STA 602 HW2

Ryan Tang

September 11th 2022

1 Excerise PH 3.3

(a)

$$
\begin{aligned}
Y_{i\in n} &\overset{iid}{\sim} Poisson(\theta) \\
p(Y|\theta) &= \prod_{i=1}^{n} \frac{1}{y_i!}\theta^{y_i}e^{-\theta} \\
&= c(y_1, y_2, \ldots, y_n)\theta^{\sum y_i}e^{-n\theta}
\end{aligned}
$$

$$
\begin{aligned}
p(\theta) &\sim Gamma(a, b) \\
&= \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}
\end{aligned}
$$

$$
\begin{aligned}
\therefore p(\theta|Y) &\propto p(Y|\theta)p(\theta) \\
&= \theta^{\sum y_i}e^{-n\theta}\theta^{a-1}e^{-b\theta} \\
&= \theta^{a-1+\sum y_i}e^{-\theta(b+n)} \\
&\sim Gamma(a + \sum_{i=1}^{n} y_i, b+n)
\end{aligned}
$$

Therefore, given $Y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6)$, $p(\theta_A|Y_A) \sim Gamma(237, 20)$. And similiarly, given $Y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)$, $p(\theta_B|Y_B) \sim Gamma(125, 14)$. The θ_A posterior has a mean of 11.85, variances of 0.593, and 95% confidence interval of $[10.39, 13.41]$. The θ_B posterior has a mean of 8.93, variances of 0.638, and 95% confidence interval of $[7.43, 10.56]$.

(b) It is hard to get the expectation of θ_B close to the θ_A posterior expectation. It only happens after $n0$ goes over 275+. Or we can also use a prior that is similar to $Gamma(14 \times n0, n0)$ to make it easier.

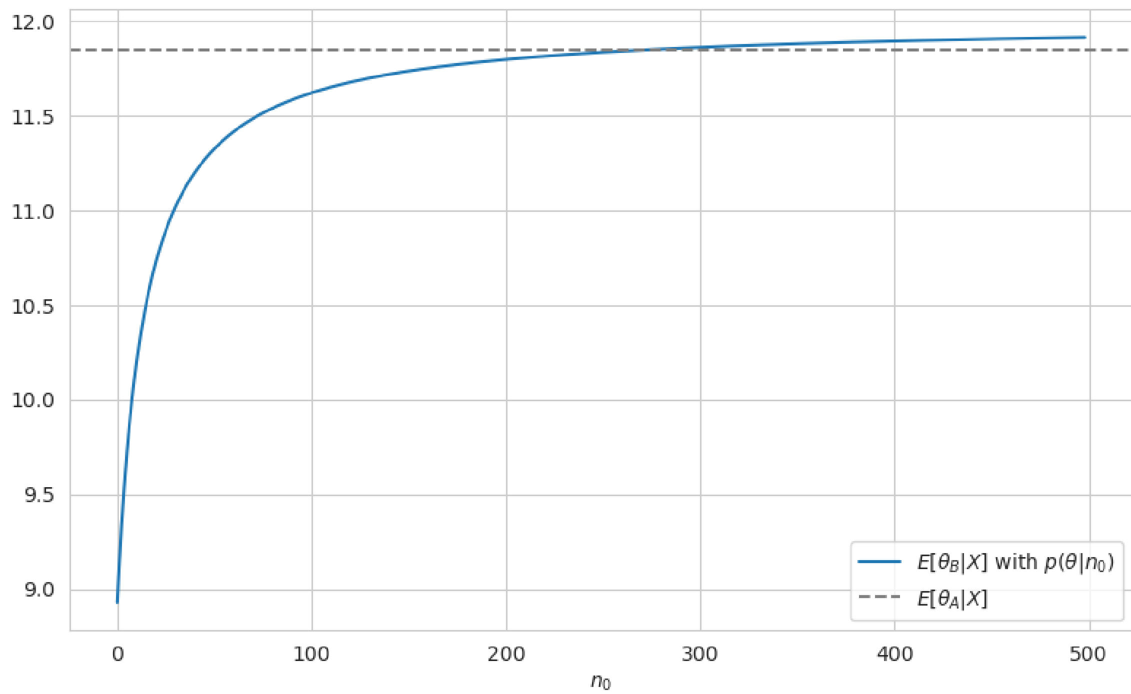

Figure 1: θ_B sensitivity to n_0

(c) I do not think it is a good idea to transfer the knowledge from θ_A to use it in θ_B inference based on the given information. They behave differently based on the data given. In other words, we can assume they are conditionally independent given the data **X**.

