QPM II - Problem Set 4

2024-11-13

1) The Negative Binomial Distribution

a. Use the integrate function in R to find the Expected A Posteriori estimate of θ. Use a β distribution for the prior with parameters $\alpha = \beta = 1$.

```r
# Given parameters
x <- c(23, 14, 24, 17, 4, 40, 17, 13, 31, 24)
s <- 1
alpha <- 1
beta <- 1

# Likelihood f(n)
likelihood <- function(theta) {
  suppressWarnings(prod(choose(x + s - 1, x) * theta^s * (1 - theta)^x))
}

# Integrals
num_posterior_integral <- integrate(function(theta) theta * dbeta(theta, alpha, beta) *
                                      likelihood(theta), 0, 1)
denom_posteriorintegral <- integrate(function(theta) dbeta(theta, alpha, beta) *
                                       likelihood(theta), 0, 1)

# EAP Estimate
E_theta <- num_posterior_integral$value / denom_posteriorintegral$value
cat("E(theta) =", E_theta, "\n")
```

```
## E(theta) = 0.5
```

b. Now execute a similar method to find the posterior variance of θ.

The posterior variance is given by $\mathrm{Var}(\theta) = \mathrm{E}(\theta^2)$ - $\mathrm{E}(\theta))$^2. We have already calculated $\mathrm{E}(\theta)$, therefore we need to calculate $\mathrm{E}(\theta^2)$. The parameter is denoted by α.

```r
# To calculate posterior variance, just square theta.
numerator <- integrate(function(theta) theta^2 *
                         dbeta(theta, alpha, beta) *
                         likelihood(theta), 0, 1)

# Denom is consistent
E_theta_squared <- numerator$value / denom_posteriorintegral$value

posterior_variance <- E_theta_squared - E_theta^2

cat("E(theta^2) =", E_theta_squared, "\n")
```

```
## E(theta^2) = 0.3333333
```

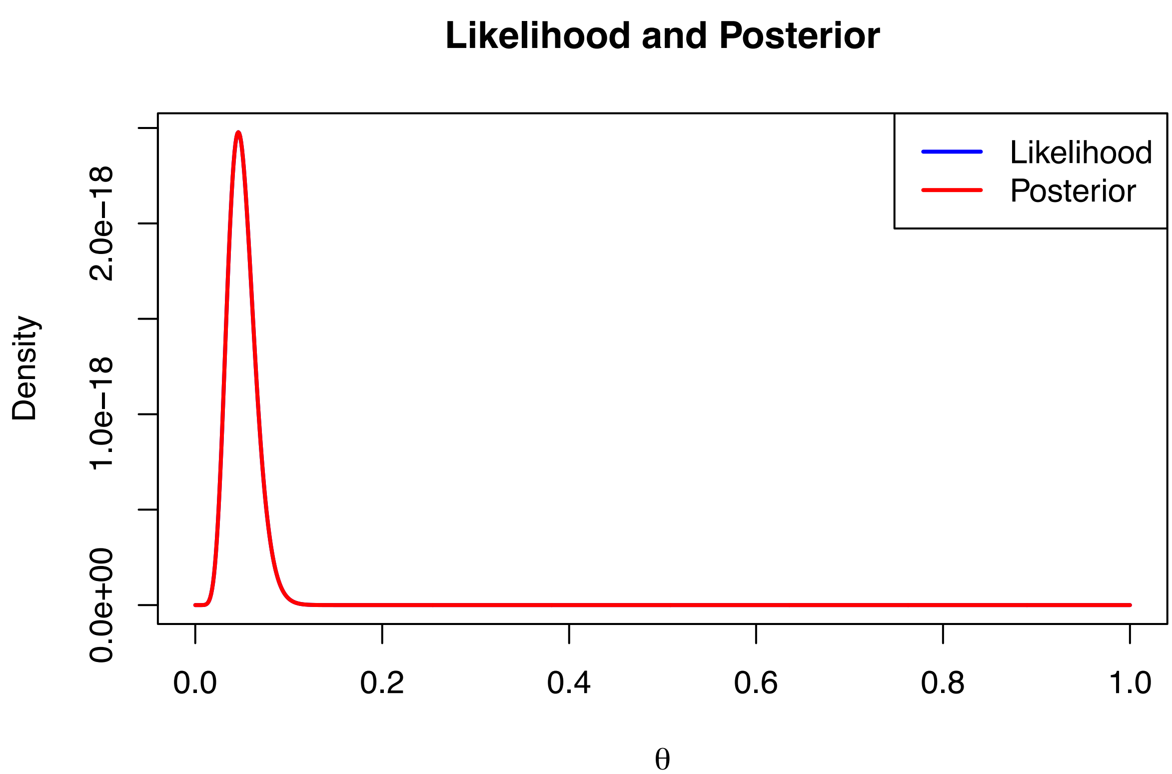
```r
cat("Posterior Variance =", posterior_variance, "\n")
```

```
## Posterior Variance = 0.08333333
```

