STAT3016/4116/7016:
Introduction to Bayesian Data Analysis

Tutorial 2

Dehua Tao

Question 1

$$
\begin{aligned}
p\left(\tilde{y}|Y\right) &= \int_0^\infty p\left(\tilde{y}|\theta, Y\right) p\left(\theta|Y\right) d\theta \\
&= \int p(\tilde{y}|\theta) p\left(\theta|Y\right) d\theta) \\
&= \int \mathrm{dpois}(\tilde{y}, \theta)\, \mathrm{dgamma}\left(\theta, \alpha + \sum_i y_i, \beta + n\right) d\theta \\
&= \int \left\{\frac{1}{\tilde{y}!}\theta^{\tilde{y}} e^{-\theta}\right\} \left\{\frac{(\beta + n)^{\alpha + \sum_i y_i}}{\Gamma\left(\alpha + \sum_i y_i\right)}\right\} \theta^{\alpha + \sum_i y_i - 1} e^{-(\beta + n)\theta} d\theta \\
&= \frac{(\beta + n)^{\alpha + \sum_i y_i}}{\Gamma(\tilde{y} + 1)\Gamma\left(\alpha + \sum_i y_i\right)} \int_0^\infty \theta^{\alpha + \sum_i y_i + \tilde{y} - 1} e^{-(\beta + n + 1)\theta} d\theta
\end{aligned}
$$

Then solve the integral above.

Question 1

$$\int_0^\infty \theta^{\alpha+\sum_i y_i+\tilde{y}-1} e^{-(\beta+n+1)\theta} d\theta = \frac{\Gamma\left(\alpha+\sum_i y_i+\tilde{y}\right)}{(\beta+n+1)^{\alpha+\sum_i y_i+\tilde{y}}}$$

For this integral, we can directly using the property of Gamma distribution.
Alternatively, we can use the Gamma function:
$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$, letting $\phi = (\beta+n+1)\theta$, the integral becomes to:

$$\frac{1}{(\beta+n+1)^{\alpha+\sum_i y_i+\tilde{y}}} \int_0^\infty \phi^{\alpha+\sum_i y_i+\tilde{y}-1} e^{-\phi} d\phi$$

Question 1

Then we have:

$$p\left(\tilde{y}|Y\right) = \frac{\Gamma\left(\alpha + \sum_i y_i + \tilde{y}\right)}{\Gamma(\tilde{y}+1)\Gamma\left(\alpha + \sum_i y_i\right)} \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum_i y_i} \left(\frac{1}{\beta+n+1}\right)^{\tilde{y}}$$

$$= \left(\begin{array}{c} \alpha + \sum_i y_i + \tilde{y} - 1 \\ \tilde{y} \end{array}\right) \left(\frac{\beta+n}{\beta+n+1}\right)^{\alpha+\sum_i y_i} \left(\frac{1}{\beta+n+1}\right)^{\tilde{y}}$$

In the form of:

$$p(x=k) = \left(\begin{array}{c} k+r-1 \\ k \end{array}\right) (1-p)^r p^k$$

with parameters (r, p)

Question 1

One expression: This is a negative binomial density with parameters $\left(\alpha + \sum_i y_i, \frac{1}{\beta+n+1}\right)$, where the probability of fail on a trial is p=$\frac{1}{\beta+n+1}$, and calculate the probability of $\tilde{y}$ successes before $\alpha + \sum_i y_i$ fails.
Mean of the negative binomial distribution: $\frac{pr}{1-p}$, variance: $\frac{pr}{(1-p)^2}$.
Another expression: This is a negative binomial density with parameters $\left(\alpha + \sum_i y_i, \frac{\beta+n}{\beta+n+1}\right)$, where the probability of success on a trial is p'=$\frac{\beta+n}{\beta+n+1}$, and calculate the probability of $\tilde{y}$ fails before $\alpha + \sum_i y_i$ successes.
Mean of the negative binomial distribution: $\frac{(1-p')r}{p'}$, variance: $\frac{(1-p')r}{p'^2}$.

