Worksheet 2

Foundations of Bayesian Methodology

Goliath : Wenje Tu, Lea Bührer, Jerome Sepin, Zhixuan Li

Spring Semester 2022

Exercise 3 (Conjugate Bayes: analytical derivation)

Assumptions:

$$\begin{aligned} y_1, \cdots, y_n \mid m &\overset{i.i.d}{\sim} \mathcal{N}(m, \kappa^{-1}) \\ m &\sim \mathcal{N}(\mu, \lambda^{-1}) \end{aligned}$$

Likelihood function:

$$\begin{aligned} f(y_1, \cdots, y_n \mid m) &= \prod_{i=1}^{n} f(y_i \mid m) \\ &= \prod_{i=1}^{n} \left(\sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{\kappa}{2}(y_i - m)^2 \right) \right) \\ &= \left(\frac{\kappa}{2\pi} \right)^{\frac{n}{2}} \exp\left(-\frac{\kappa}{2} \sum_{i=1}^{n} (y_i - m)^2 \right) \end{aligned}$$

Prior density distribution of m:

$$f(m) = \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(m - \mu)^2 \right)$$

Multiply likelihood function by prior:

$$
\begin{aligned}
&f(y_1,\cdots,y_n\mid m)f(m)\\
&=\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n}(y_i-m)^2\right)\cdot\sqrt{\frac{\lambda}{2\pi}}\exp\left(-\frac{\lambda}{2}(m-\mu)^2\right)\\
&=\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n}(y_i-m)^2-\frac{\lambda}{2}(m-\mu)^2\right)\\
&=\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n}(y_i^2-2y_im+m^2)-\frac{\lambda}{2}(m^2-2m\mu+\mu^2)\right)\\
&=\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{\kappa}{2}\left(\sum_{i=1}^{n}y_i^2-2n\bar{y}m+nm^2\right)-\frac{\lambda}{2}(m^2-2m\mu+\mu^2)\right)\\
&=\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2}\left((n\kappa+\lambda)m^2-2m(\kappa n\bar{y}+\lambda\mu)+\kappa\sum_{i=1}^{n}y_i^2+\lambda\mu^2\right)\right)\\
&=\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{n\kappa+\lambda}{2}\left(m^2-2m\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}+\frac{\kappa\sum_{i=1}^{n}y_i^2+\lambda\mu^2}{n\kappa+\lambda}\right)\right)\\
&=\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{n\kappa+\lambda}{2}\left(\left(m-\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2-\left(\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2+\frac{\kappa\sum_{i=1}^{n}y_i^2+\lambda\mu^2}{n\kappa+\lambda}\right)\right)\\
&=\underbrace{\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}}\left(\frac{\lambda}{2\pi}\right)^{\frac{1}{2}}\exp\left(\frac{\kappa\sum_{i=1}^{n}y_i^2+\lambda\mu^2}{n\kappa+\lambda}-\left(\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2\right)}_{\text{constant}}\exp\left(-\frac{n\kappa+\lambda}{2}\left(m-\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2\right)
\end{aligned}
$$

Using Bayes formula, we can write posterior as a function of prior and likelihood:

$$
\begin{aligned}
f(m\mid y_1,\cdots,y_n)&=\frac{f(y_1,\cdots,y_n\mid m)f(m)}{f(y_1,\cdots,y_n)}\\
&=\frac{f(y_1,\cdots,y_n\mid m)f(m)}{\int_{-\infty}^{\infty}f(y_1,\cdots,y_n\mid m)f(m)\mathrm{d}m}\\
&=\frac{\exp\left(-\frac{n\kappa+\lambda}{2}\left(m-\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{n\kappa+\lambda}{2}\left(m-\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2\right)dm}\\
&=\frac{\exp\left(-\frac{n\kappa+\lambda}{2}\left(m-\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2\right)}{\frac{1}{\sqrt{\frac{n\kappa+\lambda}{2\pi}}}\underbrace{\int_{-\infty}^{\infty}\sqrt{\frac{n\kappa+\lambda}{2\pi}}\exp\left(-\frac{n\kappa+\lambda}{2}\left(m-\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2\right)dm}_{\text{integrates to 1}}}\\
&=\sqrt{\frac{n\kappa+\lambda}{2\pi}}\exp\left(-\frac{n\kappa+\lambda}{2}\left(m-\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda}\right)^2\right)
\end{aligned}
$$