c. Now plot the likelihood function and the posterior and discuss how they differ.

```r
theta_values <- seq(0, 1, length.out = 1000)
likelihood_values <- sapply(theta_values, likelihood)
posterior_values <- dbeta(theta_values, alpha, beta) * likelihood_values

plot(theta_values, likelihood_values, type = "l", col = "blue",
     lwd = 2, ylab = "Density", xlab = expression(theta),
     main = "Likelihood and Posterior")
lines(theta_values, posterior_values, col = "red", lwd = 2)
legend("topright", legend = c("Likelihood", "Posterior"), col = c("blue", "red"), lwd = 2)
```



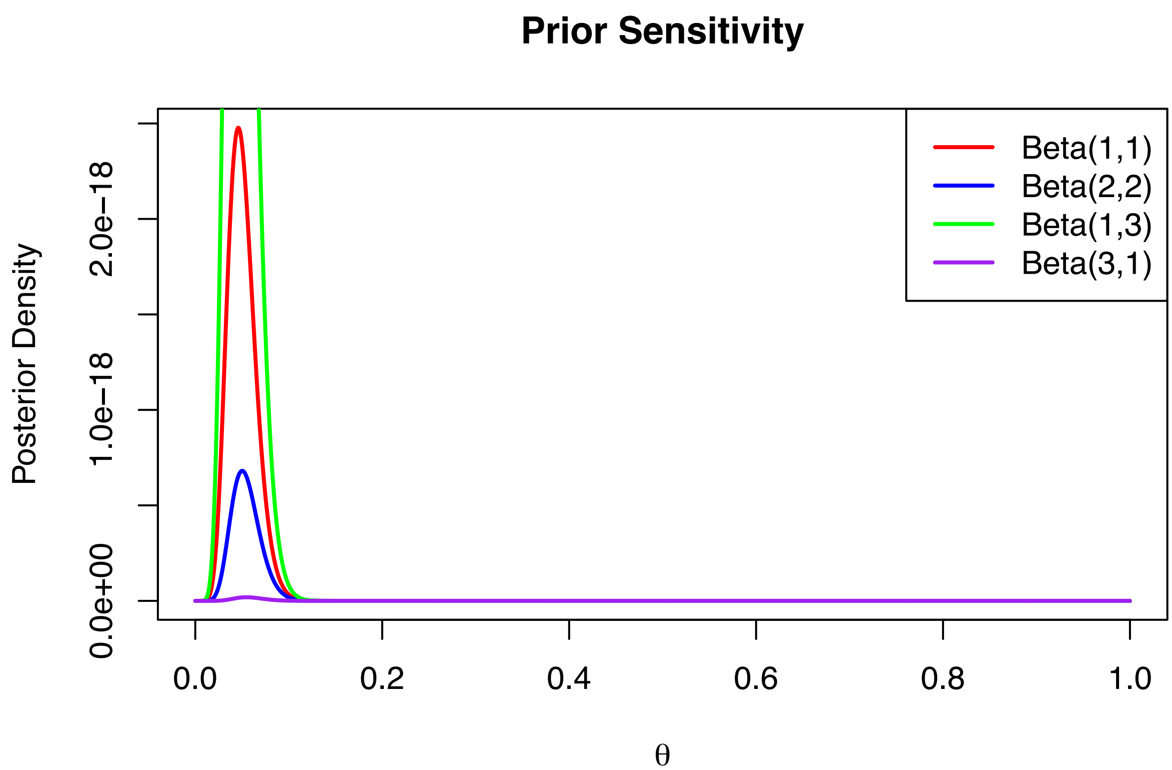
The plot is overlapping. They do not differ.

