Kelvin Niu

Homework 1

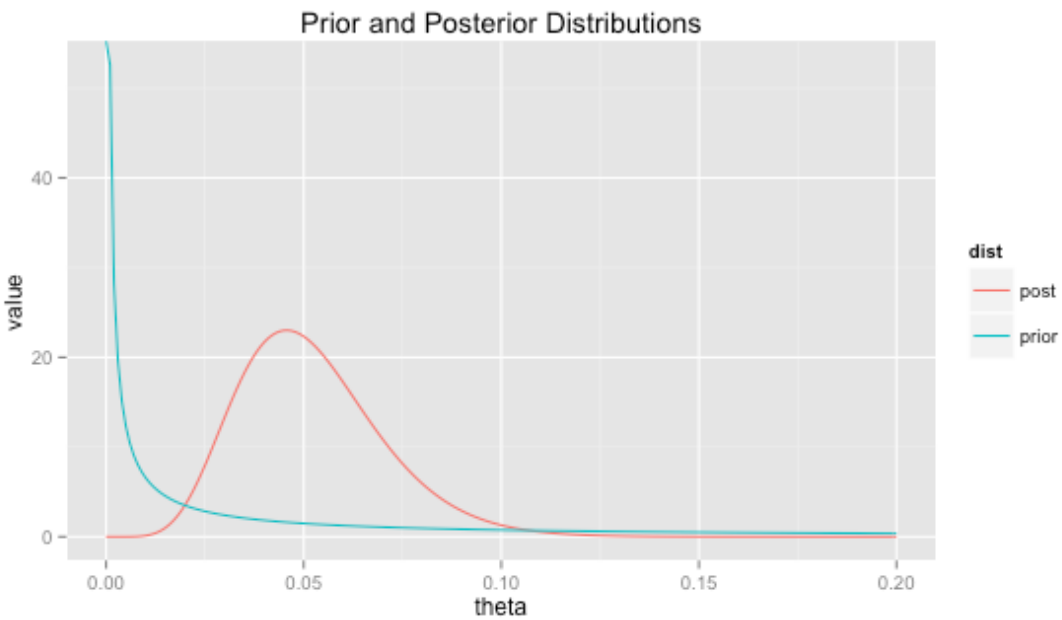
1. $$p(\theta \mid x_{1:n}) = p(x_{1:n} \mid \theta)\, p(\theta)$$
$$= \prod_{i=1}^{n} \theta e^{-\theta x_i} \mathbb{1}(x_i > 0) \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \mathbb{1}(\theta > 0)$$
$$\propto \theta^n e^{-\theta \sum_{i=1}^{n} x_i} \mathbb{1}(x_1, \dots, x_n > 0)\, \theta^{a-1} e^{-b\theta} \mathbb{1}(\theta > 0)$$
$$\propto \theta^{n+a-1} e^{-\theta(\sum_{i=1}^{n} x_i + b)} \mathbb{1}(x_1, \dots, x_n > 0) \mathbb{1}(\theta > 0)$$

$$\boxed{\begin{aligned} &p(\theta \mid x_{1:n}) \sim \text{Gamma}\left(n+a, \sum_{i=1}^{n} x_i + b\right) \\ &p(\theta \mid x_{1:n}) = \frac{(\sum_{i=1}^{n} x_i + b)^{n+a}}{\Gamma(n+a)} \theta^{n+a-1} e^{-\theta(\sum_{i=1}^{n} x_i + b)} \mathbb{1}(x_1, \dots, x_n > 0) \mathbb{1}(\theta > 0) \end{aligned}}$$

2.

Prior ~ Gamma(0.1, 1.0)

Posterior ~ Gamma(8.1, 155.5)



3. If a store owner wishes to model the wait time between customers in the morning in order to decide whether or not to take a break, an exponential model would be reasonable because customers generally come in the morning independently and at a constant rate. However, if the store owner wishes to model the wait time between customer for the whole day, an exponential model may not be appropriate since customers do not come at a constant rate throughout the day (customers may come more frequently in the morning than the afternoon). Exponential distributions can be used to model wait times between events in Poisson processes, which assume independent events occuring at a constant average rate.

4. $$\ell(s,a) = \begin{cases} 1, & \text{if } s \neq a \\ 0, & \text{if } s = a \end{cases}$$

$$\rho(a, x_{1:n}) = \mathbb{E}(\ell(\$, a) \mid x_{1:n})$$

$$= \sum_{\substack{\text{all possible} \\ \text{values of } \$}} \ell(s,a)\, \mathbb{P}(\$ = s \mid x_{1:n})$$

$$= \left[\sum_{\substack{\text{all possible} \\ \text{values of } \$}} (1)\, \mathbb{P}(\$ = s \mid x_{1:n}) \right] - \mathbb{P}(\$ = a \mid x_{1:n})$$

since the loss function does not take the value 1 only when $s = a$

$$= 1 - \mathbb{P}(\$ = a \mid x_{1:n})$$

Thus, the posterior loss $\rho(a, x_{1:n})$ is minimized when $\mathbb{P}(\$ = a \mid x_{1:n})^*$ is maximized

5. Intuitively, I would predict x_{n+1} based on the average of the observations $x_1, \ldots, x_n$

So,

$$\hat{x}_{n+1} = \begin{cases} 1, & \text{if } \bar{x} > 0.5 \\ 0, & \text{if } \bar{x} \leq 0.5 \end{cases} \qquad \text{where } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

