Guanlin Zhang
Fall '18

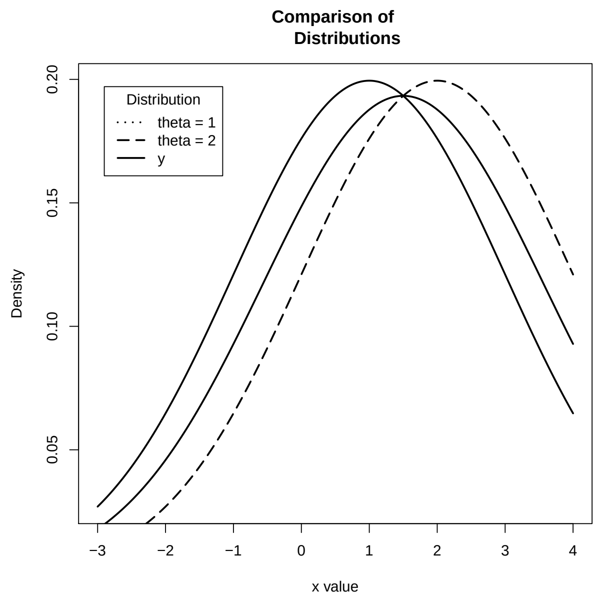
Chapter 1 Question 1 on page 27:

Solution 1. *part (a): For $\sigma = 2$, write the formula for the marginal probability density for y and sketch it.*

We have:

$$
\begin{aligned}
p(y) &= p(y, \theta = 1) + p(y, \theta = 2) \\
&= p(\theta = 1) \cdot p(y|\theta = 1) + p(\theta = 2) \cdot p(y|\theta = 2) \\
&= 0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 4}} \cdot \exp\Big\{ -\frac{1}{2 \cdot 4}(y-1)^2 \Big\} + 0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 4}} \cdot \exp\Big\{ -\frac{1}{2 \cdot 4}(y-2)^2 \Big\} \\
&= \frac{1}{4 \cdot \sqrt{2\pi}} \cdot \exp\Big\{ -\frac{1}{8}(y-1)^2 \Big\} + \frac{1}{4 \cdot \sqrt{2\pi}} \cdot \exp\Big\{ -\frac{1}{8}(y-2)^2 \Big\}
\end{aligned}
$$

```r
x <- seq(-3, 4, length = 700)
hx1 <- dnorm(x, mean = 1, sd = 2)
hx2 <- dnorm(x, mean = 2, sd = 2)
labels <- c("theta = 1", "theta = 2", "y")

plot(x, hx1, type="l", lty = 1, lwd = 2, xlab = "x value",
     ylab = "Density", main = "Comparison of
     Distributions")
lines(x, hx2, lty = 5, lwd = 2)
lines(x, 0.5*hx1 + 0.5*hx2, lty = 1, lwd = 2)
legend("topleft", inset = .05, title = "Distribution",
       labels, lwd = 2, lty = c(3, 5, 1))
```



part (b): What is $Pr(\theta = 1|y = 1)$, again supposing $\sigma = 2$?

We have the posteior distribution as:

$$\begin{aligned}
Pr(\theta = 1|y) &= \frac{Pr(y|\theta = 1) \cdot Pr(\theta = 1)}{p(y)} \\
&= \frac{\cancel{0.5} \cdot \frac{1}{\sqrt{2\pi \cdot 4}} \cdot \exp\Big\{ - \frac{1}{2 \cdot 4} \cdot (y-1)^2 \Big\}}{\cancel{0.5} \cdot \frac{1}{\sqrt{2\pi \cdot 4}} \cdot \exp\Big\{ - \frac{1}{2 \cdot 4} \cdot (y-1)^2 \Big\} + \cancel{0.5} \cdot \frac{1}{\sqrt{2\pi \cdot 4}} \cdot \exp\Big\{ - \frac{1}{2 \cdot 4} \cdot (y-2)^2 \Big\}} \\
&= \frac{\exp\Big\{ - \frac{1}{8}(y-1)^2 \Big\}}{\exp\Big\{ - \frac{1}{8}(y-1)^2 \Big\} + \exp\Big\{ - \frac{1}{8}(y-2)^2 \Big\}}
\end{aligned}$$

which gives us

$$Pr(\theta = 1|y = 1) = \frac{\exp(0)}{\exp(0) + \exp(-\frac{1}{8})} = \frac{1}{1 + \exp(-\frac{1}{8})}$$

part (c): describe how the posteior density of θ changes in shape as σ is increased and as it is decreased.