Question 1

Then we have:

$$\text{Var}\left[\tilde{Y}|Y\right] = \frac{\alpha + \sum_i y_i}{(\beta + n)^2}(\beta + n + 1) = E\left[\theta|Y\right] \times \frac{(\beta + n + 1)}{\beta + n}$$

Question 1

Interpretation:

$$\begin{aligned}
\text{Var}\left[\tilde{Y}|Y\right] &= \frac{\alpha + \sum_i y_i}{(\beta+n)^2}(\beta+n+1) \\
&= Var\left[\theta|Y\right] \times (\beta+n+1) \\
&= E\left[\theta|Y\right] \times \frac{\beta+n+1}{\beta+n} \\
&= E\left[\theta|Y\right] + Var\left[\theta|Y\right]
\end{aligned}$$

We can divide the variability into two parts, the first part is the uncertainty about the population, and another variability is the variability in sampling from the population. For a large n, $\frac{\beta+n+1}{\beta+n}$ is closer to 1 and the uncertainty about $\tilde{Y}$ is primarily from the sampling variability, for a small n the uncertainty about $\tilde{Y}$ also includes the uncertainty in θ, so that the total variability should be larger than the sampling variability.

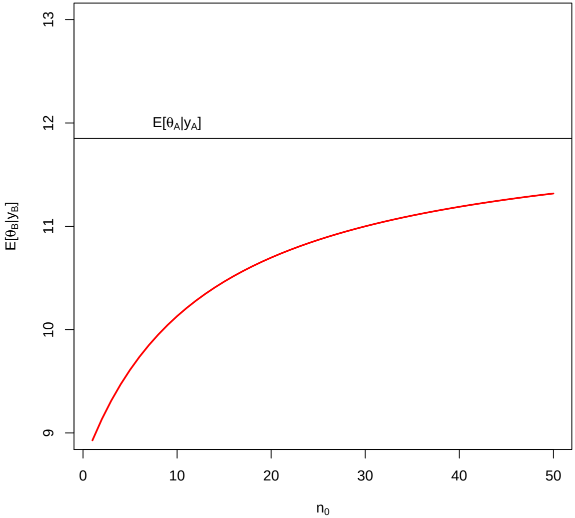
Question 2(a)

Assuming a Poisson sampling distribution for each group, the prior is Gamma, the likelihood is Poisson.

$$\theta_A \,|\mathbf{y}_\mathrm{A} \sim \text{Gamma}(237, 20); \theta_B |\, \mathrm{y}_\mathrm{B} \sim \text{Gamma}(125, 14)$$
$$E\left[\theta_A|\mathbf{y}_\mathrm{A}\right] = 237/20 = 11.9; E\left[\theta_B|\mathbf{y}_\mathrm{B}\right] = 125/14 = 8.9$$
$$\text{Var}\left[\theta_A|\mathrm{y}_\mathrm{A}\right] = 237/20^2 = 0.59; \text{Var}\left[\theta_B|\mathrm{y}_\mathrm{B}\right] = 125/14^2 = 0.64$$

95% quantile-based intervals for θ_A and θ_B are (10.4,13.4) and (7.4,10.6) respectively.
In summary, we conclude that expected counts from type B mice are lower given the data.

Question 2(b)



Question 2(b)

$$E\left[\theta_B | y_B\right] = \frac{12n_0 + 113}{n_0 + 13}$$

Shows an increasing trend which means n_0 needs to be greater than 50 in order for the posterior expectation of θ_B to be close to that of θ_A which means we would need to assume that tumor counts from Type B mice are well studied (that is, more prior data).
We want $E(\theta_B | y_B)$ close to $E(\theta_A | y_A)$ is because even if tumor count rates for type B mice are unknown, but type B mice are related to type A mice.

Question 2(c)(i)

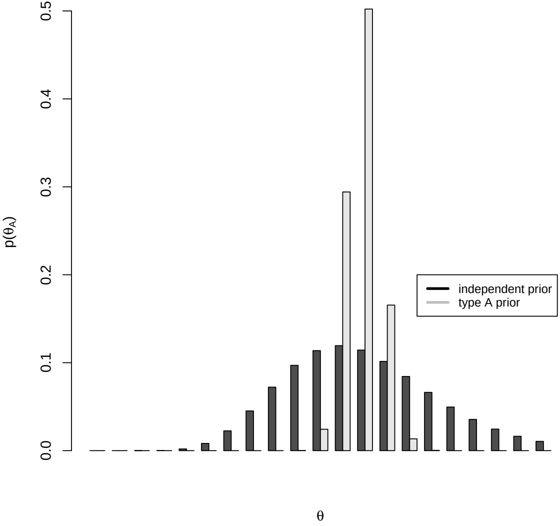
$\sum y_b = 113$. Then:

$$\tilde{y}_B \,\Big|\, y_b \sim \text{Neg Bin}\left(12 + 113, \tfrac{1}{1+1+13}\right). \quad E\left[\tilde{y}_B | y_b\right] = \frac{125 \times \frac{1}{15}}{\frac{14}{15}} = \frac{125}{14} = 8.9$$

