Master solution of group tasks

Exercise 3 (Conjugate Bayes: analytical derivation - 10 Points)

$$y_1, \dots, y_n \overset{\text{iid}}{\sim} \mathcal{N}(m, \kappa^{-1}), \tag{1}$$
$$m \sim \mathcal{N}(\mu, \lambda^{-1}), \tag{2}$$

where κ, μ and λ are fixed constants. Derive the posterior $f(m \mid y_1, \dots, y_n)$ with all constants.

The likelihood

$$L(y_1, \dots, y_n \mid m) = \prod_{i=1}^{n} \sqrt{\frac{\kappa}{2\pi}} \exp\left\{-\frac{\kappa}{2}(y_i - m)^2\right\} = \left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{\kappa}{2}\sum_{i=1}^{n}(y_i - m)^2\right\}.$$

The prior

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

Posterior by Bayes theorem

$$f(m \mid y_1, \dots, y_n) = \frac{L(y_1, \dots, y_n \mid m) f(m)}{f(y_1, \dots, y_n)}. \tag{3}$$

$$f(y_1, \dots, y_n) = \int_{-\infty}^{\infty} L(y_1, \dots, y_n \mid m) f(m) dm. \tag{4}$$

Numerator

$$\begin{aligned}
L(y_1, \dots, y_n \mid m) f(m) &= \left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \exp\left\{-\frac{\kappa}{2}\sum_{i=1}^{n}(y_i - m)^2\right\} \sqrt{\frac{\lambda}{2\pi}} \exp\left\{-\frac{\lambda}{2}(m - \mu)^2\right\} \\
&= \left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \sqrt{\frac{\lambda}{2\pi}} \exp\left\{-\frac{\kappa}{2}\sum_{i=1}^{n}(y_i - m)^2 - \frac{\lambda}{2}(m - \mu)^2\right\} \\
&= \left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \sqrt{\frac{\lambda}{2\pi}} \exp\left\{-\frac{1}{2}\left[(\lambda + \kappa n)\left(m^2 - 2m\frac{\mu\lambda + \kappa n\bar{y}}{\lambda + \kappa n}\right) + \kappa\sum_{i=1}^{n} y_i^2 + \lambda\mu^2\right]\right\} \\
= \left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \sqrt{\frac{\lambda}{2\pi}} &\exp\left\{-\frac{1}{2}\left(\kappa\sum_{i=1}^{n} y_i^2 + \lambda\mu^2 - \frac{(\lambda\mu + \kappa n\bar{y})^2}{\lambda + n\kappa}\right)\right\} \exp\left\{-\frac{\lambda + n\kappa}{2}\left(m - \frac{\lambda\mu + \kappa n\bar{y}}{\lambda + n\kappa}\right)^2\right\}
\end{aligned} \tag{5}$$

Denominator

$$\int_{-\infty}^{\infty} L(y_1,\dots,y_n \mid m) f(m) dm =$$

$$\left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \sqrt{\frac{\lambda}{2\pi}} \exp\left\{-\frac{1}{2}\left(\kappa \sum_{i=1}^{n} y_i^2 + \lambda\mu^2 - \frac{(\lambda\mu + \kappa n\bar{y})^2}{\lambda + n\kappa}\right)\right\} \frac{1}{\left(\frac{\lambda+n\kappa}{2\pi}\right)^{\frac{1}{2}}}$$
$$\underbrace{\int \left(\frac{\lambda+n\kappa}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda+n\kappa}{2}\left(m - \frac{\lambda\mu+\kappa n\bar{y}}{\lambda+n\kappa}\right)^2\right\} dm}_{=1} \tag{6}$$
$$= \left(\frac{\kappa}{2\pi}\right)^{\frac{n}{2}} \sqrt{\frac{\lambda}{2\pi}} \exp\left\{-\frac{1}{2}\left(\kappa \sum_{i=1}^{n} y_i^2 + \lambda\mu^2 - \frac{(\lambda\mu + \kappa n\bar{y})^2}{\lambda + n\kappa}\right)\right\} \frac{1}{\left(\frac{\lambda+n\kappa}{2\pi}\right)^{\frac{1}{2}}}$$

Let's plug in the numerator and denominator into equation (3) by taking the equation (4) into account. We get

$$f(m \mid y_1,\dots,y_n) = \left(\frac{\lambda+n\kappa}{2\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda+n\kappa}{2}\left(m - \frac{\lambda\mu+\kappa n\bar{y}}{\lambda+n\kappa}\right)^2\right\} \tag{7}$$