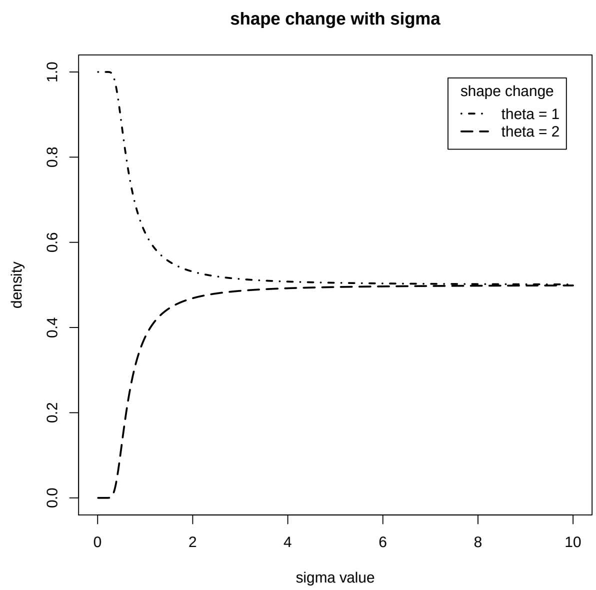
Identical to the computation as in (b), if we do not specify a particular value of σ but instead keep it as a variable, we get the posteiro distribution as:

$$\begin{aligned}
Pr(\theta = 1|y) &= \frac{\exp\Big\{ - \frac{1}{2\sigma^2}(y-1)^2 \Big\}}{\exp\Big\{ - \frac{1}{2\sigma^2}(y-1)^2 \Big\} + \exp\Big\{ - \frac{1}{2\sigma^2}(y-2)^2 \Big\}} \\
Pr(\theta = 2|y) &= \frac{\exp\Big\{ - \frac{1}{2\sigma^2}(y-2)^2 \Big\}}{\exp\Big\{ - \frac{1}{2\sigma^2}(y-1)^2 \Big\} + \exp\Big\{ - \frac{1}{2\sigma^2}(y-2)^2 \Big\}}
\end{aligned}$$

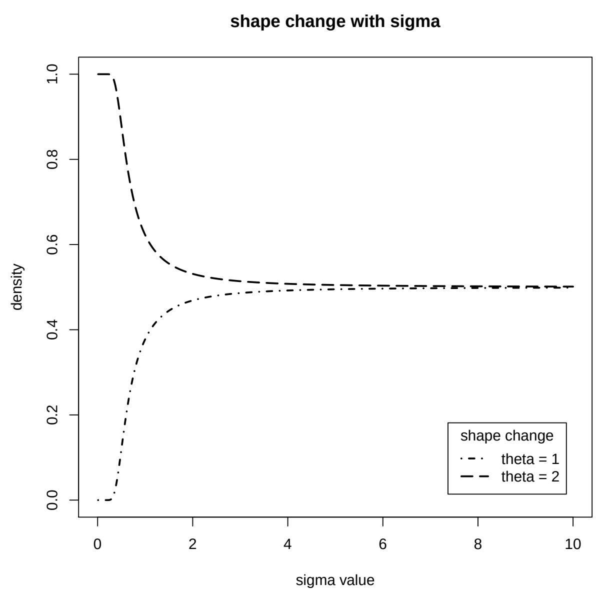
When $y < 1.5$, we have $Pr(\theta = 1|y)$ always larger than $Pr(\theta = 2|y)$, but as σ^2 increase the gap between these two get smaller and smaller and eventually both probability masses converge to very close to 0.5 (remember that these two always sum up to 1.) The following code and plot take $y = 1$ as example, and when $y < 1$ or between 1 and 1.5 the trend is very similar:

```
#part c)
sigma <-seq(0, 10, length = 1000)

#specify y value
y <- 1
post1 <- exp(-1/(2*sigma^2)*(y - 1)^2)/(exp(-1/(2*sigma^2)*(y - 1)^2)
                                        + exp(-1/(2*sigma^2)*(y - 2)^2))
post2 <- exp(-1/(2*sigma^2)*(y - 2)^2)/(exp(-1/(2*sigma^2)*(y - 1)^2)
                                        + exp(-1/(2*sigma^2)*(y - 2)^2))
labels <- c("theta = 1", "theta = 2")

plot(sigma, post1, type="l", lty = 4, lwd = 2, xlab = "sigma value",
```

