Introduction to Bayesian Data Analysis
Tutorial 2 Solutions

(1) Assume a conjugate prior $p(\theta) \propto \theta^{\alpha} e^{-\beta\theta}$

$$\begin{aligned}
p(\tilde{y}|y_1, ..., y_n) &= \int_0^\infty p(\tilde{y}|\theta, y_1, ..., y_n) p(\theta|y_1, ..., y_n) d\theta \\
&= \int p(\tilde{y}|\theta) p(\theta|y_1, ..., y_n) d\theta) \\
&= \int \text{dpois}(\tilde{y}, \theta) \text{dgamma}(\theta, \alpha + \sum_i y_i, \beta + n) d\theta \\
&= \int \left\{ \frac{1}{\tilde{y}!} \theta^{\tilde{y}} e^{-\theta} \right\} \left\{ \frac{(\beta+n)^{\alpha+\sum_i y_i}}{\Gamma(\alpha + \sum_i y_i)} \right\} \theta^{\alpha+\sum_i y_i - 1} e^{-(\beta+n)\theta} d\theta \\
&= \frac{(\beta+n)^{\alpha+\sum_i y_i}}{\Gamma(\tilde{y}+1)\Gamma(\alpha + \sum_i y_i)} \int_0^\infty \theta^{\alpha+\sum_i y_i + \tilde{y} - 1} e^{-(\beta+n+1)\theta} d\theta
\end{aligned}$$

To simplify the above integral, note that

$$1 = \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} d\theta$$

for any values $\alpha, \beta > 0$. So $\int_0^\infty \theta^{\alpha-1} e^{-\beta\theta} d\theta = \frac{\Gamma(\alpha)}{\beta^\alpha}$. Substitute in $\alpha + \sum_i y_i + \tilde{y}$ instead of α and $\beta + n + 1$ instead of β

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

Then we will have:

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

which is a negative binomial density with parameters $(\alpha + \sum_i y_i, \frac{\beta+n}{\beta+n+1})$ $(\tilde{y} \in \{0, 1, 2, ...\})$ where the probability of success on a trial is $\frac{\beta+n}{\beta+n+1}$, and we want to calculate the probability of $\tilde{y}$ failures before $\alpha + \sum_i y_i - 1$ successes. (In other words, $\alpha + \sum_i y_i - 1$ successes and $\tilde{y} - 1$ failures in $\alpha + \sum_i y_i + \tilde{y} - 1$ trials and success on the $\alpha + \sum_i y_i + \tilde{y}$ trial).

Now for $X \sim NegBin(r, p)$ (for modelling the number of failures X before r successes and the probability of success on a trial is p), we have $E[X] = \frac{r(1-p)}{p}$ and $Var[X] = \frac{r(1-p)}{p^2}$.

In our case $r = \alpha + \sum_i y_i$ and $p = \frac{\beta+n}{\beta+n+1}$ (and $1 - p = \frac{1}{\beta+n+1}$).

So we have:

$$\begin{aligned}
Var[\tilde{Y}|y_1,...,y_n] &= (\alpha + \sum_i y_i) \times \frac{1}{\beta+n+1} \times \left(\frac{\beta+n+1}{\beta+n}\right)^2 \\
&= \frac{(\alpha + \sum_i y_i)}{(\beta+n)^2} \times (\beta+n+1) \\
&= Var(\theta|y_1,...,y_n) \times (\beta+n+1) \\
&= E[\theta|y_1,...,y_n) \times \frac{\beta+n+1}{\beta+n}
\end{aligned}$$

where the third and fourth lines from the posterior distribution $\theta|y_1,...,y_n \sim Gamma(\alpha + \sum_i y_i, \beta+n)$.

