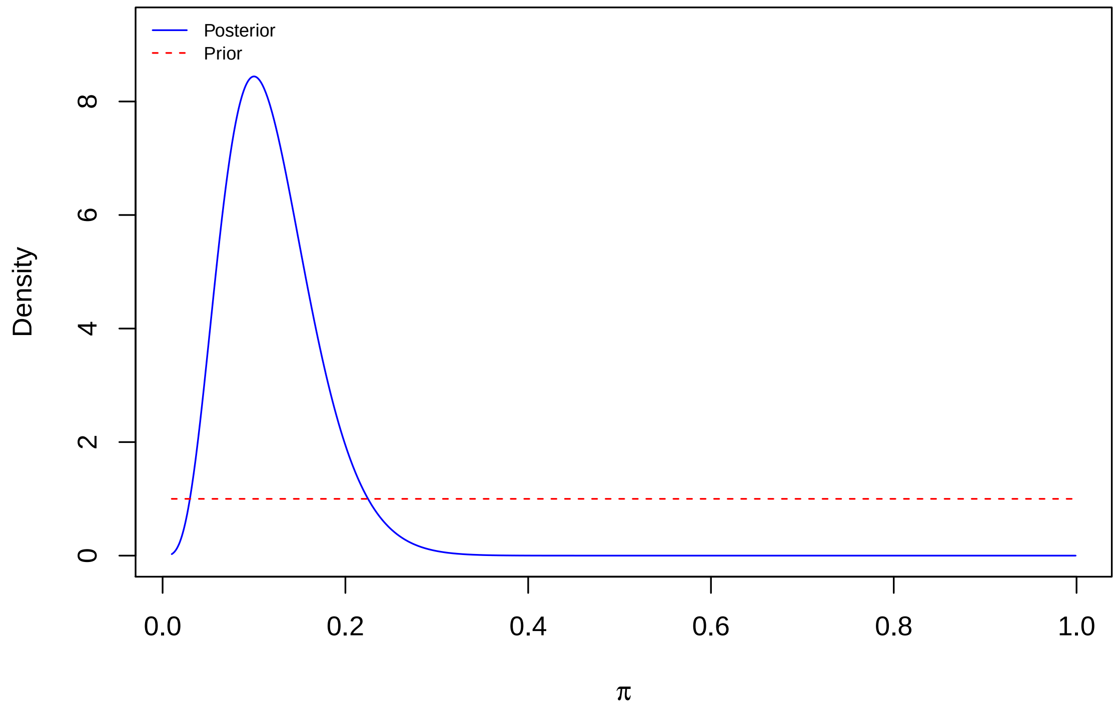
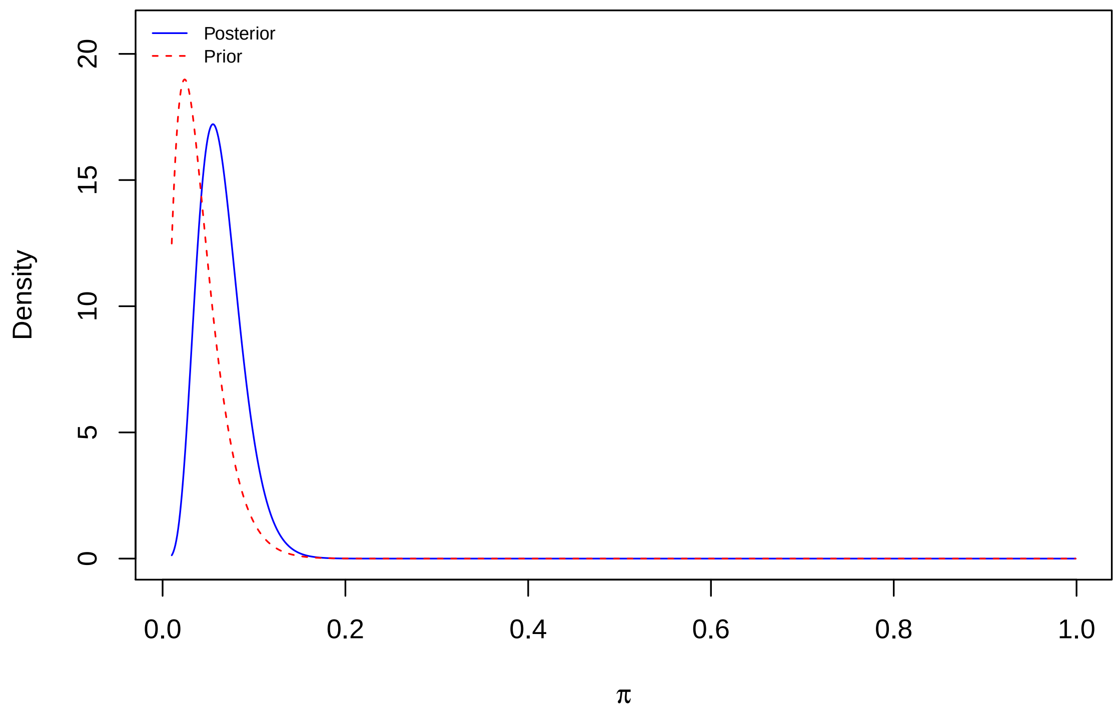
MAST20005/MAST90058: Week 11 Lab Solutions