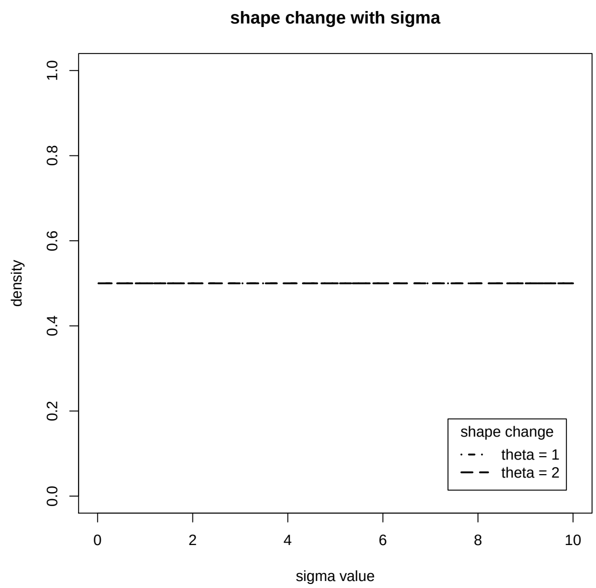
```
     ylab = "density", ylim = c(0, 1), main = "shape change with sigma")
lines(sigma, post2, lty = 5, lwd = 2)
legend("topright", inset = .05, title = "shape change",
       labels, lwd = 2, lty = c(4, 5))
```



When $y > 1.5$, it becomes the opposite as the case of $y < 1.5$: $Pr(\theta = 1|y)$ is always smaller than $Pr(\theta = 2|y)$, but the gap converges to 0 as σ^2 increase and both get really close to 0.5 when σ^2 is large. The following plot illustrates the case when $y = 2$.



When $y = 1.5$ exactly, then $Pr(\theta = 1|y) = Pr(theta = 2|y) = 0.5$ always, no matter which value σ^2 takes, as is shown in the following plot:



Chapter 2 Question 1 on page 57:

Solution 2. *From the question we know that $\theta \sim Beta(\alpha, \beta)$ with $\alpha = \beta = 4$. We also know that $x \sim Binomial(n, \theta)$ where x represents the number of heads and $n = 10$. The question gives us information that $x < 3$, namely, $x = 0, 1$ or 2.*

So the postrior distribution of θ given $x < 3$ is then:

$$\begin{aligned} p(\theta|x<3) &= \frac{p(\theta)\cdot Pr(x<3|\theta)}{\int_0^1 Pr(x<3|\theta)p(\theta)d\theta} = \frac{\cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}\theta^{\alpha-1}(1-\theta)^{\beta-1}\Big(\sum_{i=0}^{2}\binom{n}{i}\theta^i(1-\theta)^{n-i}\Big)}{\cancel{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}\int_0^1\Big(\sum_{i=0}^{2}\binom{n}{i}\theta^i(1-\theta)^{n-i}\Big)\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta} \\ &= \frac{\theta^3(1-\theta)^3\Big[(1-\theta)^{10}+10\cdot\theta(1-\theta)^9+45\cdot\theta^2(1-\theta)^8\Big]}{\int_0^1\theta^3(1-\theta)^3\Big[(1-\theta)^{10}+10\cdot\theta(1-\theta)^9+45\cdot\theta^2(1-\theta)^8\Big]d\theta} \end{aligned}$$

We would like to give a name to the proportionality constant:

$$Constant = \int_0^1 \theta^3(1-\theta)^3\Big[(1-\theta)^{10}+10\cdot\theta(1-\theta)^9+45\cdot\theta^2(1-\theta)^8\Big]d\theta$$