Now note that

$$\begin{aligned}
Var[\tilde{Y}|y_1,...,y_n] &= \frac{(\alpha + \sum_i y_i)}{(\beta+n)^2} \times (\beta+n+1) \\
&= \frac{(\alpha + \sum_i y_i)}{(\beta+n)} + \frac{(\alpha + \sum_i y_i)}{(\beta+n)^2} \\
&= E[\theta|y_1,....,y_n] + Var[\theta|y_1,....,y_n]
\end{aligned}$$

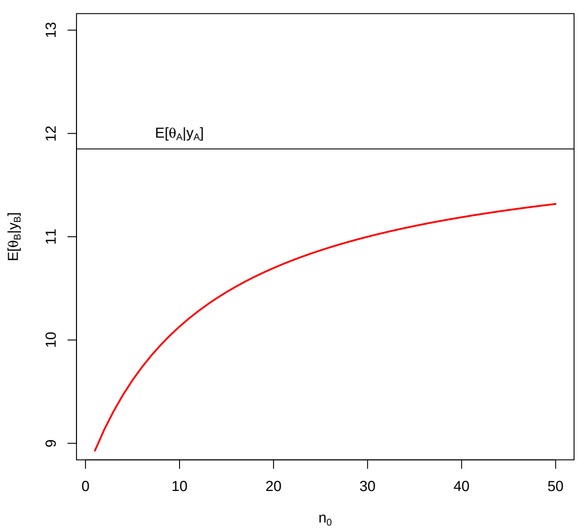
Interpretation: Uncertainty in a new prediction comes from uncertainty about the population and the individual variability in sampling a single unit from the population. So we need to adjust upwards the posterior variance $Var(\theta|y_1,...,y_n)$ by a factor of $(\beta+n+1)$ when deriving the posterior predictive variance. Alternatively when n is large, $Var[\theta|y_1,....,y_n]$ is close to zero and the predictive variability stems from $E[\theta|y_1,....,y_n]$, which represents the sampling variability, noting that θ is the sample variance of Y. For small n, uncertainty in $\tilde{Y}$ also includes uncertainty in θ.

(2) (a) $\sum_i y_{A,i} = 117$; $\sum_i y_{b,i} = 113$.
So $\theta_A|\mathbf{y_A} \sim \text{Gamma}(120+117, 10+10)$; $\theta_B|\mathbf{y_B} \sim \text{Gamma}(12+113, 1+13)$
$E[\theta_A|\mathbf{y_A}] = 237/20 = 11.9$; $E[\theta_B|\mathbf{y_B}] = 125/14 = 8.9$
$\text{Var}[\theta_A|\mathbf{y_A}] = 237/20^2 = 0.59$; $\text{Var}[\theta_B|\mathbf{y_B}] = 125/14^2 = 0.64$

```
> qgamma(c(0.025,0.975),a1+sum(ya),b1+length(ya))
[1] 10.38924 13.40545
> qgamma(c(0.025,0.975),a2+sum(yb),b2+length(yb))
[1]  7.432064 10.560308
```

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.

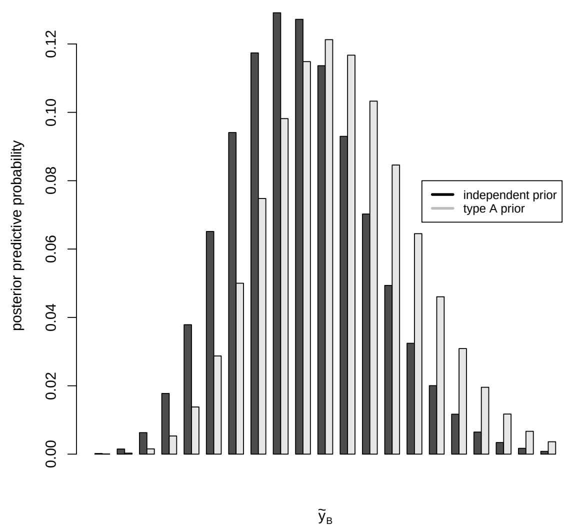
(b) n_0 needs to be greater than 50 in order for the posterior expectation of θ_B to be close to that of θ_A which means we would need to assume that tumor counts from Type B mice are well studied (that is, more prior data).