Therefore,

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

Exercise 4 (Conjugate Bayesian analysis in practice - 10 Points)

Given the following

```r
Height <- c(166, 168, 168, 177, 160, 170, 172, 159, 175, 164, 175, 167, 164)
n <- length(Height)
y_bar <- mean(Height)
kappa <- 1/900
mu <- 161
lambda <- 1/70
```

```r
###(a)
summary(Height)

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

The classical $\gamma \cdot 100\%$ confidence interval for the true mean can be calculated as

$$\mathrm{CI}_\mathrm{l} = \bar{y} - q_{t_{n-1,(1+\gamma)/2}} \frac{S}{\sqrt{n}} \tag{8}$$
$$\mathrm{CI}_\mathrm{u} = \bar{y} + q_{t_{n-1,(1+\gamma)/2}} \frac{S}{\sqrt{n}}, \tag{9}$$
$$\tag{10}$$

where S is sample standard deviation.

```
gamma <- 0.95
y_bar # sample mean

## [1] 168.0769

(S <- sd(Height)) # sample standard deviation

## [1] 5.634145

#first approach for classical CI
(CI_l <- y_bar - qt((1 + gamma) / 2, df = n - 1) * S/sqrt(n))

## [1] 164.6722

(CI_u <- y_bar + qt((1 + gamma) / 2, df = n - 1) * S/sqrt(n))

## [1] 171.4816

#second approach for classical CI
library(DescTools)
MeanCI(Height, method = "classic", conf.level = 0.95)

##     mean   lwr.ci   upr.ci
## 168.0769 164.6722 171.4816
```

Classical confidence intervals can be interpreted as follows: for repeated random samples from a distribution with unknown parameter m, a 95 % CI will cover m in 95 % of all cases (see Held and Sebanés Bové, 2020).

The prior distribution of m is drawn in Figure 2.

```
###(b)
#expectation of m
(mean_prior <- mu)

## [1] 161

#standard deviation of m
(sd_prior <- sqrt(lambda^(-1)))

## [1] 8.3666

#median of m  #lower confidence bound of m  #upper confidence bound of m
alpha <- 0.05
(quantiles_prior <- qnorm(p = c(alpha/2, 0.5, 1 - alpha / 2),
                          mean = mean_prior, sd = sd_prior))
```

```
library(MASS)
par(mfrow = c(1, 2), mai = c(0.2,0.5, 0.5, 0.5), pty = "s")
truehist(Height, main = "Histogram of height",
         xlab = "height", ylim = c(0, 0.1),
         xlim = c(151, 185), col = "grey")
abline(v = mean(Height), col = "darkgreen", lwd = 2)
legend("topleft", legend = c("Height", "sample mean"),
       col = c("grey", "darkgreen"),
      pch = c(22, NA), lwd = 2)
boxplot(Height, main = "Boxplot of height",
        horizontal = TRUE, xlab = "height")
```

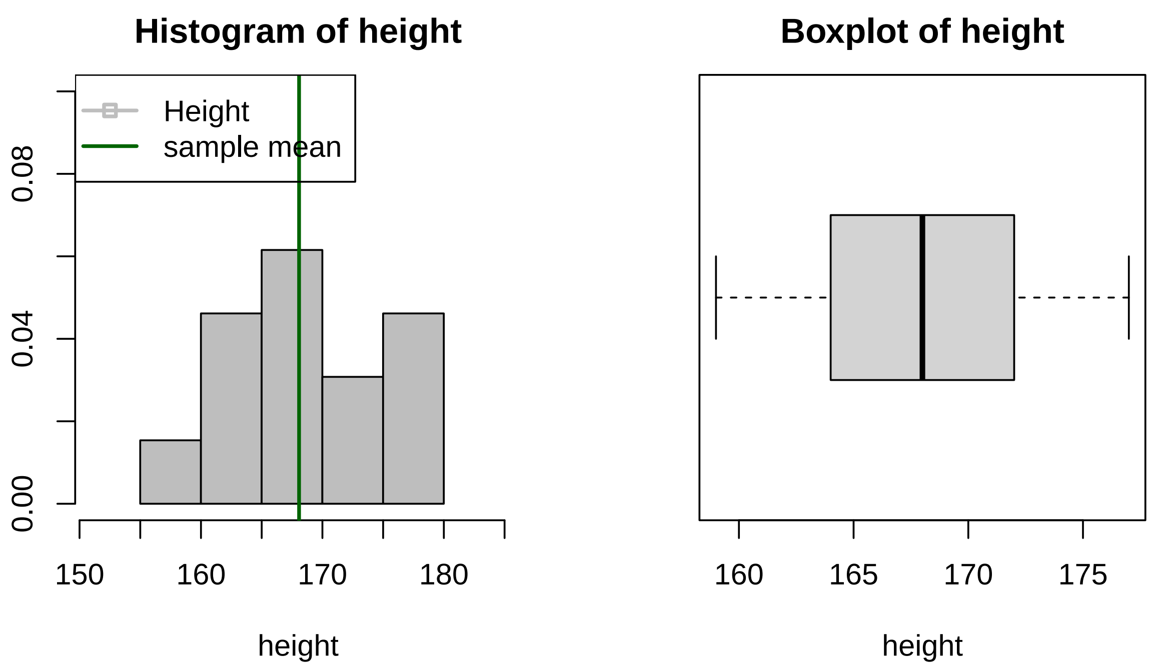

Figure 1: Descriptive graphs for the `Height` outcome.