2 Excerise PH 3.5

(a) As long as we model both the sampling distribution and the prior all within the exponential family, the resulting posterior is also an exponential family distribution. The derivation also generalizes to priors that are a mixture of exponential family distributions. The resulting posterior is simply a weighted average of each prior $p(\phi)$ that is updated by the data Y separately.

$$
\begin{aligned}
p(y|\phi) &= h(y)c(\phi)e^{\phi t(y)} \\
p(Y|\phi) &= \prod_{i=1}^{n} h(y_i)c(\phi)e^{\phi t(y_i)} && Y_{i\in n} \overset{iid}{\sim} p(y|\phi) \\
&\propto c(\phi)^n \exp(\phi \sum_{i=1}^{n} t(y_i)) \\
\tilde{p}(\phi) &= \sum_{k=1}^{K} w_k p_k(\phi) && p(\phi) = \kappa(n_0, t_0)c(\phi)^{n_0} e^{n_0 t_0 \phi}
\end{aligned}
$$

$$
\begin{aligned}
\tilde{p}(\phi|Y) &\propto p(Y|\phi)\tilde{p}(\phi) \\
&\propto c(\phi)^n \exp(\phi \sum_{i=1}^{n} t(y_i)) \sum_{k=1}^{K} w_k c(\phi)^{n_0^{(k)}} \exp(n_0^{(k)} t_0^{(k)} \phi) \\
&\propto \sum_{k=1}^{K} w_k \cdot c(\phi)^{n_0^{(k)}+n} \exp(\phi \times (n_0^{(k)} t_0^{(k)} + \sum_{i=1}^{n} t(y_i))) \\
&\propto \sum_{k=1}^{K} w_k \cdot p_k(\phi | n_0^{(k)} + n, (n_0^{(k)} t_0^{(k)} + n\bar{t}(Y))/(n_0^{(k)} + n))
\end{aligned}
$$

(b) First we write the Poisson distribution in its exponential family form as below.

$$
\begin{aligned}
p(y|\theta) &= \frac{1}{y!}\theta^y e^{-\theta} \sim Poisson(\theta) \\
&\propto \theta^y e^{-\theta} \\
&= e^{y\log(\theta)} e^{-\theta} \\
&= \exp(y\phi - e^{\phi})
\end{aligned}
$$

$$
\begin{aligned}
\therefore \quad \phi &= \log(\theta) \\
c(\phi) &= \exp(-A(\phi)) = \exp(-e^{\phi}) \\
t(y) &= y \\
p(y|\theta) &\propto \exp(-e^{\phi})e^{\phi y}
\end{aligned}
$$

Hence, its conjugate prior Gamma has the form of $p(\phi|n_0, t_0) \propto \exp(-n_0 e^{\phi})e^{n_0 t_0 \phi}$. And the posterior distribution is

$$
\tilde{p}(\phi|Y) \propto \sum_{k=1}^{K} w_k \exp(\alpha^{(k)}\phi - \beta^{(k)} e^{\phi})
$$

$$
\alpha^{(k)} = n_0^{(k)} t_0^{(k)} + n\overline{y} \qquad \beta^{(k)} = n_0^{(k)} + n \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i
$$

3 Excerise PH 3.8

(a) To encode the belief that 20% of coins are symmetric and the remainings have a bias towards either head or tail, I will arbitrarily choose a total sample size of 18 for the mixture of beta priors assuming Diaconis and Ylvisaker should have at least examed these amounts of coin for their statement. To be precise, the resulting prior looks like below.