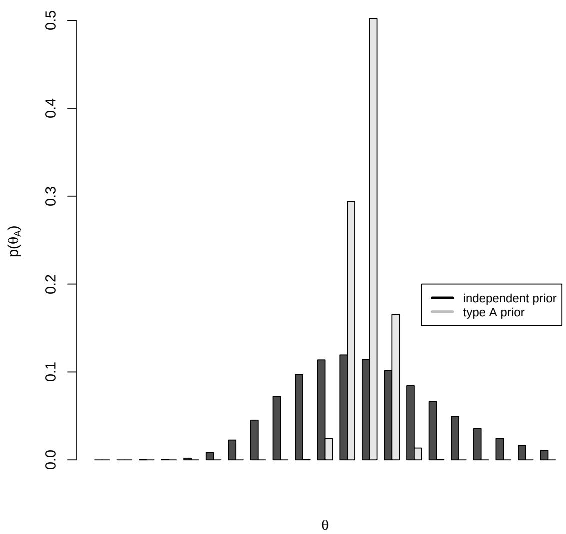


(c) (i) $\sum y_b = 113$. Then $\tilde{y}_B|y_b \sim \text{NegBin}(12+113, \frac{1+13}{1+13+1})$. $E[\tilde{y}_B|y_b] = \frac{125}{14} = 8.9$

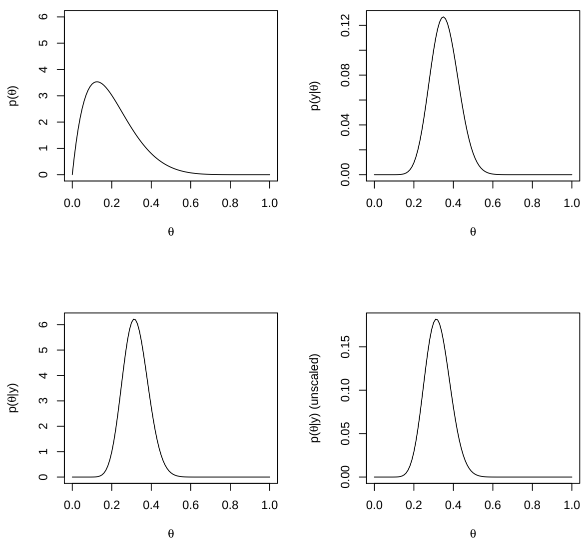
(ii) $\sum y_a = 117$. $p(\theta_b) \propto p(\theta_a|y_a)$; so $\theta_b \sim Gamma(120+117, 10+10)$. Then $\tilde{y}_B|y_b \sim \text{NegBin}(237+113, \frac{20+13}{20+13+1})$. $E[\tilde{y}_B|y_b] = \frac{350}{33} = 10.6$
We expect a higher predicted count if we assume the data from mice A form a prior distribution for the posterior of θ_B (as shown by the shift in the location of the posterior predictive distribution (see plot below).

(d) We are told that type B mice are related to type A mice. We are not told the details of the relationship but this is enough information to say that it does not make sense to assume prior independence between θ_a and θ_b. In particular, given that type A mice have been well studied, it makes more sense to use tumor rate data from type A, as a prior guess on the tumor rates of type B mice. The type A prior is a more informative prior.





(3) (a) $\theta|y \sim \text{Beta}(2+15, 8+43-15)$. $E[\theta|y] = \frac{17}{17+36} = 0.32$. Posterior mode $= \frac{17-1}{17+36-2} = 0.31$. $SD[\theta|y] = \sqrt{\frac{17\times 36}{(17+36)^2(17+36+1)}} = 0.0635$. A 95% posterior interval is (0.20,0.45)



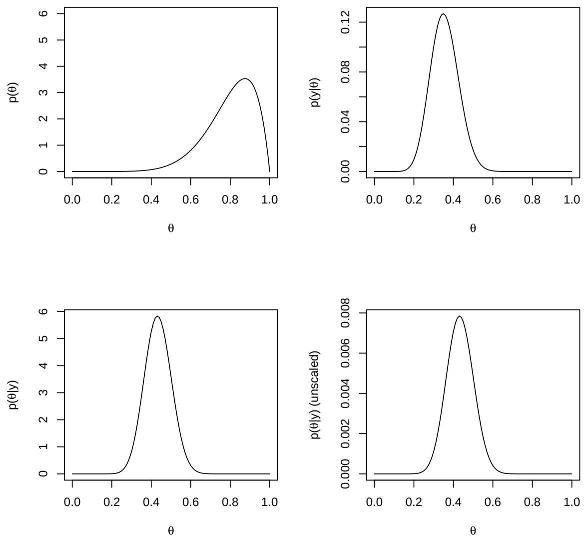

```
> qbeta(0.025,a+y,b+n-y)
[1] 0.2032978
> qbeta(0.975,a+y,b+n-y)
[1] 0.451024
```