```
## [1] 144.6018 161.0000 177.3982

 #P[m > 200]
prior_200 <- 1-pnorm(200, mean = mean_prior, sd = sd_prior)
formatC(prior_200, format = "e", digits = 2)

## [1] "1.57e-06"

###(c)
(mean_posterior <- (kappa*n*y_bar + mu*lambda)/(kappa*n+lambda))

## [1] 164.558

(sd_posterior <- sqrt((kappa*n+lambda)^(-1)))

## [1] 5.899714
```

```
(quantiles_posterior <- qnorm(c(alpha / 2, 0.5, 1 - alpha / 2),
                              mean = mean_posterior,
                              sd = sd_posterior))

## [1] 152.9948 164.5580 176.1212

###(d)
#P[m > 200 \mid y_1, ..., y_n]
posterior_200 <- 1 - pnorm(200, mean = mean_posterior, sd = sd_posterior)
formatC(posterior_200, format = "e", digits = 2)

## [1] "9.43e-10"
```

moments	prior	posterior
mean	161	164.558
sd	8.367	5.9
median	161	164.558
CrI.low	144.602	152.995
CrI.up	177.398	176.121
$m > 200$	1.57e-06	9.43e-10

Table 1: Comparison of the moments of m and the probability of $m > 200$ under prior and posterior distributions.

Table 1 shows that posterior mean and median have increased after seeing the data. Moreover, the posterior standard deviation has decreased and the credible intervals became narrower. The interpretation of Bayesian credible intervals differ from that of the classical