$$
m\mid y_1,\cdots,y_n\sim\mathcal{N}\left(\frac{\kappa n\bar{y}+\lambda\mu}{n\kappa+\lambda},\frac{1}{n\kappa+\lambda}\right)
$$

Note:

$$
\int_{-\infty}^{\infty}e^{-a(x+b)^2}\mathrm{d}x=\sqrt{\frac{\pi}{\alpha}}
$$

Source: Gaussian Integral

Exercise 4 (Conjugate Bayesian analysis in practice)

Assumptions:

$$y_1, \cdots, y_n \mid m \overset{i.i.d}{\sim} \mathcal{N}(m, \kappa^{-1})$$
$$m \sim \mathcal{N}(\mu, \lambda^{-1})$$

$$\kappa = \frac{1}{900}, \quad \mu = 161, \quad \lambda = \frac{1}{70}$$

4(a) Summary statistics:

```
height <- c(166, 168, 168, 177, 160, 170, 172, 159, 175, 164, 175, 167, 164)
summary(height)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   159.0   164.0   168.0   168.1   172.0   177.0
```

```
hist(height, breaks=length(height), freq=FALSE)
lines(density(height), col="red")
```

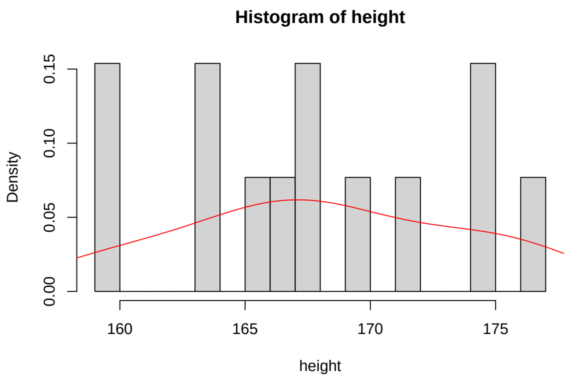

Figure 1: Histogram of the height measurements

```
library(car)
qqPlot(height, main="Normal Q-Q Plot", ylab="Sample", xlab="Norm Quantiles")
```

```
## [1] 8 4
```

```
# or alternatively without library(car)
qqnorm(height)
qqline(height)
```

In the Figure 2 (i.e. Q-Q Plot), it can be seen that all sample values lie within the area where normal distribution of the data can be assumed. Due to the fact that we only have 13 observations the 95%-CI has been calculated assuming a t-distribution.

```
## Sample size
sample.n <- length(height); sample.n
```

```
## [1] 13
```