(b) $\theta|y \sim \text{Beta}(8+15, 2+43-15)$. $E[\theta|y] = \frac{23}{23+30} = 0.43$. Posterior mode $= \frac{23-1}{23+30-2} = 0.41$. $SD[\theta|y] = \sqrt{\frac{23\times 30}{(23+30)^2(23+30+1)}} = 0.0674$. A 95% posterior interval is (0.30,0.57).

We can see the posterior summaries in (b) are higher in line with a higher prior belief of recidivism, but the uncertainty in our prior belief is similar.

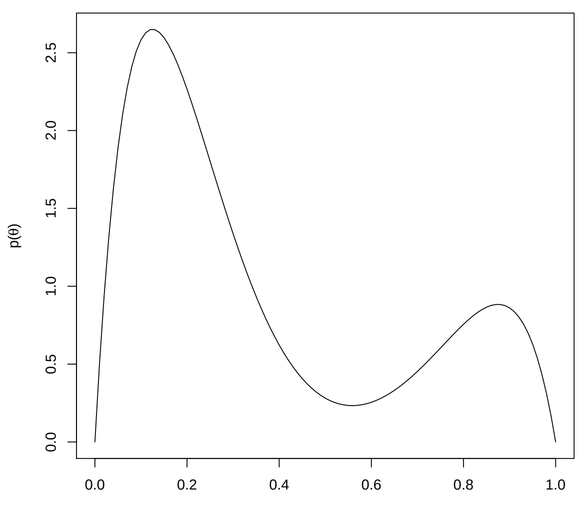
```
> qbeta(0.025,a+y,b+n-y)
[1] 0.3046956
> qbeta(0.975,a+y,b+n-y)
[1] 0.5679528
```

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




```
   theta^7*(1-theta))
plot(theta,p.theta,type="l")
```

This prior has two 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 subpopulations, one with mean recidivism rate 0.2 and the other with mean recidivism rate 0.8.



(d)

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

$$\begin{aligned}
P(\theta|y) \propto\ & 0.75\frac{\Gamma(2+8)}{\Gamma(2)\Gamma(8)}\frac{\Gamma(y+2)\Gamma(n-y+8)}{\Gamma(2+n+8)}\frac{\Gamma(2+n+8)}{\Gamma(y+2)\Gamma(n-y+8)}\theta^{y+2-1}(1-\theta)^{n-y+8-1} \\
&+ 0.25\frac{\Gamma(8+2)}{\Gamma(8)\Gamma(2)}\frac{\Gamma(y+8)\Gamma(n-y+2)}{\Gamma(8+n+2)}\frac{\Gamma(8+n+2)}{\Gamma(y+8)\Gamma(n-y+2)}\theta^{y+8-1}(1-\theta)^{n-y+2-1}
\end{aligned}$$

$$\begin{aligned}
P(\theta|y) \propto 0.75\frac{\Gamma(2+8)}{\Gamma(2)\Gamma(8)}\frac{\Gamma(y+2)\Gamma(n-y+8)}{\Gamma(2+n+8)} \times \text{Beta}(y+2, n-y+8) + \\
0.25\frac{\Gamma(8+2)}{\Gamma(8)\Gamma(2)}\frac{\Gamma(y+8)\Gamma(n-y+2)}{\Gamma(8+n+2)}\text{Beta}(y+8, n-y+2)
\end{aligned}$$

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

$$w_2 = .25\frac{\Gamma(8+2)}{\Gamma(8)\Gamma(2)}\frac{\Gamma(y+8)\Gamma(n-y+2)}{\Gamma(8+n+2)}$$

$$p_1^* = w_1/(w_1+w_2); p_2^* = 1 - p_1^*$$

$$P(\theta|y) = p_1\text{Beta}(y+2, n=y+8) + p_2\text{Beta}(y+8, n-y+2)$$

So a general formula is:

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

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

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

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 ($p_1^* = 0.98$).

```
> p1
[1] 0.9849087
> p2
[1] 0.01509134
```