Homework 3

STT 465, Bayesian Statistical Methods

Lowell Monis

October 1, 2025

Question 1

Suppose we are going to sample 100 individuals from a county and ask each sampled person whether they support policy Z or not. Let $Y_i = 1$ if person i in the sample supports the policy, and $Y_i = 0$ otherwise.

(a) 6 points

Assume $Y_1, \ldots, Y_{100}$ are conditionally i.i.d. Bernoulli random variables with parameter θ. Write down the joint density $p(y_1, \ldots, y_{100} \mid \theta)$.

Answer

For a single Bernoulli(θ) observation, the pmf is:

$$p(y_i \mid \theta) = \theta^{y_i}(1-\theta)^{1-y_i}, \quad y_i \in \{0, 1\}$$

Because the Y_i are conditionally independent given θ, the joint density is the product of the individual pmfs:

$$p(y_1, \ldots, y_{100} \mid \theta) = \prod_{i=1}^{100} \theta^{y_i}(1-\theta)^{1-y_i} = \theta^{\sum_{i=1}^{100} y_i}(1-\theta)^{100 - \sum_{i=1}^{100} y_i}$$

(b) 6 points

If we set the prior $\theta \sim \text{Beta}(a, b)$, verify that the posterior distribution is

$$\{\theta \mid y_1, \ldots, y_{100}\} \sim \text{Beta}\left(a + \sum_{i=1}^{100} y_i, ; b + 100 - \sum_{i=1}^{100} y_i\right)$$

Answer

The likelihood of the data $\mathbf{y} = (y_1, \ldots, y_{100})$, given θ, is proportional to the Binomial pmf:

$$p(\mathbf{y}|\theta) \propto \theta^S(1-\theta)^{100-S}$$

where $S = \sum_{i=1}^{100} y_i$.

The prior distribution for $\theta \sim \text{Beta}(a, b)$ has a density proportional to its kernel:

$$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$

The posterior distribution $p(\theta|\mathbf{y})$ is proportional to the product of the likelihood and the prior:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta) \cdot p(\theta)$$

Substituting the proportional expressions:

$$p(\theta|\mathbf{y}) \propto \left[\theta^S (1-\theta)^{100-S}\right] \cdot \left[\theta^{a-1}(1-\theta)^{b-1}\right]$$

Combining the terms for θ and $(1-\theta)$:

$$p(\theta|\mathbf{y}) \propto \theta^{(S+a)-1}(1-\theta)^{(100-S+b)-1}$$

This expression matches the kernel of a Beta distribution with parameters:

$$\alpha_{\text{post}} = a + S = a + \sum_{i=1}^{100} y_i$$

$$\beta_{\text{post}} = b + 100 - S = b + 100 - \sum_{i=1}^{100} y_i$$

Therefore, the posterior distribution is:

$$(\theta|y_1, \ldots, y_{100}) \sim \text{Beta}\left(a + \sum_{i=1}^{100} y_i, b + 100 - \sum_{i=1}^{100} y_i\right)$$

(c) 6 points

Suppose the survey result is $\sum_{i=1}^{100} Y_i = 57$ and we use the uniform prior $p(\theta) = 1$ on $[0, 1]$. Calculate the posterior mean and a 95% quantile-based credible interval.

Answer

Uniform prior on $[0, 1]$ is $\text{Beta}(\alpha = 1, \beta = 1)$. With $S = \sum y_i = 57$ and $n = 100$, posterior is

$$\theta \mid y \sim \text{Beta}(\alpha + S, \beta + n - 57) = \text{Beta}(1 + 57, 1 + 100 - 57) = \text{Beta}(58, 44).$$

I can now calculate the posterior mean as:

$$\mathbb{E}[\theta \mid y] = \frac{\alpha}{\alpha + \beta} = \frac{58}{58 + 44} = \frac{58}{102} \approx 0.5686275$$

We will now calculate the credible intervals using $\text{Beta}(58, 44)$.

```
S <- 57
n <- 100
alpha_prior <- 1
beta_prior <- 1
alpha_post <- alpha_prior + S
beta_post <- beta_prior + n - S
post_mean <- alpha_post / (alpha_post + beta_post)
credible_interval <- qbeta(c(0.025, 0.975), alpha_post, beta_post)
```

Thus, for the posterior distribution Beta$(58, 44)$, giving posterior mean 0.5686275, the 95% quantile-based credible interval is approximately $[0.4719136, 0.6627912]$.