d. Show that these results are sensitive to choices of α and β. Plot different posteriors, calculate the posterior mean using the numerical method above, and explain the changes.

```r
# Different priors
prior_params <- list(c(1, 1), c(2, 2), c(1, 3), c(3, 1))

plot(NULL, xlim = c(0, 1), ylim = c(0, max(posterior_values)),
     xlab = expression(theta), ylab = "Posterior Density", main = "Prior Sensitivity")

colors <- c("red", "blue", "green", "purple")
for (i in seq_along(prior_params)) {
  alpha_i <- prior_params[[i]][1]
  beta_i <- prior_params[[i]][2]
  posterior_i <- dbeta(theta_values, alpha_i, beta_i) * likelihood_values
  lines(theta_values, posterior_i, col = colors[i], lwd = 2)
}
legend("topright",
       legend = c("Beta(1,1)", "Beta(2,2)", "Beta(1,3)", "Beta(3,1)"),
       col = colors, lwd = 2)
```



Changes in the prior parameters shift the weight of the prior, affecting the posterior mean. A stronger prior like Beta(1,3) centers the distribution more along the mean of theta. Weaker distributions aren't as extremely tight around the mean.

e. Perform a non-parametric bootstrap to find the standard error of the MLE

```r
mle_theta <- function(data) {
  mean(data / (data + 1))
```

```r
}

# Bootstrap
iterations <- 1000
bootstrap_mle <- replicate(iterations, mle_theta(sample(x, replace = TRUE)))

# SE of MLE
bootstrap_se <- sd(bootstrap_mle)
cat("Booststrap Standard Error of MLE:", bootstrap_se)
```

```
## Booststrap Standard Error of MLE: 0.01544707
```

Problem Set 4

Dan Char

2) One Sample T-test

$$f(x_i \mid \theta, \theta^2) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left(-\frac{(x_i-\theta_0)^2}{2\theta^2}\right)$$

$$\mathcal{L} = \prod^n \frac{1}{\sqrt{2\pi\theta^2}} \exp\left(-\frac{(x_i-\theta_0)^2}{2\theta^2}\right) = \frac{1}{(2\pi\,\theta^2)^{n/2}} \exp\left(-\frac{1}{2\theta^2}\sum(x_i-\theta_0)^2\right.$$

a) $\log\mathcal{L} = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\theta^2) - \frac{1}{2\theta^2}\sum(x_i-\theta_0)^2$

b) $\partial\ell/\partial\theta^2 = -\frac{n}{2\theta^2} + \frac{1}{2(\theta^2)^2}\sum(x_i-\theta_0)^2 = 0$ $\quad * 2(\theta^2)^2$

$$-n\theta^2 + \sum(x_i-\theta_0)^2 = 0$$

$$\boxed{\theta^{2*} = \frac{1}{n}\sum(x_i-\theta_0)^2}$$

c) $\ell(\theta_0, \theta^{2*} \mid x) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{1}{n}\sum(x_i-\theta_0)^2\right) - \frac{1}{2\left(\frac{1}{n}\cancel{\sum(x_i-\theta_0)^2}\right)}\cancel{\sum(x_i-\theta_0)^2}$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\frac{1}{n}\sum(x_i-\theta_0)^2\right) - \frac{n}{2}$$

