STAT 638: Solution for Homework #3

3.14

a. $Y_1, \cdots, Y_n \sim$ i.i.d. binary(θ)
Likelihood function, $L(\theta) = \prod_{i=1}^n P(Y_i|\theta) = \prod_{i=1}^n \theta^{y_i}(1-\theta)^{1-y_i}$

$$
\begin{aligned}
l(\theta|y) &= \sum \log P(y_i|\theta) = \sum y_i \log\theta + \left(n - \sum y_i\right)\log(1-\theta) \\
\Leftrightarrow \frac{\partial l(\theta|y)}{\partial\theta} &= \frac{\sum y_i}{\theta} - \frac{n-\sum y_i}{1-\theta}
\end{aligned}
$$

Setting first derivative equal to zero implies that

$$\hat{\theta}_{\text{MLE}} = \frac{\sum y_i}{n}.$$

$$
\begin{aligned}
\frac{\partial^2 l(\theta|y)}{\partial\theta^2} &= -\frac{\sum y_i}{\theta^2} - \frac{n-\sum y_i}{(1-\theta)^2} \\
J(\theta) &= -\frac{\partial^2 l(\theta|y)}{\partial\theta^2} = \frac{\sum y_i}{\theta^2} + \frac{n-\sum y_i}{(1-\theta)^2} \\
\frac{J(\hat{\theta})}{n} &= \frac{\bar{y}}{\theta^2} + \frac{1-\bar{y}}{(1-\theta)^2} = \frac{1}{\bar{y}} + \frac{1}{1-\bar{y}}
\end{aligned}
$$

b.

$$
\begin{aligned}
\log P_u(\theta) &= \frac{l(\theta|y)}{n} + c = \bar{y}\log\theta + (1-\bar{y})\log(1-\theta) + c \\
P_u(\theta) &= e^c\theta^{\bar{y}}(1-\theta)^{1-\bar{y}} \\
\frac{\partial \log P_u(\theta)}{\partial\theta} &= \frac{\bar{y}}{\theta} - \frac{1-\bar{y}}{(1-\theta)^2} \\
J(\theta) &= -\frac{\partial^2 \log P_u(\theta)}{\partial\theta^2} = \frac{\bar{y}}{\theta^2} + \frac{1-\bar{y}}{(1-\theta)^2}
\end{aligned}
$$

c.

$$
\begin{aligned}
P_u(\theta) \times P(Y_1, \cdots, Y_n|\theta) &\propto \theta^{\bar{y}}(1-\theta)^{1-\bar{y}}\theta^{\sum y_i}(1-\theta)^{n-\sum y_i} \\
&= \theta^{\bar{y}}(1-\theta)^{1-\bar{y}}\theta^{n\bar{y}}(1-\theta)^{n-n\bar{y}} \\
&= \theta^{\bar{y}+n\bar{y}}(1-\theta)^{1-\bar{y}+n-n\bar{y}}
\end{aligned}
$$

Hence, this is Beta($\bar{y}+n\bar{y}+1, 2-\bar{y}+n-n\bar{y}$), which can be considered as a posterior distribution for θ.

d. Since $P(y|\theta)$ is Poisson distribution, likelihood function $L(\theta|y) = \prod_{i=1}^n \frac{e^{-\theta}\theta^{y_i}}{y_i!}$.

log-likelihood $l(\theta|y) = -n\theta + \sum y_i \log\theta - \sum \log y_i!$

$$
\begin{aligned}
\frac{\partial l(\theta|y)}{\partial\theta} &= 0 \Rightarrow -n + \frac{\sum y_i}{\theta} = 0 \Rightarrow \hat{\theta}_{MLE} = \frac{\sum y_i}{n} \\
\frac{\partial^2 l(\theta|y)}{\partial\theta^2} &= -\frac{\sum y_i}{\theta^2} \\
J(\theta) &= -\frac{\partial^2 l(\theta|y)}{\partial\theta^2} = \frac{\sum y_i}{\theta^2} \quad \text{and} \quad \frac{J(\hat{\theta})}{n} = \frac{\bar{y}}{\hat{\theta}^2} = \frac{1}{\bar{y}}
\end{aligned}
$$

$$
\begin{aligned}
\log P_u(\theta) &= \frac{l\left(\theta|y\right)}{n} + c \propto -\theta + \bar{y}\log\theta + c \\
P_u(\theta) &= e^c e^{-\theta}\theta^{\bar{y}} \propto \\
\frac{\partial \log P_u(\theta)}{\partial\theta} &= -1 + \frac{\bar{y}}{\theta} \\
\frac{\partial^2 \log P_u(\theta)}{\partial\theta^2} &= -\frac{\bar{y}}{\theta^2} \\
J(\theta) &= \frac{\bar{y}}{\theta^2}
\end{aligned}
$$

$$
P_u(\theta) \times P\left(Y_1, \cdots, Y_n|\theta\right) \propto e^{-\theta}\theta^{\bar{y}}e^{-n\theta}\theta^{\sum y_i} = e^{-\theta}\theta^{\bar{y}}e^{-n\theta}e^{n\bar{y}} = e^{-\theta(1+n)}\theta^{\bar{y}(1+n)}
$$

Hence, this is distributed as Gamma($(1+n)\bar{y}+1, 1+n$), and can be considered as a posterior distribution for θ.