The following algebra simplifies the expression of the "Constant":

$$\begin{aligned}
Constant &= \int_0^1 \theta^3(1-\theta)^{13} + 10\theta^4(1-\theta)^{12} + 45\theta^5(1-\theta)^{11} d\theta \\
&= \frac{\Gamma(4)\Gamma(14)}{\Gamma(18)} \int_0^1 \frac{\Gamma(18)}{\Gamma(4)\Gamma(14)}\theta^3(1-\theta)^{13} d\theta + 10 \cdot \frac{\Gamma(5)\Gamma(13)}{\Gamma(18)} \int_0^1 \frac{\Gamma(18)}{\Gamma(5)\Gamma(13)}\theta^4(1-\theta)^{12} d\theta \\
&\quad + 45 \cdot \frac{\Gamma(6)\Gamma(12)}{\Gamma(18)} \int_0^1 \frac{\Gamma(18)}{\Gamma(6)\Gamma(12)}\theta^5(1-\theta)^{11} d\theta \\
&= \frac{\Gamma(4)\Gamma(14)}{\Gamma(18)} \cdot 1 + 10 \cdot \frac{\Gamma(5)\Gamma(13)}{\Gamma(18)} \cdot 1 + 45 \cdot \frac{\Gamma(6)\Gamma(12)}{\Gamma(18)} \cdot 1 \\
&= B(4,14) + 10B(5,13) + 45B(6,12)
\end{aligned}$$

then the posterior distribution can be expressed as:

$$p(\theta|x<3) = \frac{1}{Constant}\theta^3(1-\theta)^{13} + \frac{10}{Constant}\theta^4(1-\theta)^{12} + \frac{45}{Constant}\theta^5(1-\theta)^{11}$$

It can be further normalized into the form as:

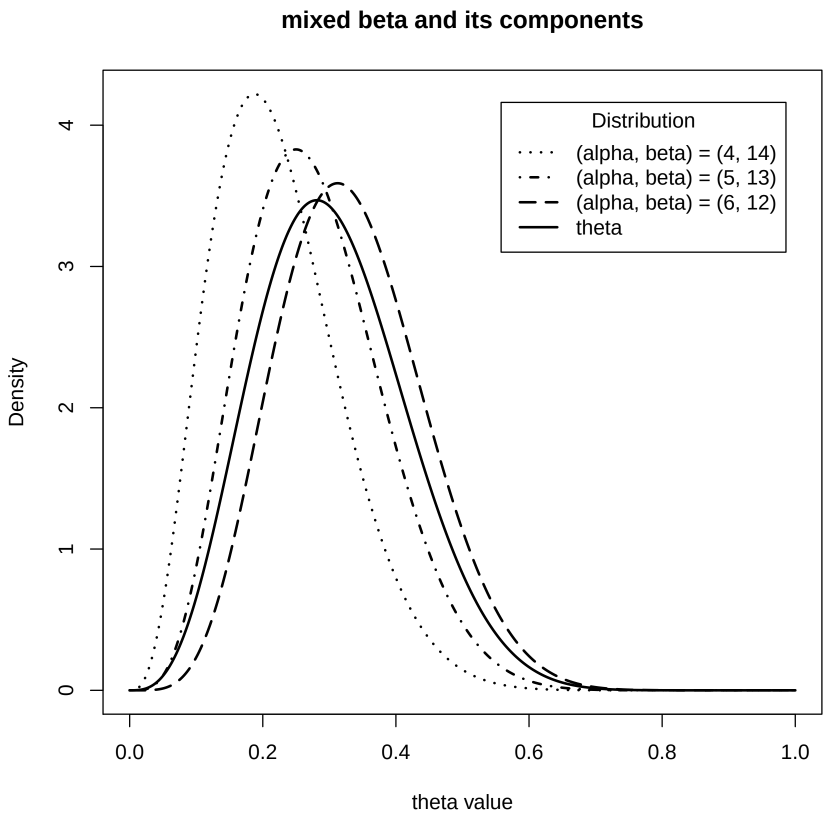
$$\begin{aligned}
p(\theta|x<3) &= \frac{\Gamma(4)\Gamma(14)}{Constant \cdot \Gamma(18)} \cdot \frac{\Gamma(18)}{\Gamma(4)\Gamma(14)}\theta^3(1-\theta)^{13} + \frac{10 \cdot \Gamma(5)\Gamma(13)}{Constant \cdot \Gamma(18)} \cdot \frac{\Gamma(18)}{\Gamma(5)\Gamma(13)}\theta^4(1-\theta)^{12} \\
&\quad + \frac{45 \cdot \Gamma(6)\Gamma(12)}{Constant \cdot \Gamma(18)} \cdot \frac{\Gamma(18)}{\Gamma(6)\Gamma(12)}\theta^5(1-\theta)^{11} \\
&= \frac{\Gamma(4)\Gamma(14)}{Constant \cdot \Gamma(18)} Beta(\theta|4,14) + \frac{10 \cdot \Gamma(5)\Gamma(13)}{Constant \cdot \Gamma(18)} Beta(\theta|5,13) \\
&\qquad + \frac{45 \cdot \Gamma(6)\Gamma(12)}{Constant \cdot \Gamma(18)} Beta(\theta|6,12) \\
&= \frac{B(4,14)}{Constant} Beta(\theta|4,14) + \frac{10 \cdot B(5,13)}{Constant} Beta(\theta|5,13) + \frac{45 \cdot B(6,12)}{Constant} Beta(\theta|6,12)
\end{aligned}$$