$$= -\frac{n}{2}\left(\log(2\pi) + \log\left(\frac{1}{n}\sum(x_i-\theta_0)^2\right) + 1\right)$$

$$= -\frac{n}{2}\log\left(2\pi \cdot \frac{1}{n}\sum(x_i-\theta_0)^2 \cdot e\right)$$

$$= \exp\left[-\frac{n}{2}\log\left(2\pi \cdot \frac{1}{n}\sum(x_i-\theta_0)^2 \cdot e\right)\right]$$

$$= \left(\frac{2\pi}{n}\sum(x_i-\theta_0)^2\right)^{-n/2} e^{-n/2} \quad \checkmark$$

d) $\partial\ell/\partial\theta^2 = \theta^{2*} = \frac{1}{n}\sum(x_i-\bar{x})^2$

$\downarrow$ Same steps as (c). Only difference is that $\theta_0 \to \bar{x}$

$$\sup\{L(\theta,\theta^2)\} = \left(\frac{2\pi}{n}\sum(x_i-\bar{x})^2\right)^{-n/2} e^{-n/2}$$

e) $\lambda = \frac{\sup_{\theta\in\Theta_0} L(\theta,\theta^2)}{\sup_{\theta\in\Theta} L(\theta,\theta^2)} = \frac{\left(\frac{2\pi}{n}\sum(x_i-\theta_0)^2\right)^{-n/2} e^{-n/2}}{\left(\frac{2\pi}{n}\sum(x_i-\bar{x})^2\right)^{-n/2} e^{-n/2}}$

$$\lambda = \left(\frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\theta_0)^2}\right)^{n/2}$$

f) $$\lambda = \left(\frac{\Sigma(x_i - \bar{x})^2}{\Sigma(x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2} \right)^{n/2}$$

$$\lambda = \left(1 + \frac{\Sigma(x_i - \bar{x})^2}{n(\bar{x} - \theta_0)^2} \right)^{n/2} = \left(1 + \frac{n(\bar{x} - \theta_0)^2}{\Sigma(x_i - \bar{x})^2} \right)^{-n/2}$$

g) $$1 + \frac{n(\bar{x} - \theta_0)^2}{\Sigma(x_i - \bar{x})^2} \leq k^{-2/n}$$

$$\frac{n(\bar{x} - \theta_0)^2}{\Sigma(x_i - \bar{x})^2} \leq k^{-2/n} - 1 \Rightarrow n(\bar{x} - \theta_0)^2 \geq (k^{-2/n} - 1)\Sigma(x_i - \bar{x})^2$$

$$(\bar{x} - \theta_0)^2 \geq \leq \frac{(k^{-2/n} - 1)\Sigma(x_i - \bar{x})^2}{n}$$

$$|\bar{x} - \theta_0| \geq \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}} \cdot \sqrt{(k^{-2/n} - 1)}$$

$$\frac{|\bar{x} - \theta_0|}{\sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}} \geq \sqrt{(k^{-2/n} - 1)}$$

Ho can be rejected when

h) $$S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$$

$$S^2(n-1) = \Sigma(x_i - \bar{x})^2 \Rightarrow \frac{|\bar{x} - \theta_0|}{\sqrt{\frac{S^2(n-1)}{n}}} > c$$