$$p(\theta) = 0.5 \cdot Beta(a = 6, b = 12) + 0.5 \cdot Beta(a = 12, b = 6)$$

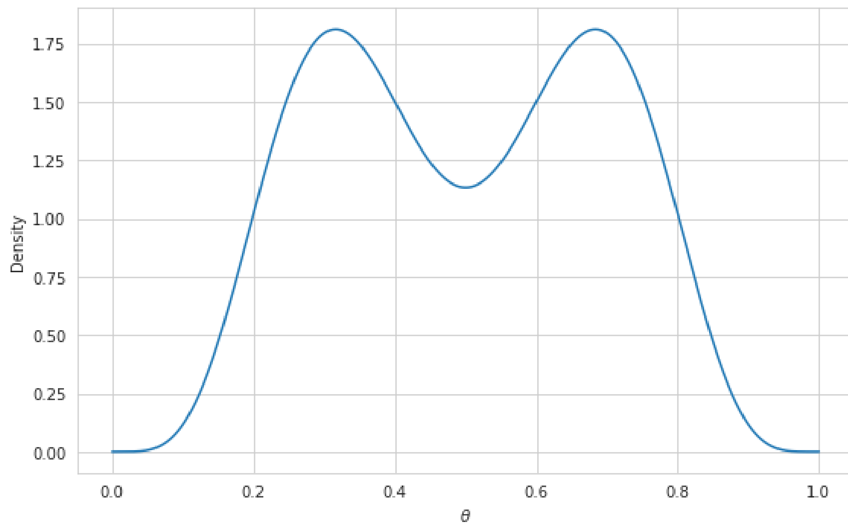

Figure 2: Coin Prior of θ

(b) I had a year-1999 penny tossed 50 times and obtained 29 heads, surprisingly.

(c) After incorporating the experiment data with a binomial likelihood, the resulting θ posterior for the year-1999 penny looks like the following.

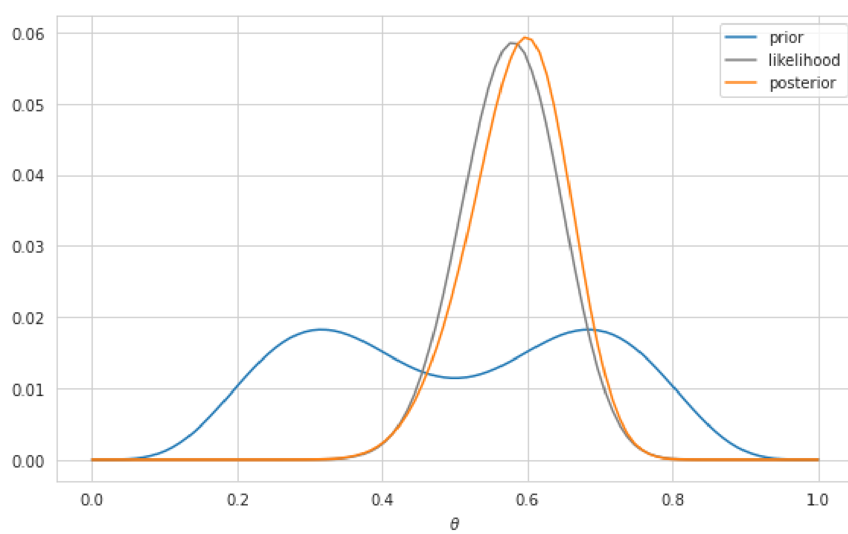

Figure 3: Coin Posterior of θ — 1999 penny

(d) After seeing the experiment result, I started to believe in Diaconis and Ylvisaker's statement more strongly. Nevertheless, with only one sample, I do not think it is appropriate to believe that the coin denomination or the year has anything to do with the bias. Hence, I only 1.5x times the sample size from 18 to 27.

This time with nickel from 2013, I obtained 21 heads instead out of 50 tosses. Therefore, here is the θ posterior for this nickel.

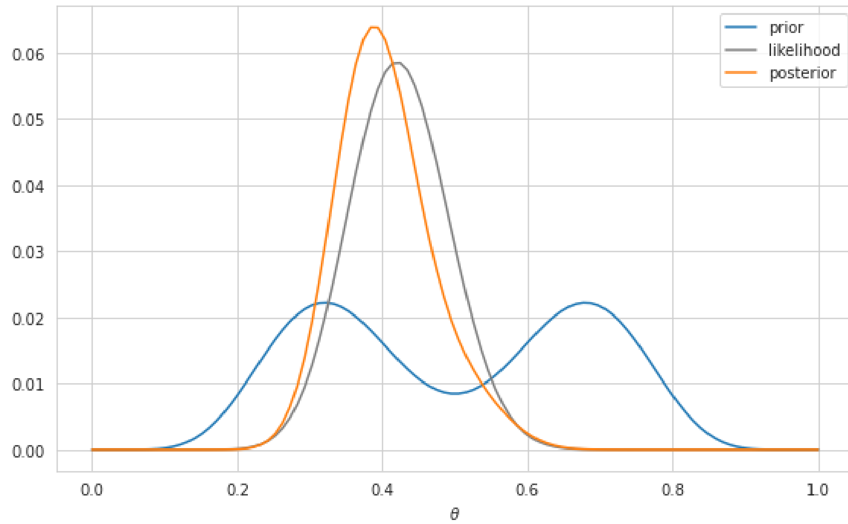

Figure 4: Coin Posterior of θ — 2013 nickel

4 Excerise PH 4.3

(a) Overall the Poisson generative model matches quite well with the data, Y_A.