This shows that the posterior distribution is a mixed beta distribution, with the combination coefficients and the beta parameters given as above.

The following code plot each component of the above mixed beta distribution, as well as the mixed beta distribution itself.

```r
#define constant
C <- beta(4, 14) + 10*beta(5, 13) + 45*beta(6, 12)
#define domain of beta distribution
x <- seq(0, 1, length = 1000)
#define the components of the mixed beta distribution
hx1 <- dbeta(x, 4, 14)
hx2 <- dbeta(x, 5, 13)
hx3 <- dbeta(x, 6, 12)
labels <- c("(alpha, beta) = (4, 14)", "(alpha, beta) = (5, 13)",
            "(alpha, beta) = (6, 12)", "theta")
```

```
plot(x, hx1, type="l", lty = 3, lwd = 2, xlab = "theta value",
   ylab = "Density", main = "mixed beta and its components")
lines(x, hx2, lty = 4, lwd = 2)
lines(x, hx3, lty = 5, lwd = 2)
lines(x, (beta(4, 14)*hx1 + 10*beta(5, 13)*hx2+45*beta(6, 12)*hx3)/C,
    lty = 1, lwd = 2)
legend("topright", inset = .05, title = "Distribution",
     labels, lwd = 2, lty = c(3, 4, 5, 1))
```



Chapter 2 Question 5 on page 58:

Solution 3. *part (a): If the prior distribution for θ is uniform on $[0, 1]$, then the prior predictive* *distribution for y (the marginal distribution for y) is:*

$$\begin{aligned}
Pr(y = k) &= \int_0^1 Pr(y = k|\theta)d\theta \\
&= \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} d\theta \\
&= \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \int_0^1 \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-+1)} \cdot \theta^{(k+1)-1}(1-\theta)^{[(n-k+1)-1]} d\theta \\
&= \binom{n}{k} B(k+1, n-k+1) = \frac{n!}{k!(n-k)!} B(k+1, n-k+1)
\end{aligned}$$

The above holds true for any $k = 0, 1, \ldots, n$.

part (b): suppose θ has prior distribution $Beta(\alpha, \beta)$. We need to show that the mean of the posterior distribution of θ lies between $\frac{alpha}{alpha+beta}$ and $\frac{y}{n}$.

For the posterior distribution of θ, we have:

$$\begin{aligned}
& p(\theta|y) \propto p(\theta) \cdot p(y|\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \binom{n}{y} \theta^y (1-\theta)^{n-y} \\
& \Longrightarrow p(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}
\end{aligned}$$

So the posterior distribution of θ is $Beta(y+\alpha, n-y+\beta)$, thus the mean is

$$\frac{y+\alpha}{n+\alpha+\beta}$$

and we need to show it is between $\frac{\alpha}{\alpha+\beta}$ and $\frac{y}{n}$.