$$= \frac{|\bar{x} - \theta_0|}{S\sqrt{\frac{(n-1)}{n}}} > c$$

$$= \frac{|\bar{x} - \theta_0|}{S/\sqrt{n}} > c \cdot \sqrt{\frac{n-1}{n}}$$

i. Conduct the likelihood ratio test for the null hypotheses using the results above

```
# Given data
data <- c(1, 3, 2, 4, -1, 7, 19, 3, -4, -5, -8)
n <- length(data)
theta0 <- 0
alpha <- 0.10

# T-statistic
x_bar <- mean(data)
S <- sd(data)
t_stat <- (x_bar - theta0) / (S / sqrt(n))

critical_value <- qt(1 - alpha, df = n - 1)

# Perform the hypothesis test
reject_null <- abs(t_stat) > critical_value
cat("Reject H0:", reject_null,"\n", "T-Statistic value:", abs(t_stat), "\n", "Critical Value:", critica
```

```
## Reject H0: FALSE
##  T-Statistic value: 0.8823841
##  Critical Value: 1.372184
```

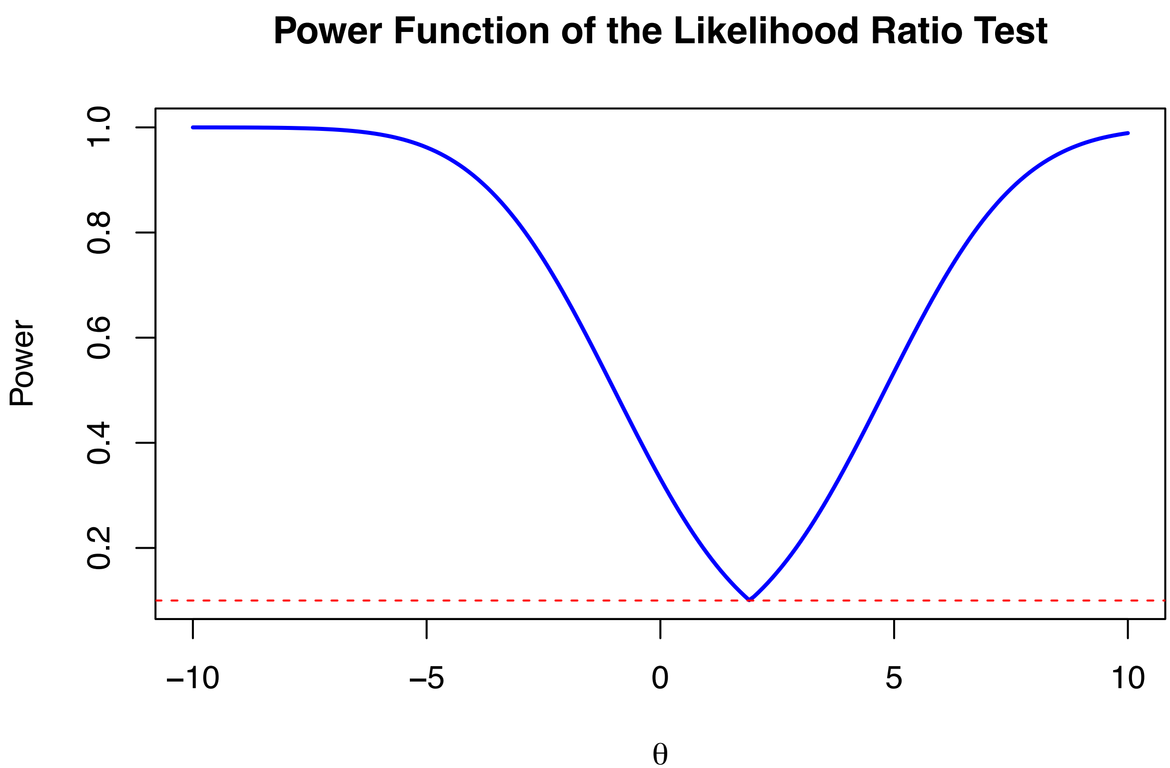
j. Find the power function for this test.

```
# power = Pr(rejecting H0 (t > critical value))
power_function <- function(theta, x_bar, S, n, critical_value) {
  non_central_t <- (x_bar - theta) / (S / sqrt(n))
  power <- pt(critical_value, df = n - 1, ncp = abs(non_central_t), lower.tail = FALSE)
  return(power)
  }
```