```
par(pty = "s")
curve(dnorm(x, mean = mu, sd = sqrt((lambda^(-1)))),
      xlab = expression(m),
      ylab = "density",
      from = 140, to = 190, lwd = 2, col = "black", ylim = c(0,0.1))
curve(dnorm(x,
            mean = mean_posterior,
            sd = sd_posterior),
      from = 140, to = 190, lwd = 2, add = TRUE, col = "red")
legend("topleft", legend = c("prior", " posterior"), col = c("black", "red"), lwd = 2)
```

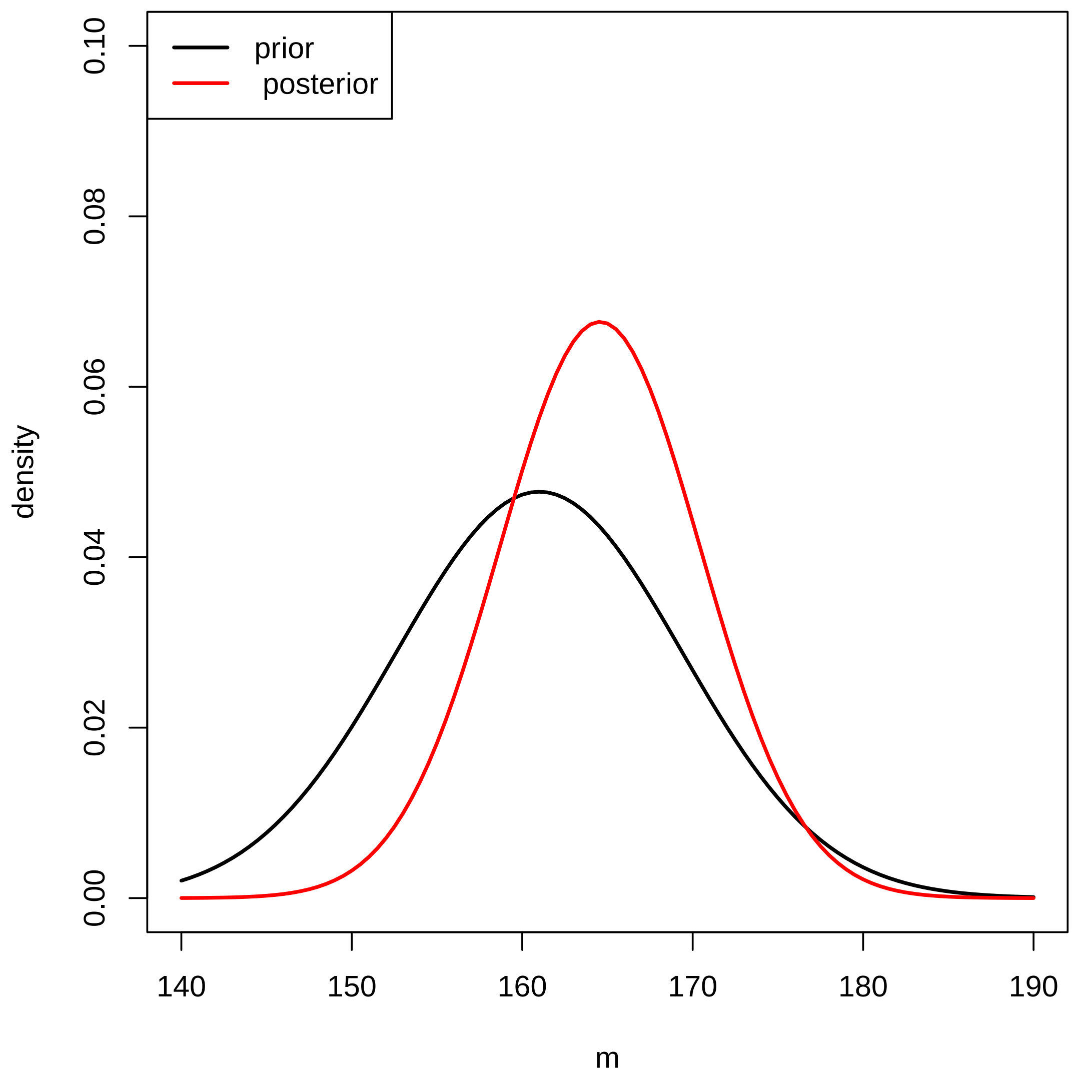

Figure 2: Prior and Posterior distributions

confidence intervals. We can say that the posterior m lies between 152.995 and 176.121 with probability 95% when prior $\mathcal{N}(161, 70)$ is assumed. Although the mean under the posterior is shifted to the right, the probability of m being larger than 200 is smaller under the posterior because the spread is reduced.

Exercise 5 (Bayesian learning - 10 Points)

```
x_interim <- 3
n_interim <- 12
x_final <- 14
n_final <- 64
prior1_shape1 <- prior1_shape2 <- 0.5
prior2_shape1 <- 8
prior2_shape2 <- 24

posterior_shape1 <- function(alpha, beta, n_y_bar) {
   alpha+n_y_bar
}
posterior_shape2 <- function(alpha, beta, n, n_y_bar) {
   beta + n - n_y_bar
}

(post_prior1_interim_shape1 <-
      posterior_shape1(alpha = prior1_shape1 ,
                       beta=prior1_shape2, n_y_bar = x_interim))

## [1] 3.5

(post_prior1_interim_shape2 <-
      posterior_shape2(alpha = prior1_shape1 ,
                       beta=prior1_shape2, n = n_interim, n_y_bar = x_interim))

## [1] 9.5

(post_prior1_final_shape1 <-
      posterior_shape1(alpha = prior1_shape1 ,
                       beta=prior1_shape2, n_y_bar = x_final))

## [1] 14.5
```

```
(post_prior1_final_shape2 <-
      posterior_shape2(alpha = prior1_shape1 ,
                       beta=prior1_shape2, n = n_final, n_y_bar = x_final))

## [1] 50.5

(post_prior2_interim_shape1 <-
      posterior_shape1(alpha = prior2_shape1 ,
                       beta=prior2_shape2, n_y_bar = x_interim))

## [1] 11

(post_prior2_interim_shape2 <-
      posterior_shape2(alpha = prior2_shape1 ,
                       beta=prior2_shape2, n = n_interim, n_y_bar = x_interim))

## [1] 33

(post_prior2_final_shape1 <-
      posterior_shape1(alpha = prior2_shape1 ,
                       beta=prior2_shape2, n_y_bar = x_final))

## [1] 22

(post_prior2_final_shape2 <-
      posterior_shape2(alpha = prior2_shape1 ,
                       beta=prior2_shape2, n = n_final, n_y_bar = x_final))

## [1] 74
```

The posterior mean is $\frac{\alpha_{post}}{\alpha_{post}+\beta_{post}}$. The posterior at interim analysis with prior Beta(0.5, 0.5) is Beta(3.5, 9.5). Therefore the mean for posterior distribution at this stage is 0.269.