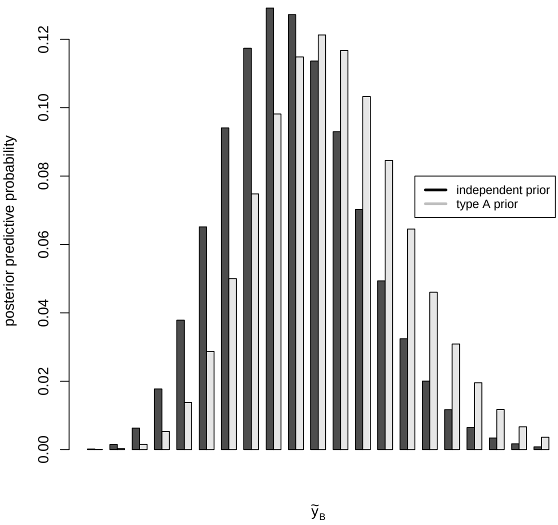
Question 2(c)(ii)

$p(\theta_B) = p(\theta_A|y_A)$, then $\theta_B \sim Gamma(237, 20)$.
Then
$\tilde{y}_B \Big| y_b \sim \text{Neg Bin}(237 + 113, \frac{1}{1+20+13}).E\left[\tilde{y}_B|y_b\right] = \frac{350}{33} = \frac{350\times\frac{1}{34}}{\frac{33}{34}} = 10.6$
The predicted count is higher by using the prior in (ii).
If $\theta|y \sim Gamma(\alpha^*, \beta^*)$, then $\tilde{y}|y \sim NegBin(\alpha^*, \frac{1}{1+\beta^*})$

Question 2(c)(ii)



Question 2(c)(ii)



Question 2(c)(ii)

Type A prior is more informative compared to independent prior. The location of the predictive distribution shifts to the right, which means the posterior predictive mean using the Type A prior is higher than that using independent prior. In other words, showing a more related result.

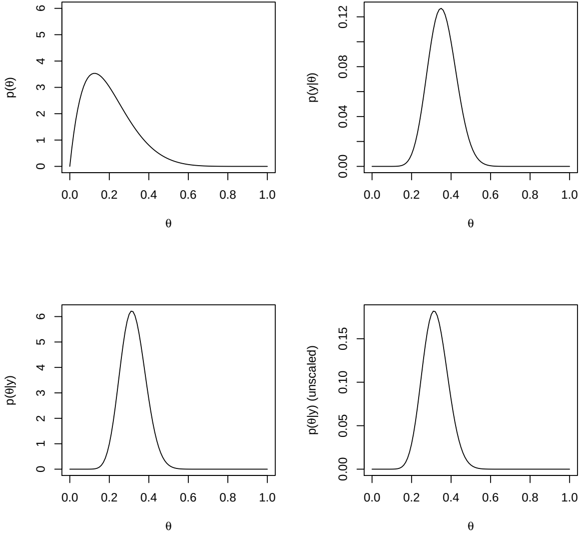
Question 2(d)

We are told that type B mice are related to type A mice. So it is not appropriate to assume independent prior between θ_A and θ_B. Compared to applying independent prior between two parameters, it makes more sense to let the posterior distribution of θ_A as the prior of θ_B to show a more related result between A and B.

Question 3(a)

- posterior distribution: $\theta|y \sim \text{Beta}(2+15, 8+43-15)$.
- posterior mean: $E[\theta|y] = \frac{17}{17+36} = 0.32$.
- posterior mode: $\frac{17-1}{17+36-2} = 0.31$.
- posterior standard deviation: $SD[\theta|y] = \sqrt{\frac{17\times36}{(17+36)^2(17+36+1)}} = 0.0635$.
- prior mode: $\frac{2-1}{2+8-2} = 0.125$.