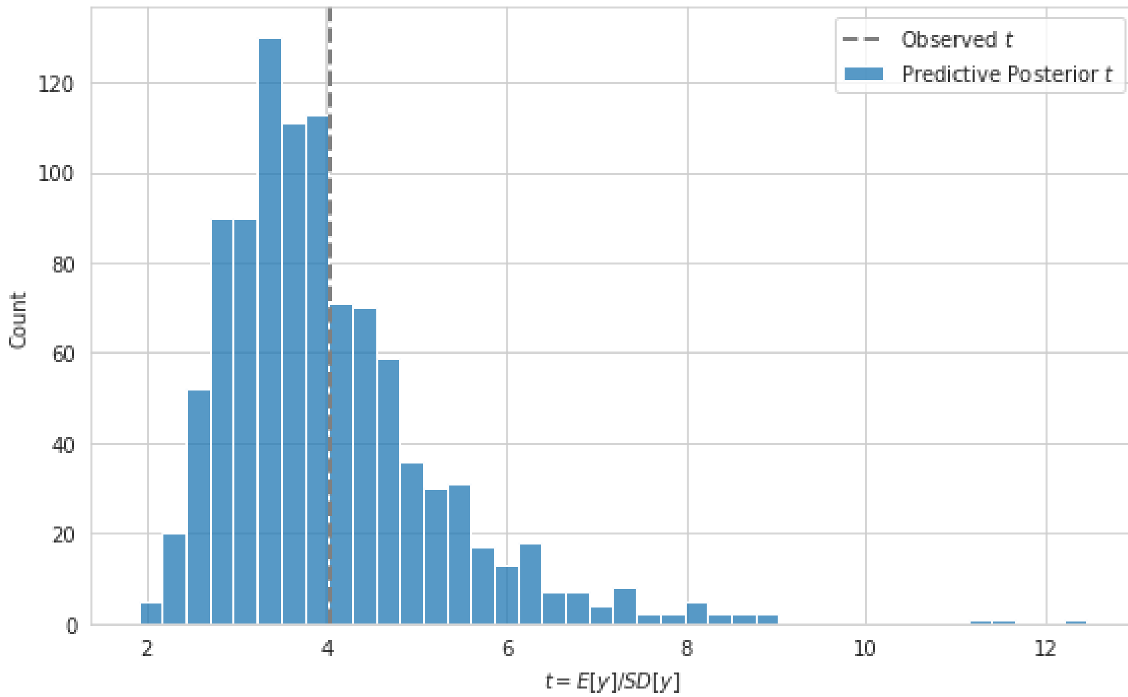

Figure 5: Posterior Predictive Model Check for Group A

(b) However, the model does not work well with the group B data, Y_B. It could be that we have a terrible prior, not enough data size, or perhaps Poisson is just not a good model for the data.