4.1 From previous homework, the posterior distribution of θ_1 is Beta(58,44). Since the prior distribution is uniform(0,1), posterior distribution of θ_2 is

$$
P\left(\theta|Y\right) \propto P(\theta)P\left(Y|\theta\right) = \binom{50}{30}\theta^{30}(1-\theta)^{20}.
$$

Hence, the posterior distribution of θ_2 is Beta(31,21). $P\left(\theta_1 < \theta_2|\text{data and prior}\right) \approx 0.6302$

4.2

a. From homework 2, the posterior distribution of θ_A is Gamma(237,20) and that of θ_B is Gamma(125,14). From Monte Carlo sampling, $P\left(\theta_B < \theta_A|y_A, y_B\right) \approx 0.9942$.

b. The posterior distribution of θ_A is Gamma(237,20), and that of θ_B is Gamma($120n_0$+113, n_0+13). The plot shows that probability of $\theta_B < \theta_A$ drastically decreases when n_0 is relatively small. When n_0 is greater than 400, the probability of $\theta_B < \theta_A$ decreases very slowly. Therefore, the probability of $\theta_B < \theta_A$ is sensitive to the prior when n_0 is relatively small.

c. There are two ways to solve this problem. One is to compute the posterior predictive distribution and then sample from it. The other is to use the method which is suggested in 4.3 of Hoff.

$$
\begin{aligned}
P\left(\tilde{y}|y\right) &= \int_0^\infty P\left(\tilde{y}|\theta\right)P\left(\theta\right)d\theta = \int_0^\infty \frac{\theta^{\tilde{y}}e^{-\theta}}{\tilde{y}!} \times \frac{(n+\beta)^{\sum y_i+\alpha}}{\Gamma\left(y_i+\alpha\right)}\theta^{\sum y_i+\alpha-1}e^{-(n+\beta)^\theta}d\theta \\
&= \frac{(n+\beta)^{\sum y_i+\alpha}}{\Gamma\left(\sum y_i+\alpha\right)\Gamma\left(\tilde{y}+1\right)}\int_0^\infty \theta^{\tilde{y}+\sum y_i+\alpha-1}e^{-(n+\beta+1)^\theta}d\theta \\
&= \frac{(n+\beta)^{\sum y_i+\alpha}}{\Gamma\left(\sum y_i+\alpha\right)\Gamma\left(\tilde{y}+1\right)}\frac{\Gamma\left(\tilde{y}+\sum y_i+\alpha\right)}{(n+\beta+1)^{\tilde{y}+\sum y_i+\alpha}} \\
&= \frac{\Gamma\left(\tilde{y}+\sum y_i+\alpha\right)}{\Gamma\left(\sum y_i+\alpha\right)\Gamma\left(\tilde{y}+1\right)}\left(\frac{n+\beta}{n+\beta+1}\right)^{\sum y_i+\alpha}\left(\frac{1}{n+\beta+1}\right)^{\tilde{y}}
\end{aligned}
$$

Hence, the posterior predictive distribution is negative binomial with parameters ($\beta+n, \alpha+\sum y_i$).

$$
P(\tilde{y}_B < \tilde{y}_A|y_A, y_B) \approx 0.692.
$$

The plot shows that probability of $\tilde{y}_B < \tilde{y}_A$ drastically decreases when n_0 is small (from 0.7 to 0.51 when n_0 is between 1 to 40). When n_0 is large, probability of $\tilde{y}_B < \tilde{y}_A$ decreases very slowly. Therefore, the probability of $\tilde{y}_B < \tilde{y}_A$ is sensitive to the prior when n_0 is small.

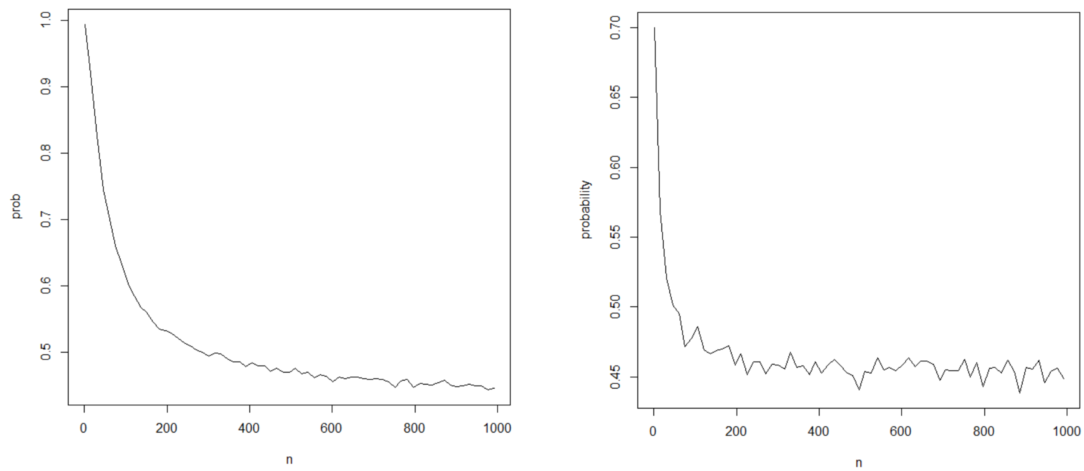

Figure 1: The left plot is $P\left(\theta_B < \theta_A | \mathbf{y}_A, \mathbf{y}_B\right)$ vs. n_0, and the right plot is $P\left(\tilde{y}_B < \tilde{y}_A | \, \mathbf{y}_A, \mathbf{y}_B\right)$ vs. n_0

4.8

a.