From Question 4, we know that the action x_{n+1} that maximizes $\mathbb{P}(X_{n+1} = x_{n+1} \mid x_{1:n})$ is the action that minimizes posterior expected loss, thus is the action chosen by Bayes procedure

$$\hat{x}_{n+1} = \underset{x_{n+1}}{\arg\max}\ \mathbb{P}(X_{n+1} = x_{n+1} \mid x_{1:n})$$

Since $X_{n+1} \sim \text{Bern}(\theta)$, it only can take two values: 1 and 0

Thus we are comparing

$\mathbb{P}(X_{n+1} = 1 \mid x_{1:n}) = \theta$ and $\mathbb{P}(X_{n+1} = 0 \mid x_{1:n}) = 1 - \theta$

Since θ is unknown, our best estimate is the expected value of the posterior

$$\theta \mid x_{1:n} \sim \text{Beta}\left(a + \sum_{i=1}^{n} x_i,\ b + n - \sum_{i=1}^{n} x_i\right)$$

$$\begin{aligned}
\mathbb{E}(\theta \mid x_{1:n}) &= \int_0^1 \theta\, p(\theta \mid x_{1:n})\, d\theta \\
&= \int_0^1 \theta \frac{\theta^{a + \sum_{i=1}^{n} x_i - 1} (1-\theta)^{b + n - \sum_{i=1}^{n} x_i - 1}}{B\left(a + \sum_{i=1}^{n} x_i,\ b + n - \sum_{i=1}^{n} x_i\right)}\, d\theta \\
&= \frac{1}{B\left(a + \sum_{i=1}^{n} x_i,\ b + n - \sum_{i=1}^{n} x_i\right)} \int_0^1 \theta^{a + \sum_{i=1}^{n} x_i} (1-\theta)^{b + n - \sum_{i=1}^{n} x_i - 1}\, d\theta \\
&= \frac{B\left(a + \sum_{i=1}^{n} x_i + 1,\ b + n - \sum_{i=1}^{n} x_i\right)}{B\left(a + \sum_{i=1}^{n} x_i,\ b + n - \sum_{i=1}^{n} x_i\right)} \\
&= \frac{\Gamma\left(a + \sum_{i=1}^{n} x_i + 1\right) \Gamma\left(b + n - \sum_{i=1}^{n} x_i\right)}{\Gamma(a + b + n + 1)} \cdot \frac{\Gamma(a + b + n)}{\Gamma\left(a + \sum_{i=1}^{n} x_i\right) \Gamma\left(b + n - \sum_{i=1}^{n} x_i\right)} \\
&= \frac{a + \sum_{i=1}^{n} x_i}{a + b + n}
\end{aligned}$$

Thus, the resuting action is

$$\boxed{\hat{x}_{n+1} = \begin{cases} 1, & \text{if } \dfrac{a + \sum_{i=1}^{n} x_i}{a + b + n} > 0.5 \\ 0, & \text{if } \dfrac{a + \sum_{i=1}^{n} x_i}{a + b + n} \leq 0.5 \end{cases}}$$

6. when $a = b = 0$, the Bayes procedure is the same as my intuitive procedure.

A larger a makes the Bayes procedure more likely to predict 1 while a larger b makes the Bayes procedure more likely to predict 0

7. From Question 5, the posterior mean is

$$\mathbb{E}(\theta \mid x_{1:n}) = \frac{a + \sum_{i=1}^{n} x_i}{a+b+n}, \qquad \theta \mid x_{1:n} \sim \text{Beta}\left(a + \sum_{i=1}^{n} x_i,\ b+n-\sum_{i=1}^{n} x_i\right)$$

The prior mean is

$$\mathbb{E}(\theta) = \frac{a}{a+b}, \qquad \theta \sim \text{Beta}(a, b)$$

We are looking for a $t \in [0,1]$ such that

$$t\bar{x} + (1-t)\frac{a}{a+b} = \frac{a + \sum_{i=1}^{n} x_i}{a+b+n}$$

$$t\frac{1}{n}\sum_{i=1}^{n} x_i + (1-t)\frac{a}{a+b} = \left(\frac{1}{a+b+n}\right)a + \left(\frac{1}{a+b+n}\right)\sum_{i=1}^{n} x_i$$

Since a, b, n have no direct impact on individual x_i's, the coefficients of $\sum_{i=1}^{n} x_i$ must match on the left and right side of the equation, thus

$$t\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{a+b+n}\sum_{i=1}^{n} x_i$$

$$\boxed{t = \frac{n}{a+b+n}}$$

To check,

$$\frac{n}{a+b+n}\bar{x} + \left(1 - \frac{n}{a+b+n}\right)\frac{a}{a+b} = \frac{n}{a+b+n}\frac{1}{n}\sum_{i=1}^{n} x_i + \left(\frac{a+b}{a+b+n}\right)\frac{a}{a+b}$$

$$= \left(\frac{1}{a+b+n}\right)\sum_{i=1}^{n} x_i + \left(\frac{1}{a+b+n}\right)a$$

$$= \frac{a + \sum_{i=1}^{n} x_i}{a+b+n} \checkmark$$