1. The observed data are $y = 4$ out of a sample size of $n = 40$. The uniform prior gives the posterior $\theta \mid y \sim \text{Beta}(5, 37)$. The informative prior is $\theta \sim \text{Beta}(2.4, 57.6)$, which gives the posterior $\theta \mid y \sim \text{Beta}(6.4, 93.6)$.

   (a)
```r
library(Bolstad)
par(mar = c(4, 4, 1, 1))
binobp(4, 40)   # using uniform prior
```




```r
binobp(4, 40, 2.4, 57.6)   # using informative prior
```



(b)
```r
qbeta(c(0.025, 0.975), 5, 37)  # using uniform prior
```

```
## [1] 0.04080673 0.23131455
```

```r
qbeta(c(0.025, 0.975), 6.4, 93.6)  # using informative prior
```

```
## [1] 0.02505801 0.11930557
```

(c) We need to calculate $\Pr(\theta > 0.03 \mid y)$, which is a straightforward calculation using the cdf of the posterior.

```r
pbeta(0.03, 1, 1, lower.tail = FALSE)  # using uniform prior
```

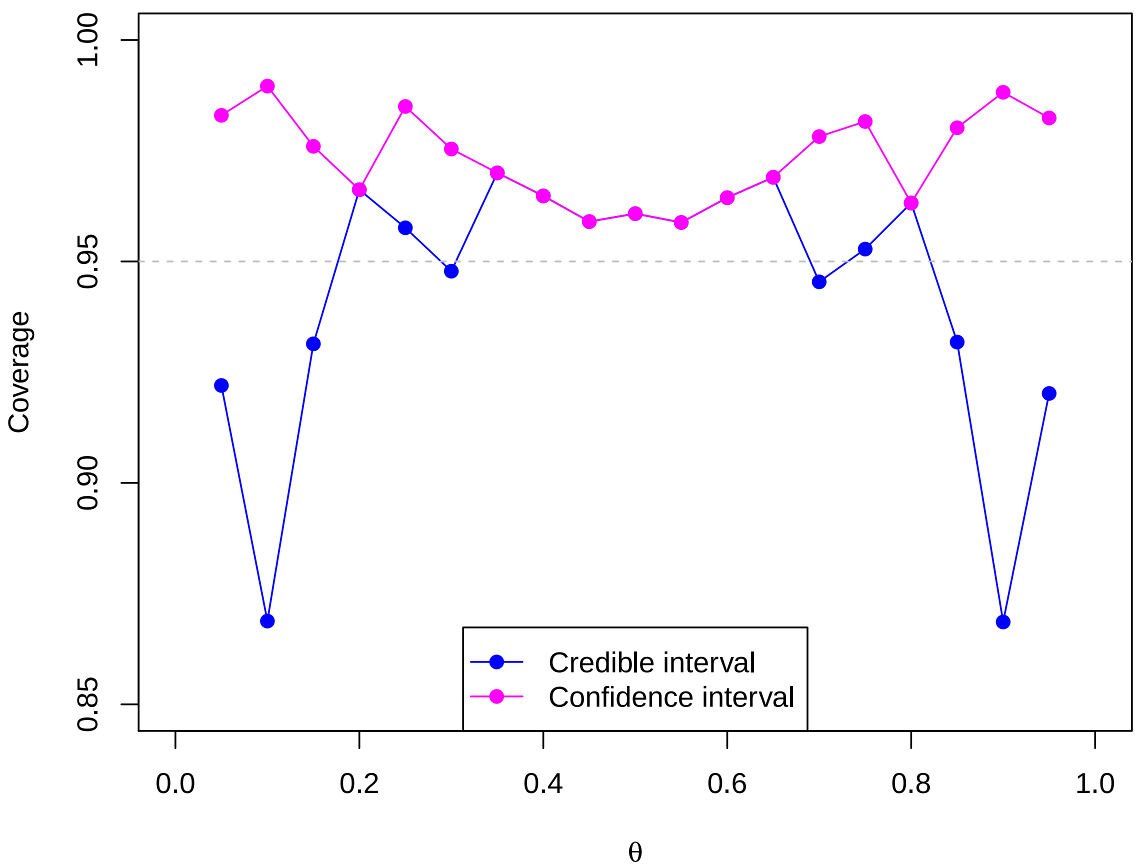
```
## [1] 0.97
```

```r
pbeta(0.03, 6.4, 93.6, lower.tail = FALSE)  # using informative prior
```

```
## [1] 0.9461301
```