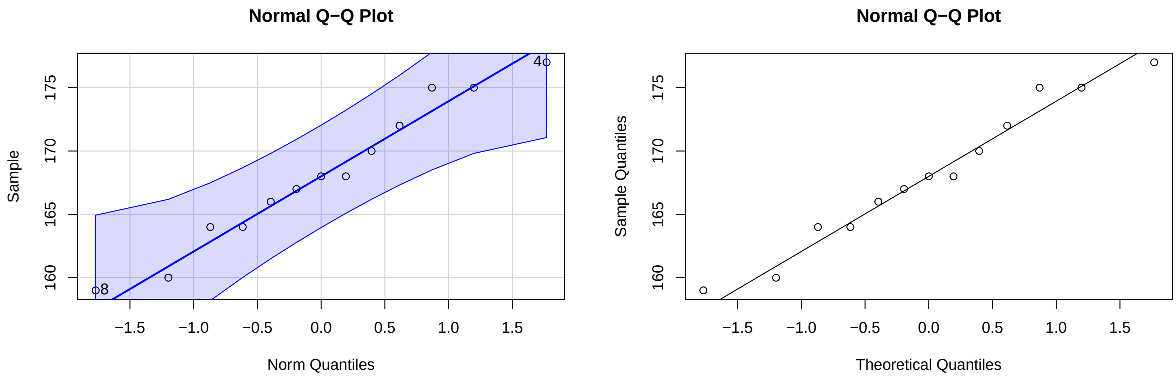

Figure 2: Q-Q Plot of the Measurements

```
## Sample median
sample.med <- median(height); sample.med
```

```
## [1] 168
```

```
## Sample meam
sample.mean <- mean(height); sample.mean
```

```
## [1] 168.0769
```

```
## Sample standard deviation
sample.sd <- sd(height); sample.sd
```

```
## [1] 5.634145
```

```
## Sample standard error
sample.se <- sample.sd / sqrt(sample.n); sample.se
```

```
## [1] 1.562631
```

```
## Significance level
alpha <- 0.05

## Degrees of freedom
df <- sample.n - 1

## t-score
t.score <- qt(p=alpha/2, df=df, lower.tail=FALSE); t.score
```

```
## [1] 2.178813
```

```
## Contract the 95% confidence interval
lower.bound <- sample.mean - t.score * sample.se
upper.bound <- sample.mean + t.score * sample.se
print(c(lower.bound, upper.bound))
```

```
## [1] 164.6722 171.4816
```

Interpretation: For repeated random samples from a normal distribution with unknown but fixed mean, the 95% confidence interval $(164.6722, 171.4816)$ will cover the true unknown mean in 95% of all cases.

Table 1: Summary statistics of the sample distribution

	Mean	Standard deviation	Median	95% CI
$y_1, \cdots, y_n$	168.0769	5.6341	168.0000	$(164.6722, 171.4816)$

4(b) We first plot the prior distribution of m:

```
## Plot of the Prior Distribution:
m <- seq(101, 221)
plot(m, dnorm(m, mean=161, sd=sqrt(70)), type="l", ylab="density", col="red",
     ylim=c(0,0.07), main="Prior distribution", lwd=2)
legend("topright", legend=c("prior: N(161,70)"), lwd=2,
       lty=1, col=c("red"), cex=.8)
```

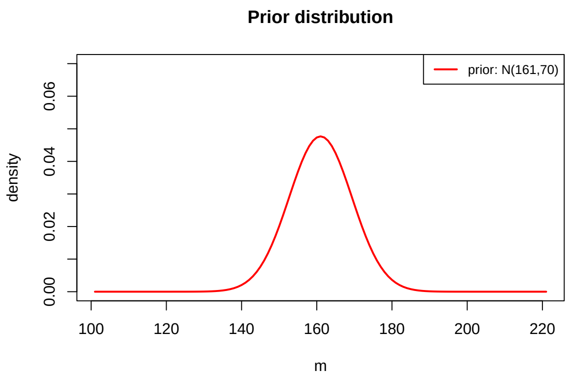

Figure 3: Prior Distribution

The expectation and the standard deviation can be explicitly obtained from the given information. Hence, we only need to compute the median and equi-tailed 95% interval. We use `qnorm()` function in `R` to serve this purpose. Regarding the estimation of $P[m > 200]$, we simply use `pnorm()` in `R` to obtain the corresponding probability.

$$P[m > 200] = 1 - P[m \leq 200]$$

```
## 2.5%, 50%(i.e. median), 97.5% quantiles of N(161,70)
qnorm(c(0.025, 0.5, 0.975), mean=161, sd=sqrt(70))
```

```
## [1] 144.6018 161.0000 177.3982
```

```
## P[m>200]
pnorm(200, mean=161, sd=sqrt(70), lower.tail=FALSE)
```

```
## [1] 1.570393e-06
```