Without loss of generality, we can first assume that $\frac{\alpha}{\alpha+\beta} < \frac{y}{n}$, then we have:

$$\begin{aligned}
& \begin{cases} \alpha n < y(\alpha+\beta) \\ y(\alpha+\beta) > \alpha n \end{cases} \Longrightarrow \begin{cases} yn + \alpha n < yn + y\alpha + y\beta \\ y(\alpha+\beta) + \alpha(\alpha+\beta) > \alpha n + \alpha(\alpha+\beta) \end{cases} \\
& \Longrightarrow \begin{cases} (y+\alpha) \cdot n < y(n+\alpha+\beta) \\ (y+\alpha)(\alpha+\beta) > \alpha(n+\alpha+\beta) \end{cases} \Longrightarrow \begin{cases} \frac{y+\alpha}{n+\alpha+\beta} < \frac{y}{n} \\ \frac{y+\alpha}{n+\alpha+\beta} > \frac{\alpha}{\alpha+\beta} \end{cases} \Longrightarrow \frac{\alpha}{\alpha+\beta} < \frac{y+\alpha}{n+\alpha+\beta} < \frac{y}{n}
\end{aligned}$$

If it is the other case that $\frac{\alpha}{\alpha+\beta} > \frac{y}{n}$, then we repeat the same algebra by reversing all inequalities above, then we get

$$\frac{y}{n} < \frac{y+\alpha}{n+\alpha+\beta} < \frac{\alpha}{\alpha+\beta}$$

Either case, we have proved our claim.

part (c): suppose the prior for θ is uniform, then $\theta \sim U(0,1)$ and we have the prior variance as

$$Var(\theta) = \frac{1}{12}$$

We need to show that for the posterior variance of θ, we have:

$$Var(\theta|y) < \frac{1}{12} \text{ for any } y \text{ and } n.$$

We can first compute the posterior distribution:

$$p(\theta|y) \propto p(y|\theta)p(\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$
$$\Longrightarrow p(\theta|y) \propto \theta^y(1-\theta)^{n-y}$$

Thus we know the posterior distribution: $p(\theta|y) \sim Beta(y+1, n-y+1)$, and hence the posterior varinace is:

$$Var(\theta|y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{(y+1)(n-y+1)}{(n+2)^2(n+3)} = \frac{-(y+1)[y-(n+1)]}{(n+2)^2(n+3)}$$

Notice that the numerator is the equation for a downward parabola with roots at $y = -1$ and $y = n+1$, so the maximum of the numerator is achieved at the center of the two roots: $y = \frac{1}{2}(-1+n+1) = \frac{n}{2}$, where the maximum value of the numerator is

$$-(y+1)[y-(n+1)]\Big|_{y=\frac{n}{2}} = -(\frac{n}{2}+1)(\frac{n}{2}-n-1) = (\frac{n}{2}+1)^2 = \frac{1}{4}(n+2)^2$$

So we have:

$$Var(\theta|y) = \frac{-(y+1)[y-(n+1)]}{(n+2)^2(n+3)} \le \frac{\frac{1}{4}(n+2)^2}{(n+2)^2(n+3)} \le \frac{1}{4(n+3)} \le \frac{1}{12}$$

Since $\frac{1}{4(n+3)}$ is strictly decreasing with respect to n, and the equality is only achieved when n starts with 0, thus in reality the inequality is always strict.

part (d), we need to find an example of y, n, α, β such that given the $Beta(\alpha, \beta)$ prior distribution, the posterior variance of θ is higher than the prior variance

Notice that the prior variance is:

$$Var(\theta) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

and the posterior variance is

$$Var(\theta|y) = \frac{(y+\alpha)(n-y+\beta)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)}$$

If we consider n, α and β as already being fixed and y is the only variable, then

$$(y+\alpha)(n-y+\beta)\Big|_{\max} = (y+\alpha)(n-y+\beta)\Big|_{y=\frac{n+\beta-\alpha}{2}} = \frac{1}{4}(n+\beta+\alpha)^2$$

So

$$Var(\theta|y)\Big|_{max} = \frac{1}{4(\alpha+\beta+1)}$$

So we just need to choose α, β *wisely such that*

$$\frac{\alpha\beta}{(\alpha+\beta)^2} < \frac{1}{4}$$

From the inequality $(\alpha+\beta)^2 \geq 4\alpha\beta$ *and the attaining of equality when* $\alpha = \beta$*, we know that we need to set apart the values of* α *and* β *in order to reach our goal. After some experiment we choose the following:*

Let $\alpha = 2, \beta = 8, n = 4$ *and* $y = \frac{n+\beta-\alpha}{2} = 5$*, then we have*

$$Var(\theta|y) = \frac{1}{60} \simeq 0.0167$$
$$Var(\theta) = \frac{16}{100 \cdot 11} \simeq 0.0145$$

So this example meets the requirement.