```
#prior level - before seeing any data
1 - pbeta(0.4,  shape1 = 0.5, shape2 = 0.5) #for prior: Beta(0.5, 0.5)

## [1] 0.5640942

1 - pbeta(0.4,  shape1 = 8, shape2 = 24) #for prior: Beta(8, 24)

## [1] 0.03298768

#interim analysis with prior Beta(0.5, 0.5)
#mean, median and 95%CrI
3.5 / (3.5 + 9.5) #mean for posterior Beta(3.5, 9.5)
```

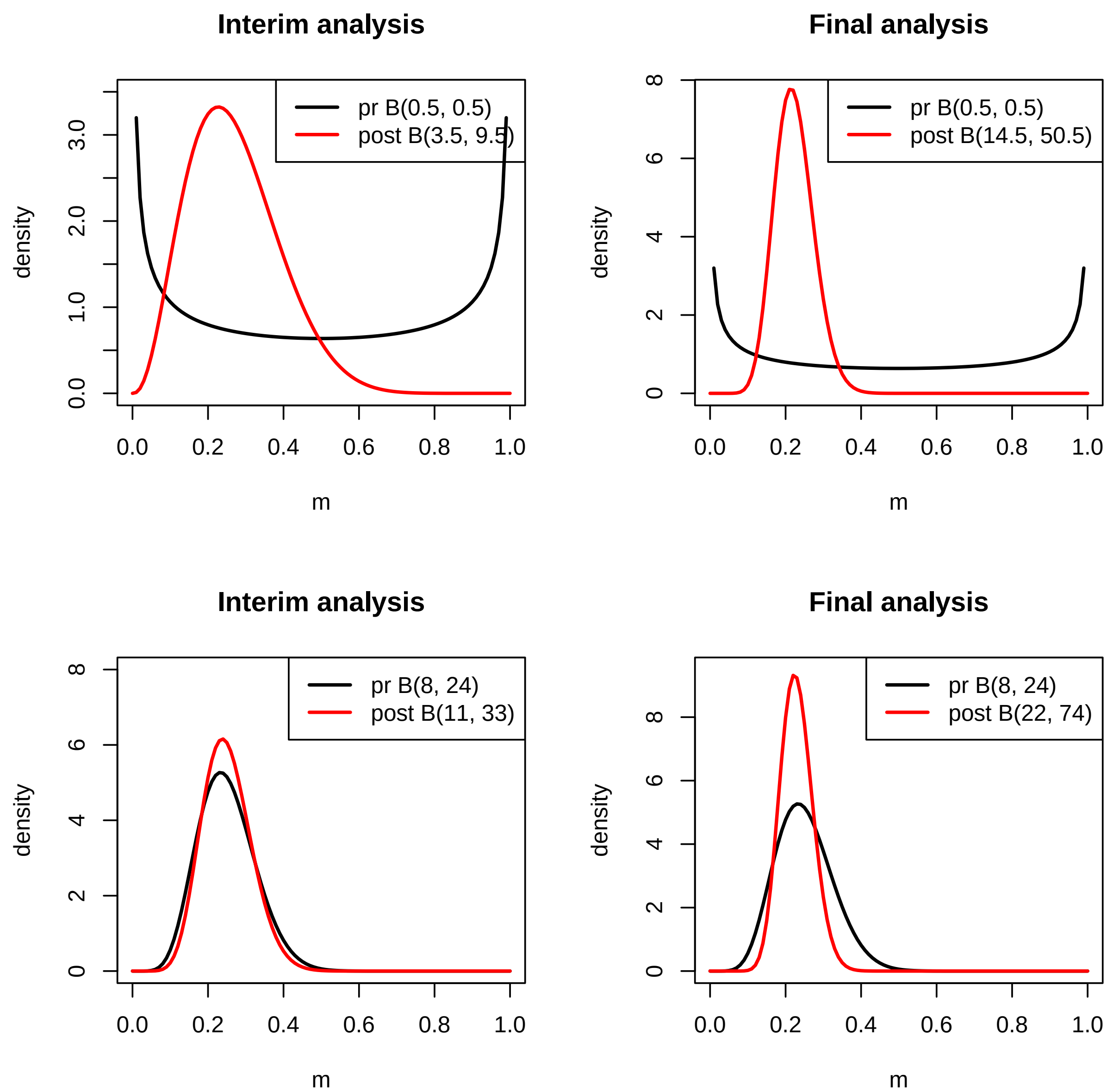

Figure 3: Prior and posterior distributions with two different priors at interim and final stages of the study.