Question 3(a)



p(θ)
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
θ
p(y|θ)
0.00
0.04
0.08
0.12
0.0
0.2
0.4
0.6
0.8
1.0
θ
p(θ|y)
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
θ
p(θ|y) (unscaled)
0.00
0.05
0.10
0.15
0.0
0.2
0.4
0.6
0.8
1.0
θ

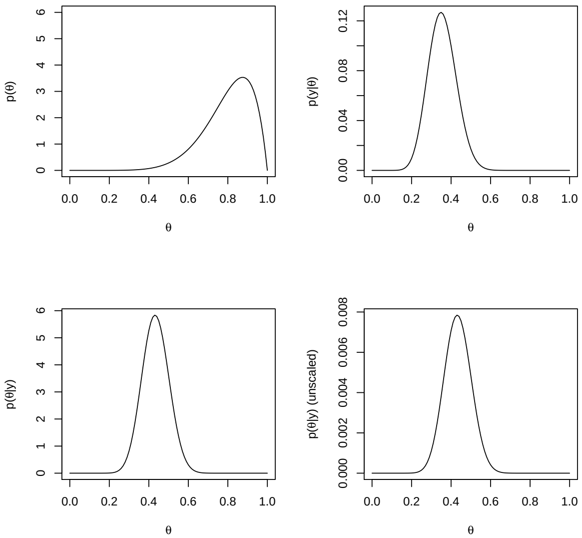
Question 3(a)

```
> qbeta(c(0.025,0.975),a+y,b+n-y)
[1] 0.2032978 0.4510240
> quantile(rbeta(10000,a+y,b+n-y),c(0.025,0.975))
  2.5%     97.5%
  0.2014589 0.4513895
```

Question 3(b)

- posterior distribution: $\theta|y \sim \text{Beta}(8 + 15, 2 + 43 - 15)$.
- posterior mean: $E[\theta|y] = \frac{23}{23+30} = 0.43$.
- posterior mode: $\frac{23-1}{23+30-2} = 0.43$.
- posterior standard deviation: $SD[\theta|y] = \sqrt{\frac{23\times 30}{(23+30)^2(23+30+1)}} = 0.0674$.
- prior mode: $\frac{8-1}{8+2-2} = 0.875$.

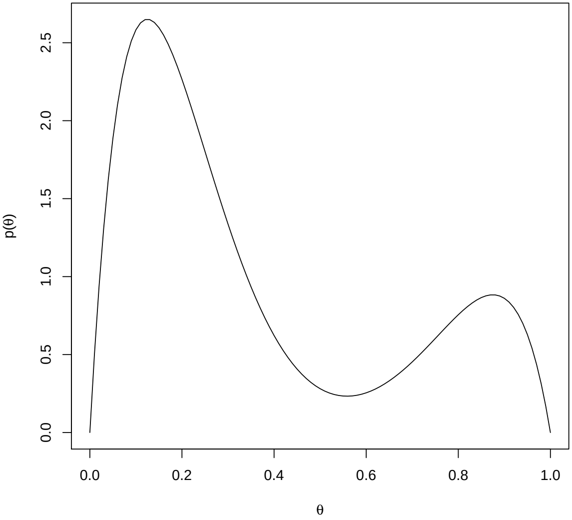
Question 3(b)


p(θ)
p(y|θ)
p(θ|y)
p(θ|y) (unscaled)
θ
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
0.00
0.04
0.08
0.12
0.000
0.002
0.004
0.006
0.008

Question 3(b)

```
> qbeta(c(0.025,0.975),a+y,b+n-y)
[1] 0.3046956 0.5679528
> quantile(rbeta(10000,a+y,b+n-y),c(0.025,0.975))
     2.5%     97.5%
0.3054893 0.5666899
```

Question 3(c)



Question 3(c)

This is a bimodal distribution, the prior distribution has more than one peaks. The prior could be based on reviews of recidivism rates in previous years or other locations, and the prior information could indicate the presence of two sub-populations, one with mean recidivism rate 0.2 and the other with mean recidivism rate 0.8.