k. Plot the power function for this test.

```r
theta_values <- seq(-10, 10, by = 0.1)  # Range of theta values
power_values <- sapply(theta_values, power_function,
                       x_bar = x_bar, S = S, n = n, critical_value = critical_value)

# Plotting of power function with the significance line == 2
plot(theta_values, power_values, type = "l", col = "blue", lwd = 2,
     xlab = expression(theta), ylab = "Power",
     main = "Power Function of the Likelihood Ratio Test")
abline(h = 0.10, col = "red", lty = 2)
```



3) $k' = \mu_0 + z_\alpha \, \theta/\sqrt{n}$

$H_0: \mu = 7$
$H_a: \mu > 7$

$\theta^2 = 5$
$n = 20$

$k' = 7 + 1.645\left(\sqrt{5/20}\right)$
$k' = 7.8225$

a) Reject H_0 if $k' > 7.8225$, Do not reject H_0 otherwise

b) Power $= P(\bar{Y} > 7.8225 \mid \mu = \mu_a) \rightarrow N(\mu_0, 5/n)$

$$Z = \frac{\bar{Y} - \mu_a}{\sqrt{5/n}}$$

$$\text{Power} = \left(\frac{\bar{Y} - \mu_a}{\sqrt{5/n}} > \frac{7.8225 - \mu_a}{\sqrt{5/n}}\right) \qquad \Phi =$$

$$= 1 - \Phi\left(\frac{7.8225 - \mu_a}{\sqrt{5/n}}\right)$$

3) Power Calculations

b + c

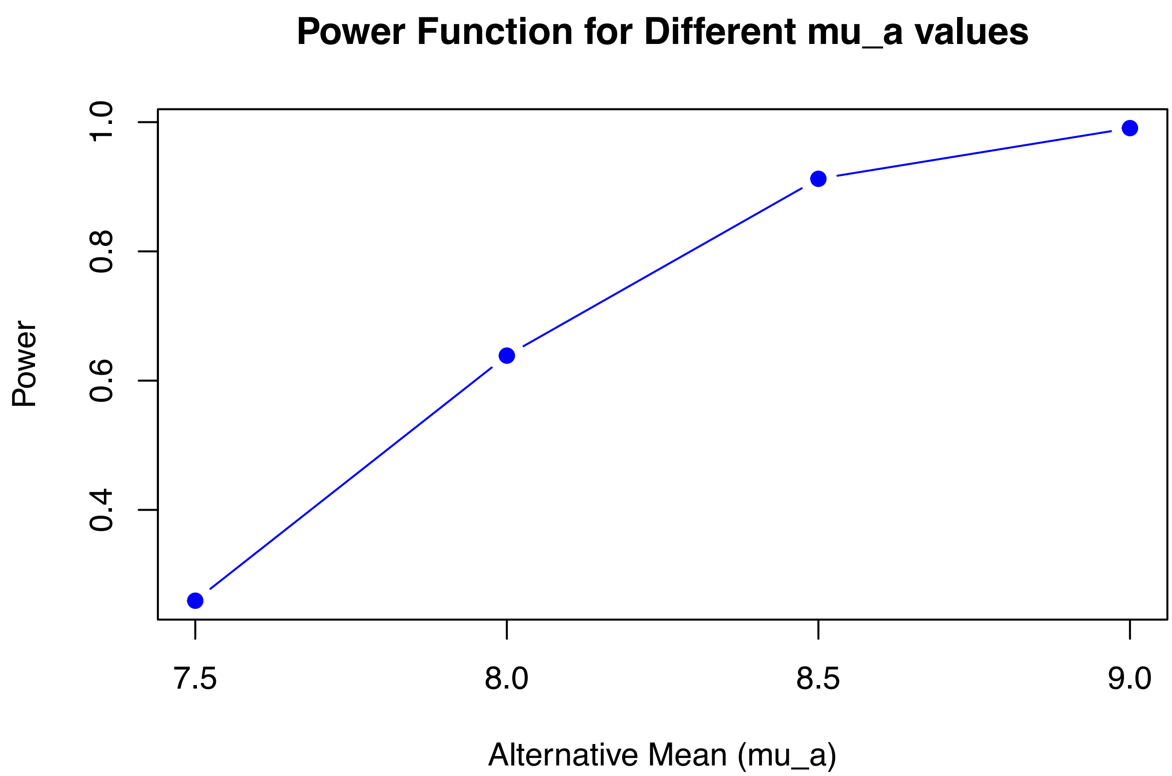
```r
# Power Function Calculation for different mu_a values
mu_0 <- 7
mu_a_values <- c(7.5, 8, 8.5, 9)
n <- 20
critical_value <- 7.8225

power_function <- function(mu_a, critical_value, n) {
  z_score <- (critical_value - mu_a) / sqrt(5 / n)
  power <- 1 - pnorm(z_score)
  return(power)
```

```r
}

# Powers for i - iv
powers <- sapply(mu_a_values, power_function, critical_value = critical_value, n = n)

# Plot
plot(mu_a_values, powers, type = "b", pch = 19, col = "blue",
     xlab = "Alternative Mean (mu_a)", ylab = "Power",
     main = "Power Function for Different mu_a values")
```