Summary statistics: see Table 2

4(c)

$$\begin{aligned} y_1, \cdots, y_n \mid m &\overset{i.i.d}{\sim} \mathcal{N}(m, \kappa^{-1}) \\ m &\sim \mathcal{N}(\mu, \lambda^{-1}) \\ m \mid y_1, \cdots, y_n &\sim \mathcal{N}\left(\frac{\kappa n \bar{y} + \lambda\mu}{n\kappa + \lambda}, (n\kappa + \lambda)^{-1}\right) \end{aligned}$$

```
## Sum of measurements
(sum(height))
```

```
## [1] 2185
```

```
## Posterior mean
((1/900)*2185+(1/70)*161 ) / (13*(1/900)+(1/70) )
```

```
## [1] 164.558
```

```
## Posterior variance
(13*(1/900)+(1/70))^(-1)
```

```
## [1] 34.80663
```

$$\begin{aligned} y_1, \cdots, y_n &\sim \mathcal{N}(m, 900) \\ m &\sim \mathcal{N}(161, 70) \\ m \mid y_1, \cdots, y_n &\sim \mathcal{N}(164.558, 34.80663) \end{aligned}$$

```
## Visualization of prior and posterior
m <- seq(101, 221)
plot(m, dnorm(m, mean=161, sd=sqrt(70)), type="l", ylab="density", col="red",
     ylim=c(0,0.08), main="Prior vs Posterior", lwd=2)
lines(m, dnorm(m, mean=164.558, sd=sqrt(34.80663)), col="blue", lwd=2)
legend("topright", legend=c("prior: N(161,70)", "posterior: N(164.56, 34.81)"),
       lty=1, col=c("red", "blue"), cex=.8, lwd=2)
```

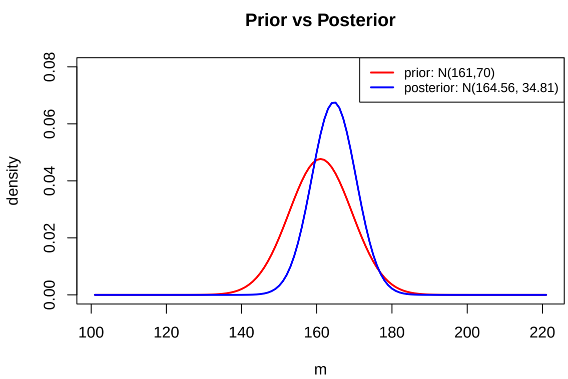

Figure 4: Prior vs. Posterior

```
qnorm(c(0.025, 0.5, 0.975), mean=164.558, sd=sqrt(34.80663))
```

```
## [1] 152.9948 164.5580 176.1212
```

Interpretation: the posterior belief about the mean `Height` m lies between 152.9948 and 176.1212 with a probability of 95%, given a $\mathcal{N}(161, 70)$ prior is assumed.

4(d)

$$P[m > 200 \mid y_1, ..., y_n]$$

Table 2: Summary statistics of the sample, prior and posterior distributions

	Mean	Standard deviation	Median	Equi-tailed 95% CI/CrI
$y_1, \cdots, y_n$	168.0769	5.6341	168.0000	$(164.6722, 171.4816)$
m	161.0000	8.3666	161.0000	$(144.6018, 177.3982)$
$m \mid y_i, \cdots, y_n$	164.5580	5.8997	164.5580	$(152.9948, 176.1212)$

```
## P[m>200|y1,...,yn]
pnorm(200, mean=164.558, sd=sqrt(34.80663), lower.tail=FALSE)
```

```
## [1] 9.425552e-10
```

```
# or
1 - pnorm(200, mean=164.558, sd=sqrt(34.80663))
```

```
## [1] 9.425551e-10
```

The posterior probability that an adult Swiss female has a height larger than 200 is 9.426×10^{-10}.