Question 3(c)

```
> theta<-seq(0,1,0.01)
> p.theta<-0.75*dbeta(theta,a1,b1)+0.25*dbeta(theta,a2,b2)
> 0.25*gamma(10)/(gamma(2)*gamma(8))*(3*theta*(1-theta)^7+
theta^7*(1-theta))
> plot(theta,p.theta,type="l")
```

Question 3(d)(i)

$$
\begin{aligned}
P(\theta|y) &\propto p(y|\theta)p(\theta) \\
&= p(y|\theta)(0.75 \text{ beta } (2,8) + 0.25 \text{ beta } (8,2)) \\
&= \theta^y(1-\theta)^{n-y}(0.75\,\text{beta}(2,8) + 0.25\,\text{beta}(8,2)) \\
&= \theta^y(1-\theta)^{n-y}[0.75\frac{\Gamma(2+8)}{\Gamma(2)\Gamma(8)}\theta^{2-1}(1-\theta)^{8-1} \\
&+ 0.25\frac{\Gamma(8+2)}{\Gamma(8)\Gamma(2)}\theta^{8-1}(1-\theta)^{2-1}] \\
&= 0.75\frac{\Gamma(2+8)}{\Gamma(2)\Gamma(8)}\theta^{y+2-1}(1-\theta)^{n-y+8-1} \\
&+ 0.25\frac{\Gamma(8+2)}{\Gamma(8)\Gamma(2)}\theta^{y+8-1}(1-\theta)^{n-y+2-1}
\end{aligned}
$$

Question 3(d)(ii)

$$P(\theta|y) \propto w_1 \times \text{Beta}(y+2, n-y+8) + w_2 \times \text{Beta}(y+8, n-y+2)$$

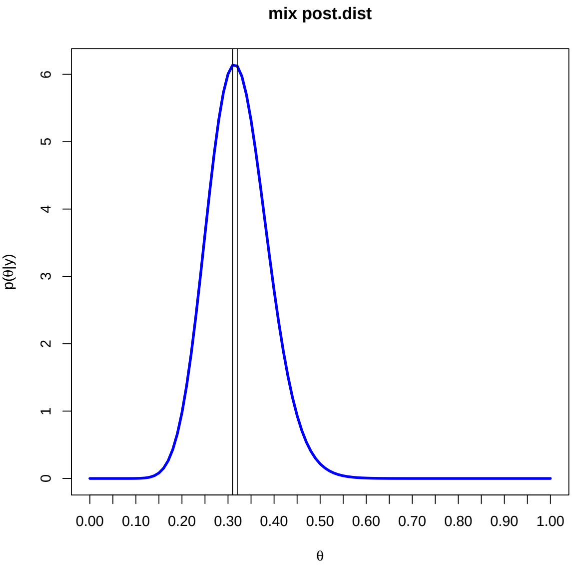
where

$$w_1 = 0.75 \frac{\Gamma(2+8)}{\Gamma(2)\Gamma(8)} \frac{\Gamma(y+2)\Gamma(n-y+8)}{\Gamma(2+n+8)}$$
$$w_2 = 0.25 \frac{\Gamma(8+2)}{\Gamma(8)\Gamma(2)} \frac{\Gamma(y+8)\Gamma(n-y+2)}{\Gamma(8+n+2)}$$

then normalize:

$$p_1^* = w_1 / (w_1 + w_2)\,; p_2^* = 1 - p_1^*$$
$$P(\theta|y) = p_1^* \,\text{Beta}(y+2, n-y+8) + p_2^* \,\text{Beta}(y+8, n-y+2)$$

Question 3(d)(iii)



Question 3(d)(iii)

The posterior mode is around 0.315.
Similar to the posterior mode in (a), which makes sense given the higher posterior weighting on the prior in (a) in the mixture distribution, with $p_1^* = 0.9849087, p_2^* = 0.01509134$).

Question 3(e)

General form:

$$w_j = p_j \frac{\Gamma(a_j + b_j)}{\Gamma(a_j)\Gamma(b_j)} \frac{\Gamma(y + a_j)\Gamma(n - y + b_j)}{\Gamma(a_j + n + b_j)}$$

$$p_j^* = w_j / \sum_j w_j$$

$$P(\theta|y) = \sum_j p_j^* P_j(\theta|y)$$