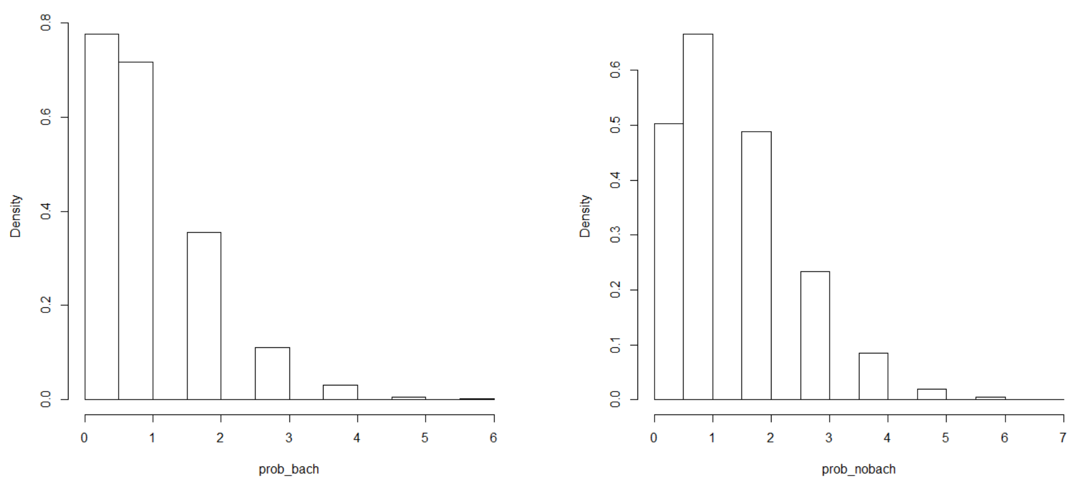

Figure 2: The plots are the Monte Carlo approximations to posterior predictive distributions.

b. 95% quantile-base posterior confidence interval for $\theta_B - \theta_A = (0.161, 0.737)$
   95% quantile-base posterior confidence interval for $\tilde{y}_B - \tilde{y}_A = (-2, 4)$

c. The distribution for the non-bachelor group is significantly different from Poisson with parameter $\hat{\theta} = 1.4$. Therefore, the Poisson distribution is not a good fit.

d. The empirical value for the non-bachelor group with zero and one child is (74, 49). It lies outside the Poisson model (located on the right lower corner). Therefore, the Poisson model is inadequate to describe this case.

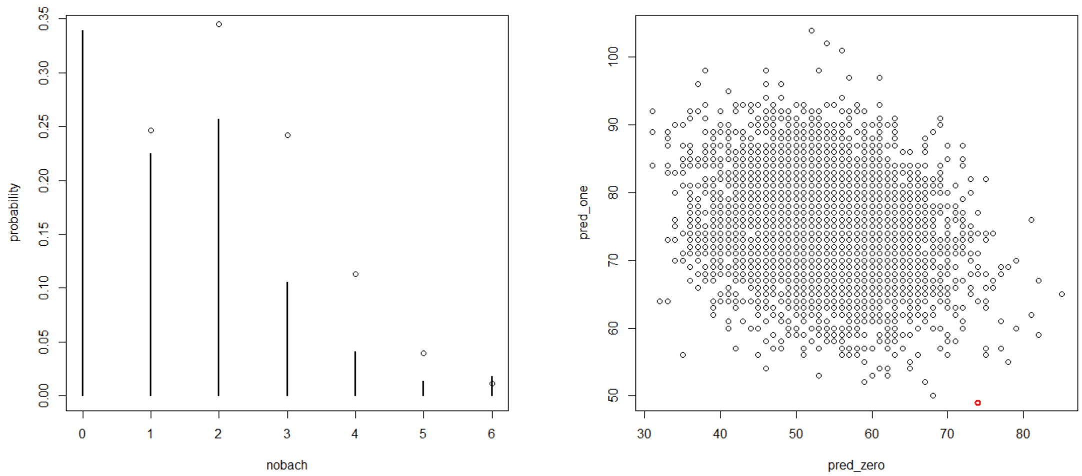

Figure 3: The left plot is the empirical distribution in group B and the Poisson distribution with mean $\hat{\theta} = 1.4$. (Solid lines are empirical distribution and dots are Poisson distribution.) The right plot corresponds to 5000 datasets drawn from the posterior predictive distribution. (The red dot is the observed datum.)