4(e)

$$\text{Prior} \rightarrow \text{Posterior}$$
$$\mathcal{N}(161, 70) \rightarrow \mathcal{N}(164.558, 34.80663)$$
$$P[m > 200] \rightarrow P[m > 200 \mid y_1, \cdots, y_n]$$
$$1.570393 \times 10^{-6} \rightarrow 9.425552 \times 10^{-10}$$

From prior to posterior, we see an increase in the mean of m from 161 to 164.558 and a decrease in the variance of m from 70 to 34.80663. Figure 5 displays a huge overlap between the prior distribution and the likelihood density, it is not surprising that the posterior mean lies somewhere between the prior mean and the sample mean and that the posterior variance lies somewhere between the prior variance and sample variance. Since both prior and likelihood mostly agree, we see a more concentrated posterior distribution with light tails. Thus, the probability of observing a Swiss female with a height greater than 200 also decreases.

```
m <- seq(101, 221)
plot(m, dnorm(m, mean=161, sd=sqrt(70)), type="l", ylab="density", col="red",
     ylim=c(0,0.08), main="Prior vs Likelihood vs Posterior", lwd=2)
lines(m, dnorm(m, mean=164.558, sd=sqrt(34.80663)), col="blue", lwd=2)
lines(density(height), col="orange", lwd=2)
legend("topright", lty=1, col=c("red", "blue", "orange"), cex=.8, lwd=2,
       legend=c("prior: N(161,70)", "posterior: N(164.56, 34.81)", "likelihood density"))
```

Exercise 5 (Bayesian learning)

Prior distribution:

$$p \sim \text{Beta}(\alpha, \beta)$$

Posterior distribution:

$$p \mid y_1, \cdots, y_n \sim \text{Beta}(\alpha + n\bar{y}, \beta + n - n\bar{y})$$

5(a) $\alpha = \beta = 0.5, \quad \text{Beta}(0.5, 0.5)$

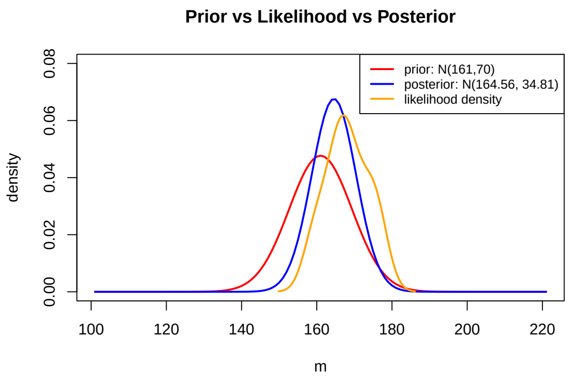

Figure 5: Prior vs. Likelihood vs. Posterior

Table 3: Evidence P(response rate > 0.4) for each stage under different priors

Prior $B(\alpha, \beta)$	Stage	# n	Responders x (%)	Posterior $B(\alpha + x, \beta + n - x)$	Evidence P(response rate > 0.4)
	No data	0	0 (0%)	$B(0.5, 0.5)$	0.5640942
$B(0.5, 0.5)$	Interim	12	3 (25%)	$B(3.5, 9.5)$	0.1437649
	Final	64	14 (21.875%)	$B(14.5, 50.5)$	0.001075757
	No data	0	0 (0%)	$B(8, 24)$	0.03298768
$B(8, 24)$	Interim	12	3 (25%)	$B(11, 33)$	0.01621346
	Final	64	14 (21.875%)	$B(22, 74)$	0.0001727695

