c(1,1,1),col=c(1,2,3))
  ```

5. (Gelman, Chapter 2, Problem 14)

- **b.**

  For $k \leq n$, denote $\mathbf{y_k} = (y_1, \ldots, y_k)$, $\theta_k = \frac{\frac{\sigma^2}{k}\mu + \tau^2\bar{\mathbf{y_k}}}{\frac{\sigma^2}{k} + \tau^2}$, and $\sigma_k^2 = \frac{\sigma^2\tau^2}{\sigma^2 + k\tau^2}$

  For the case of $\mathbf{y_1}$, it is already shown in the lecture notes. Using mathematical induction, suppose the formula holds for all $\mathbf{y_k},\ k \leq m-1$, then

$$\begin{aligned}
p(\theta|\mathbf{y_m}) &\propto p(y_m|\theta)p(\theta|\mathbf{y_{m-1}}) \\
&\propto exp(-\frac{(y_m-\theta)^2}{2\sigma^2})exp(-\frac{(\theta-\theta_{m-1})^2}{2\sigma_{m-1}^2}) \\
&\propto exp(-\frac{\theta^2 - 2\frac{\sigma_{m-1}^2 y_m + \sigma^2\theta_{m-1}}{\sigma_{m-1}^2+\sigma^2}\theta}{2\frac{\sigma^2\sigma_{m-1}^2}{\sigma_{m-1}^2+\sigma^2}}) \\
&\propto exp(-\frac{(\theta-\theta_m)^2}{2\sigma_m^2})
\end{aligned}$$

  Hence $p(\theta|\mathbf{y_n}) \sim N(\theta_n, \sigma_n^2)$