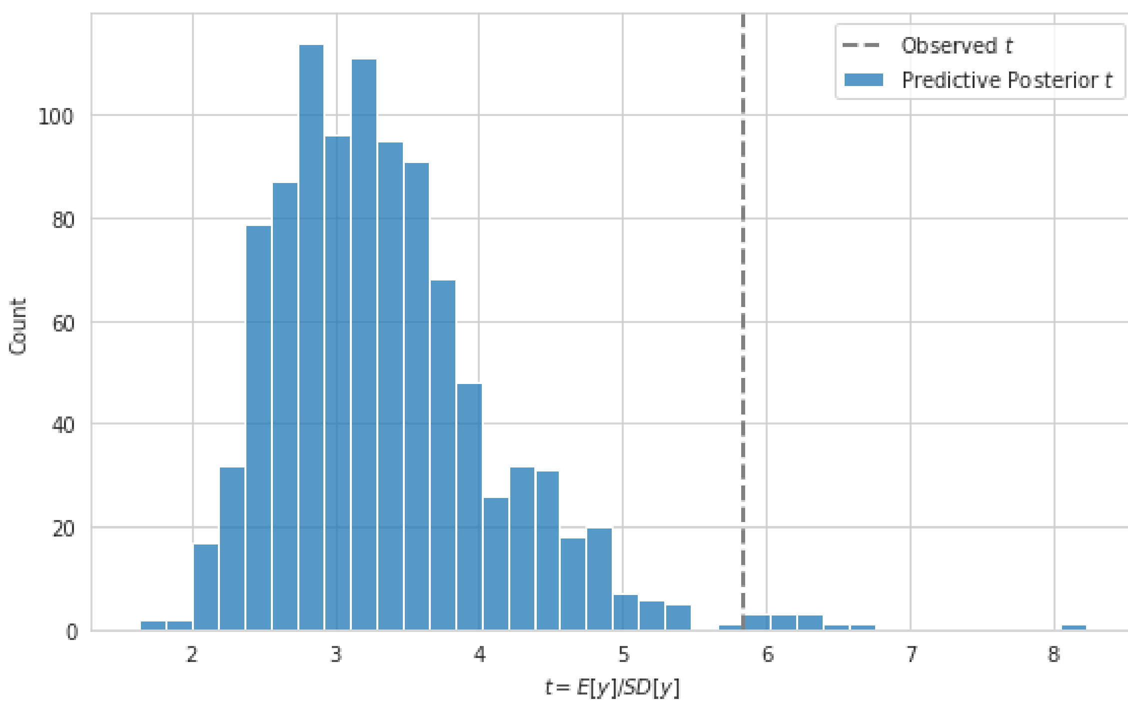

Figure 6: Posterior Predictive Model Check for Group B

5 Excerise 5

(a) Given the prior $\theta \sim Gamma(a = 10, b = 1)$ and the sampling model $X_{i \in n} \overset{iid}{\sim} Poisson(\theta)$, the posterior is also a gamma distribution. Now we know we received a total of 200 call from 10 different days, the Posterior

$$\theta|\mathbf{X} \sim Gamma(a = 210, b = 11)$$

The θ estimate that minimizes the L2 loss is the mean, 19.09. And the θ estimate that minimizes the L1 loss is the median, 19.06.

(b) The model is reasonable because observing 20 calls on average is a typical belief shown in the posterior. However, given only one data point, it is hard to tell the truth. The underlying data points might well be $\{0, 0, \ldots, 0, 200\}$. In such a case, our model performs poorly instead. In other words, if all we care about is the average, the model is fine.

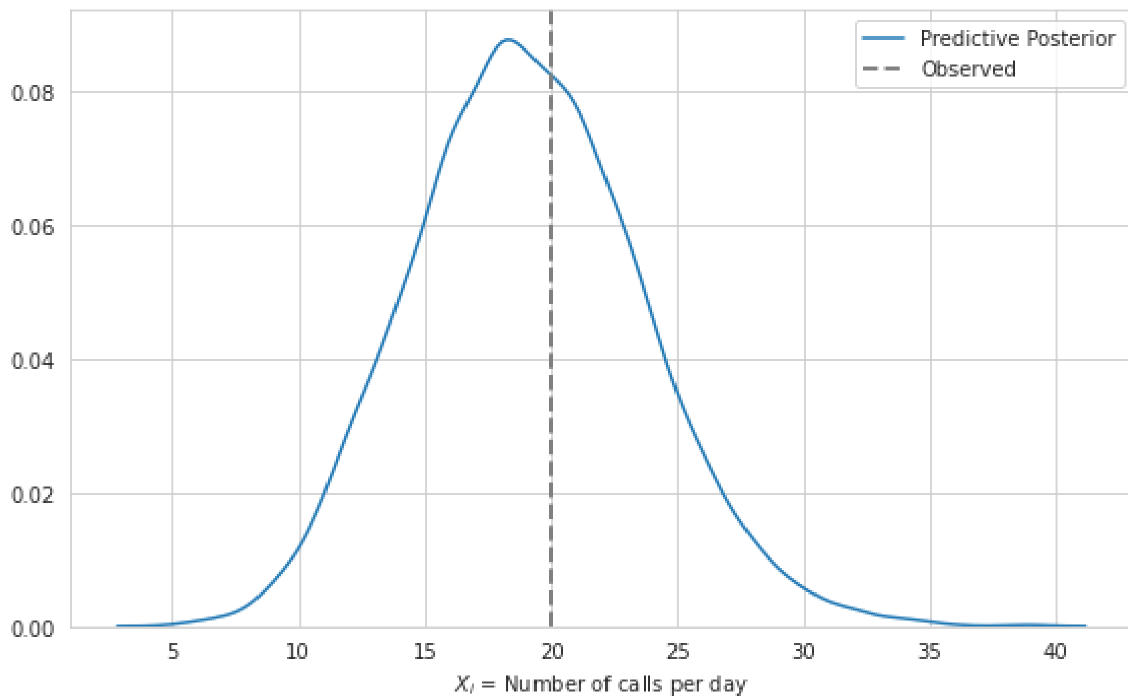

Figure 7: Posterior Predictive Model Check