Table 4: Summary statistics of posterior distributions

Prior	Stage	Posterior $B(\alpha, \beta)$	Mean $\frac{\alpha}{\alpha+\beta}$	Median 50% quantile	Equi-tailed 95% CrI
	No data	$B(0.5, 0.5)$	0.5000	0.5000	$(0.0015, 0.9985)$
$B(0.5, 0.5)$	Interim	$B(3.5, 9.5)$	0.2692	0.2571	$(0.0759, 0.5292)$
	Final	$B(14.5, 50.5)$	0.2231	0.2202	$(0.1313, 0.3310)$
	No data	$B(8, 24)$	0.2500	0.2447	$(0.1186, 0.4110)$
$B(8, 24)$	Interim	$B(11, 33)$	0.2500	0.2461	$(0.1352, 0.3863)$
	Final	$B(22, 74)$	0.2292	0.2273	$(0.1512, 0.3178)$

```r
p <- seq(1e-3,1, length=200)

plot(p, dbeta(p, 0.5, 0.5), type="l", ylab="density", col="red",
     ylim=c(0, 10), main="Intermediate study", lwd=2)
lines(p, dbeta(p, 3.5, 9.5), ylab="density", col="blue", lwd=2)

legend("topright", legend=c("prior: B(0.5,0.5)", "posterior: B(3.5,9.5)"),
       col=c("red", "blue"), lty=1, cex=.8, lwd=2)

plot(p, dbeta(p, 0.5, 0.5), type="l", ylab="density", col="red",
     ylim=c(0, 10), main="Final study", lwd=2)
lines(p, dbeta(p, 14.5, 50.5), ylab="density", col="blue", lwd=2)

legend("topright", legend=c("prior: B(0.5,0.5)", "posterior: B(14.5,50.5)"),
       col=c("red", "blue"), lty=1, cex=.8, lwd=2)
```

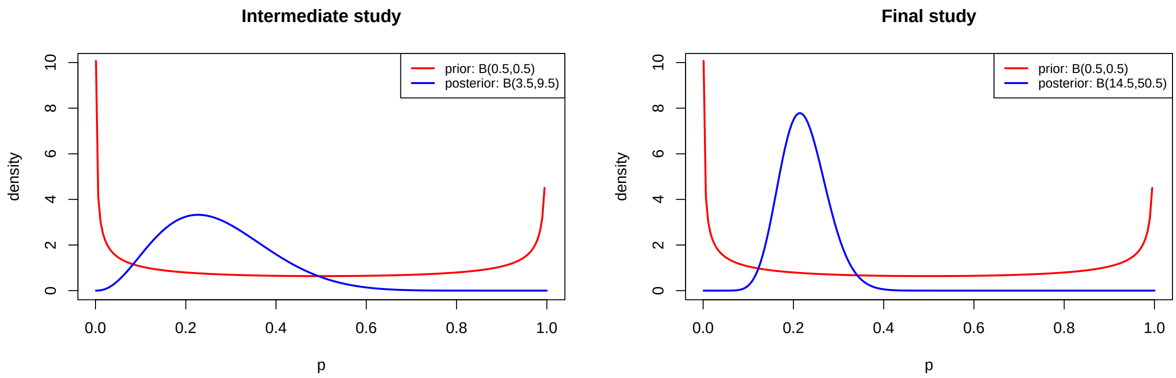

Figure 6: Prior B(0.5,0.5)

$$P[\text{response rate} > 0.4] = P[p > 0.4] = 1 - P[p \leq 0.4]$$

```r
# Before seeing any data (i.e. only prior belief)
pbeta(0.4, 0.5, 0.5, lower.tail=FALSE)
```

```
## [1] 0.5640942
```

```r
# Intermediate study
pbeta(0.4, 3.5, 9.5, lower.tail=FALSE)
```

```
## [1] 0.1437649
```

```r
# Final study
pbeta(0.4, 14.5, 50.5, lower.tail=FALSE)
```

```
## [1] 0.001075757
```

```r
# Before observing data
## 2.5%, 50%, 97.5% quantiles of Beta(0.5, 0.5)
qbeta(c(0.025, 0.5, 0.975), 0.5, 0.5)
```

```
## [1] 0.001541333 0.500000000 0.998458667
```

```r
# Intermediate study
## 2.5%, 50%, 97.5% quantiles of Beta(3.5, 9.5)
qbeta(c(0.025, 0.5, 0.975), 3.5, 9.5)
```

```
## [1] 0.07594233 0.25711895 0.52919108
```

```r
# Final study
## 2.5%, 50%, 97.5% quantiles of Beta(14.5, 50.5)
qbeta(c(0.025, 0.5, 0.975), 14.5, 50.5)
```

```
## [1] 0.1312669 0.2202242 0.3310055
```