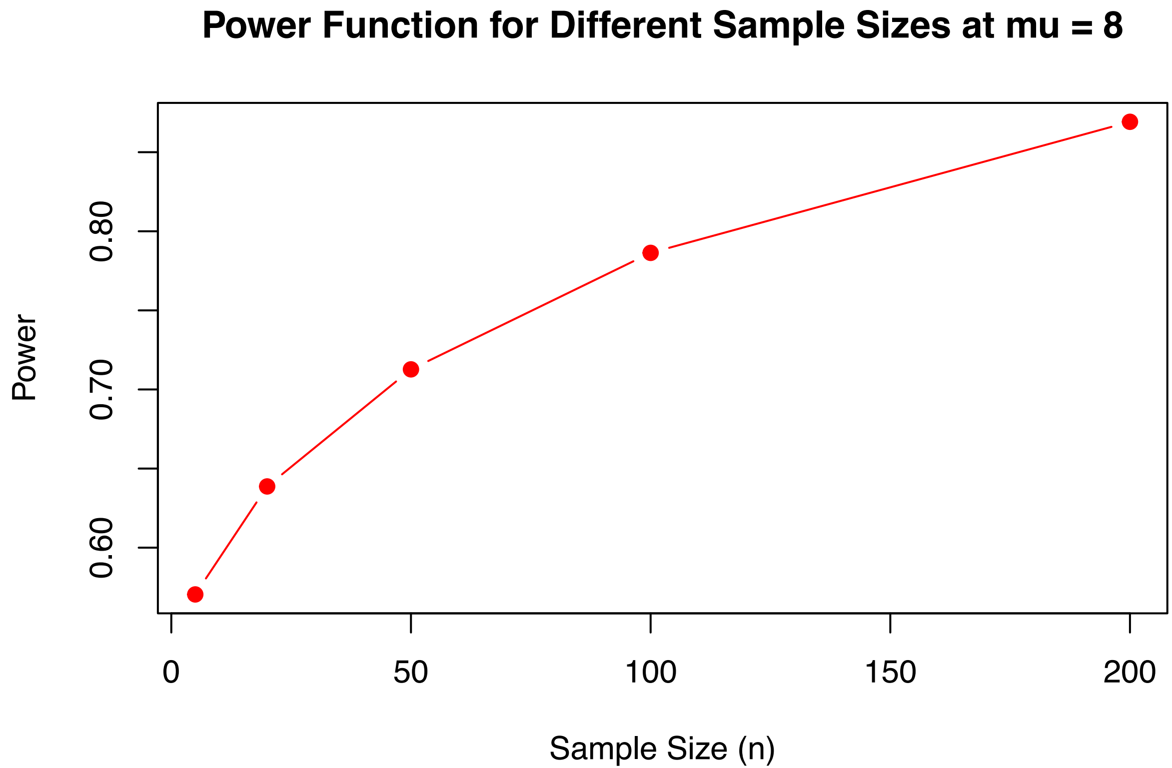


d + e.

```r
mu_a <- 8
sample_sizes <- c(5, 20, 50, 100, 200)

powers_size <- sapply(sample_sizes, power_function, mu_a = mu_a, critical_value = critical_value)

plot(sample_sizes, powers_size, type = "b", pch = 19, col = "red",
     xlab = "Sample Size (n)", ylab = "Power",
     main = "Power Function for Different Sample Sizes at mu = 8")
```


Power Function for Different Sample Sizes at mu = 8
Power
0.60
0.70
0.80
0
50
100
150
200
Sample Size (n)

4) a) False Discovery Rate $= P(H_0 = \text{True} \mid \text{Reject } H_0)$

$$= \frac{P(\text{Reject } H_0 \mid H_0 = \text{True})\, P(H_0 = \text{True})}{P(\text{Reject } H_0)}$$

$P(\text{Reject } H_0 \mid H_0 = \text{True}) = \alpha$

$P(H_0 = \text{True}) = \phi$

$P(\text{Reject } H_0) = \alpha\phi + (1-\beta)(1-\phi)$

$$\text{False Discovery Rate} = \frac{\alpha\phi}{\alpha\phi + (1-\beta)(1-\phi)}$$

4) Simulation b - d.

```r
alpha <- 0.05

false_discovery_rate_fn <- function(alpha, beta, phi) {
  false_discovery_rate <- (alpha * phi) / (alpha * phi + (1 - beta) * (1 - phi))
  return(false_discovery_rate)
}

false_discovery_rate_b <- false_discovery_rate_fn(alpha, beta = 0.75, phi = (1/6))