```
## [1] 0.2692308

qbeta(p = c(0.025, 0.5, 0.975), shape1 = 3.5, shape2 = 9.5)

## [1] 0.07594233 0.25711895 0.52919108

1 - pbeta(0.4, shape1 = 3.5, shape2 = 9.5)

## [1] 0.1437649

#interim analysis with prior Beta(8,24)
11 / (11 + 33) #mean for posterior Beta(11, 33)

## [1] 0.25

qbeta(p = c(0.025, 0.5, 0.975), shape1 = 11, shape2 = 33)

## [1] 0.1351860 0.2461854 0.3863082

1 - pbeta(0.4,  shape1 = 11, shape2 = 33)

## [1] 0.01621346

#final analysis with prior Beta(0.5, 0.5)
14.5 / (14.5 + 50.5) #mean for posterior Beta(14.5, 50.5)

## [1] 0.2230769

qbeta(p = c(0.025, 0.5, 0.975), shape1 = 14.5, shape2 = 50.5)

## [1] 0.1312669 0.2202242 0.3310055

1-pbeta(0.4,  shape1 = 14.5, shape2 = 50.5)

## [1] 0.001075757

#final analysis with prior Beta(8,24)
22 / (22 + 74) #mean for posterior Beta(22, 74)

## [1] 0.2291667

qbeta(p = c(0.025, 0.5, 0.975), shape1 = 22, shape2 = 74)

## [1] 0.1511774 0.2272801 0.3178360

1 - pbeta(0.4, shape1 = 22, shape2 = 74)

## [1] 0.0001727695
```

Let's discuss the difference of the impacts of the assumed priors on the posterior at interim and final stages of the analysis. Figure 4 shows that as the sample size increases in the final stage of analysis, the posterior distributions become independent of the prior assumed.

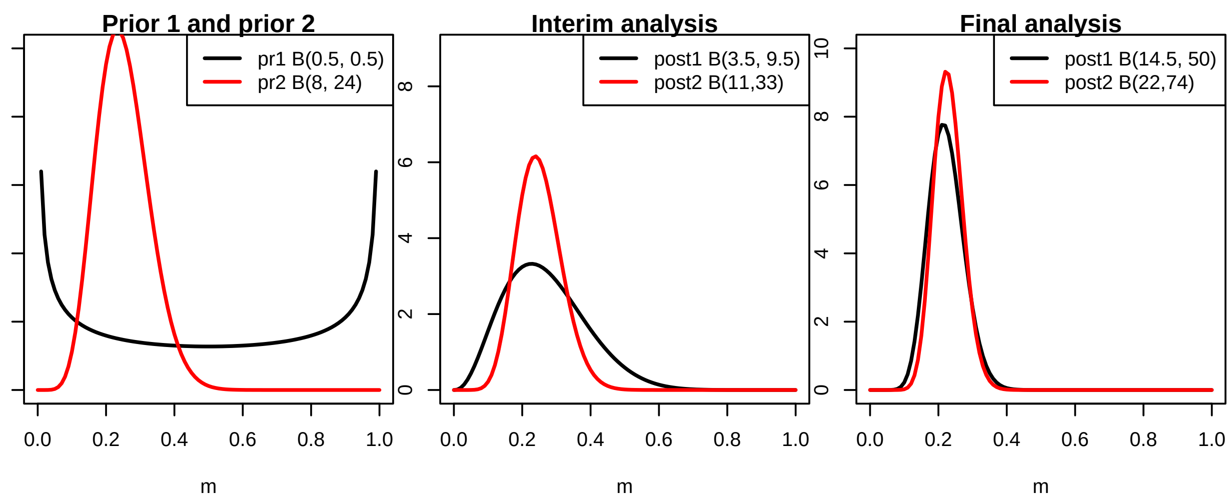

Figure 4: Two different priors are assumed (left). The posterior distributions at interim analysis with two different priors (middle). The posterior distributions at final analysis with two different priors (right).