5(b) $\alpha = 8, \beta = 24, \quad \text{Beta}(8, 24)$

$$P[\text{response rate} > 0.4] = P[p > 0.4] = 1 - P[p \leq 0.4]$$

```r
p <- seq(1e-3,1, length=200)

plot(p, dbeta(p, 8, 24), type="l", ylab="density", col="red",
     ylim=c(0, 10), main="Intermediate study", lwd=2)
lines(p, dbeta(p, 11, 33), ylab="density", col="blue", lwd=2)

legend("topright", legend=c("prior: B(8,24)", "posterior: B(11,33)"),
       col=c("red", "blue"), lty=1, cex=.8, lwd=2)

plot(p, dbeta(p, 8, 24), type="l", ylab="density", col="red",
     ylim=c(0, 10), main="Final study", lwd=2)
lines(p, dbeta(p, 22, 74), ylab="density", col="blue", lwd=2)

legend("topright", legend=c("prior: B(8,24)", "posterior: B(22,74)"),
       col=c("red", "blue"), lty=1, cex=.8, lwd=2)
```

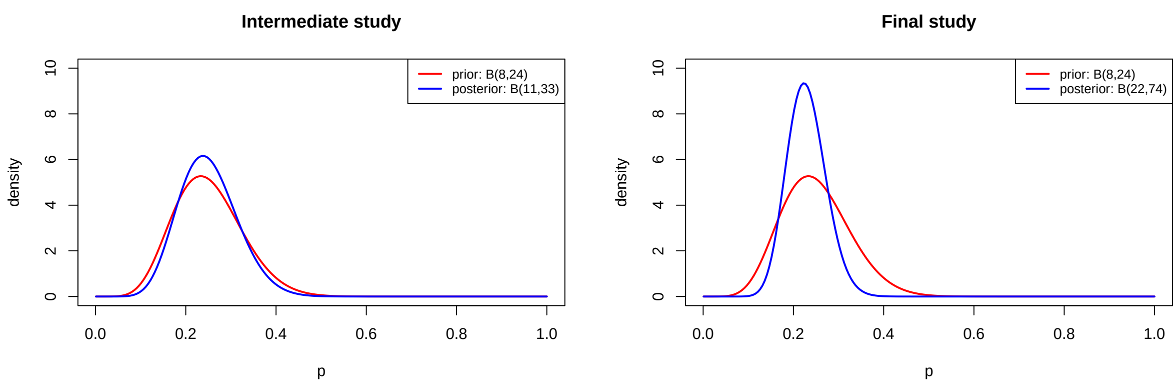

Figure 7: Prior B(8,24)

```r
# Before seeing any data
pbeta(0.4, 8, 24, lower.tail=FALSE)
```

```
## [1] 0.03298768
```

```r
# Intermediate study
pbeta(0.4, 11, 33, lower.tail=FALSE)
```

```
## [1] 0.01621346
```

```r
# Final study
pbeta(0.4, 22, 74, lower.tail=FALSE)
```

```
## [1] 0.0001727695
```

```r
# Before observing data
## 2.5%, 50%, 97.5% quantiles of Beta(8, 24)
qbeta(c(0.025, 0.5, 0.975), 8, 24)
```

```
## [1] 0.1185640 0.2447417 0.4109639
```

```r
# Intermediate study
## 2.5%, 50%, 97.5% quantiles of Beta(11, 33)
qbeta(c(0.025, 0.5, 0.975), 11, 33)
```

```
## [1] 0.1351860 0.2461854 0.3863082
```

```r
# Final study
## 2.5%, 50%, 97.5% quantiles of Beta(22, 74)
qbeta(c(0.025, 0.5, 0.975), 22, 74)
```

```
## [1] 0.1511774 0.2272801 0.3178360
```