cat("b. False Discovery Rate:", false_discovery_rate_b, "\n")
```

```
## b. False Discovery Rate: 0.03846154
```

```r
false_discovery_rate_c <- false_discovery_rate_fn(alpha, beta = 0.75, phi = (1/21) )
cat("c. False Discovery Rate:", false_discovery_rate_c, "\n")
```

```
## c. False Discovery Rate: 0.00990099
```

```r
false_discovery_rate_d <- false_discovery_rate_fn(alpha, beta = 0.6, phi = (1/41))
cat("d. False Discovery Rate:", false_discovery_rate_d, "\n")
```

```
## d. False Discovery Rate: 0.003115265
```

5) $H_0: P(\text{Heads}) = 0.5$
$H_a: P(\text{Heads}) \neq 0.5$

a)
Prior : Beta distribution , assume fair $\rightarrow$ Beta(1, 1)

$$L = P(\text{Data} \mid p) = \binom{20}{18} p^{18} (1-p)^2$$

$$\text{Posterior} = \frac{P(\text{Data} \mid p)\, P(p)}{P(\text{Data})}$$

b) $P(p=0.5 \mid \text{Data}) = \binom{20}{18} (0.5)^{18} (0.5)^2$ $\leftarrow$ $P(p=0.5) = 1$, $P(p \neq 0.5) = 0$

$$= 190 (0.5)^{20}$$

$$\boxed{P(p=0.5 \mid \text{Data}) = 0.00018}$$

c) $P(p) = \text{Beta}(p; 1, 1)$

$P(p \mid \text{Data}) = \text{Beta}(p; 1+18, 1+2)$
$= \text{Beta}(p; 19, 3)$

$$\mu = \frac{\alpha}{\alpha+\beta} = \frac{19}{19+3} = \boxed{0.86}$$

posterior probability of coin being biased