2.
```r
thetas <- seq(0.05, 0.95, 0.05) # true parameter values
n <- 20                         # sample size
a <- 2                          # prior parameter (alpha)
b <- 2                          # prior parameter (beta)
nsim <- 5000                    # number of simulations
coverage.cred <- numeric(length(thetas))
coverage.conf <- numeric(length(thetas))
width.cred    <- numeric(length(thetas))
width.conf    <- numeric(length(thetas))
for (i in seq_along(thetas)) {
    # Set parameters and simulate data.
    theta <- thetas[i]  # true parameter value for current simulation
    y <- rbinom(nsim, size = n, prob = theta)

    # Calculate limits for credible interval.
    l1 <- qbeta(0.025, a + y, b + n - y)
    u1 <- qbeta(0.975, a + y, b + n - y)

    # Calculate limits for confidence interval.
    p.tilde <- (y + 2) / (n + 4)
    se <- sqrt(p.tilde * (1 - p.tilde) / n)
    l2 <- p.tilde - 1.96 * se
    u2 <- p.tilde + 1.96 * se

    # Calculate coverage probabilities.
    coverage.cred[i] <- mean(l1 < theta & theta < u1)
    coverage.conf[i] <- mean(l2 < theta & theta < u2)

    # Calculate interval widths.
    width.cred[i] <- mean(u1 - l1)
    width.conf[i] <- mean(u2 - l2)
}
```

```r
plot(thetas, coverage.cred, type = "o", pch = 19, col = "blue",
     xlim = c(0   , 1), xlab = expression(theta),
     ylim = c(0.85, 1), ylab = "Coverage")
lines(thetas, coverage.conf, type = "o", pch = 19, col = "magenta")
abline(h = 0.95, lty = 2, col = "grey")
legend("bottom", c("Credible interval", "Confidence interval"),
       lty = 1, pch = 19, col = c("blue", "magenta"))
```



The coverage of the credible interval is higher when the true value of θ is towards the middle (between 0.2 and 0.8), which is as expected since this is where the prior places more weight. The confidence interval has more consistent coverage across different values of θ and the coverage is generally higher than our nominal confidence level of 95%, which means we could make the interval narrower to get the coverage closer to 95%.

Overall, the coverage of the confidence interval is similar or greater than for the credible interval. However, it is also on average a wider interval, as shown in the plot on the next page. If we wanted a 'fairer' comparison of the coverage, we would want to first modify the intervals so they have the same average width.

```r
plot(thetas, width.cred, type = "o", pch = 19, col = "blue",
     xlim = c(0  , 1  ), xlab = expression(theta),
     ylim = c(0.2, 0.5), ylab = "Interval width")
lines(thetas, width.conf, type = "o", pch = 19, col = "magenta")
abline(h = 0.95, lty = 2, col = "grey")
legend("bottom", c("Credible interval", "Confidence interval"),
       lty = 1, pch = 19, col = c("blue", "magenta"))
```