(d) 6 points

Repeat (c) but using priors Beta$(1, 3)$ and Beta$(3, 1)$. Compare results to (c) and explain what you observe.

Answer

Consider prior Beta$(1, 3)$. Here, posterior distribution will be Beta$(1 + 57, 3 + 100 - 57) =$ Beta$(58, 46)$. Thus, the posterior mean is 0.5576923.

```
alpha_prior <- 1
beta_prior <- 3
alpha_post <- alpha_prior + S
beta_post <- beta_prior + n - S
post_mean <- alpha_post / (alpha_post + beta_post)
credible_interval <- qbeta(c(0.025, 0.975), alpha_post, beta_post)
```

Thus, for the posterior distribution Beta$(58, 44)$, giving posterior mean 0.5576923, the 95% quantile-based credible interval is approximately $[0.4618439, 0.6514381]$.

Consider prior Beta$(3, 1)$. Here, posterior distribution will be Beta$(3 + 57, 1 + 100 - 57) =$ Beta$(60, 44)$. Thus, the posterior mean is 0.5769231.

```
alpha_prior <- 3
beta_prior <- 1
alpha_post <- alpha_prior + S
beta_post <- beta_prior + n - S
post_mean <- alpha_post / (alpha_post + beta_post)
credible_interval <- qbeta(c(0.025, 0.975), alpha_post, beta_post)
```

Thus, for the posterior distribution Beta$(58, 44)$, giving posterior mean 0.5769231, the 95% quantile-based credible interval is approximately $[0.4812264, 0.6698164]$.

The data is informative enough that the posterior means under all three priors (uniform, Beta$(1, 3)$, Beta$(3, 1)$) are similar (about 0.558 to 0.577, sample proportion being 0.57). The prior shifts the posterior mean slightly: Beta$(1, 3)$ (which favors smaller θ) pulls the posterior mean down a bit towards the distribution mean (0.25); Beta$(3, 1)$ (favoring larger θ) pulls it up a bit towards the distribution mean (0.75). However, this pull is minor. The credible intervals overlap substantially, and the posterior means are all very close to the sample proportion of 0.57. This demonstrates that for a sample size of $n = 100$, the likelihood for the data strongly dominates the prior information. The Beta distribution parameters can be interpreted as effective counts. The priors used here, where $\alpha + \beta$ is small (e.g., 1+3=4 or 3+1=4), have a small effective sample size or concentration. Because this effective size is much smaller than the actual sample size $n = 100$, the likelihood overwhelms the prior, leading to very similar posterior inferences across all three prior choices. The data provides a consensus that cannot be easily overridden by these relatively weak priors.

(e) 6 points

Continue from (c). We are interested in the log-odds $g(\theta) = \log \frac{\theta}{1-\theta}$. Use Monte Carlo to calculate the posterior mean and a 95% quantile-based credible interval for $g(\theta)$. Plot smooth kernel density approximations to the prior and posterior distributions of $g(\theta)$ and explain the differences.

Answer

We continue from (c): posterior distribution is $\theta \sim \text{Beta}(58, 44)$.

We use the Monte Carlo approach:

1. We will draw a large number of samples, say M from $\theta \sim \text{Beta}(58, 44)$.
2. We will then transform them via $g = \log(\theta/(1-\theta))$ (i.e., the logit transform).
3. We will then compute the sample mean of g and the 95% credible interval.
4. To compare prior versus posterior distributions of g, we will sample from the prior (here uniform $\text{Beta}(1, 1)$), transform, and then plot kernel density estimates of prior-g and posterior-g.

```r
set.seed(465)
M <- 100000
post_theta <- rbeta(M, 58, 44)
post_g <- log(post_theta / (1 - post_theta))

post_mean_g <- mean(post_g)
post_ci_g <- quantile(post_g, c(0.025, 0.975))

prior_theta <- rbeta(M, 1, 1) # same as runif(), but in beta form
prior_g <- log(prior_theta / (1 - prior_theta))

prior_mean_g <- mean(prior_g)
prior_ci_g <- quantile(prior_g, c(0.025, 0.975))
```

The Monte Carlo posterior mean for $g(\theta)$ is 0.2791498. The Monte Carlo 95% Credible Interval for $g(\theta)$ is [-0.1150814, 0.6757992]

Now, I generate smooth kernel density approximation plots for both distributions.

```r
plot(density(prior_g),
     main = "Density of log-odds, prior vs posterior", xlab = "g",
     ylim = c(0, max(density(prior_g)$y, density(post_g)$y)), col='red')
lines(density(post_g), col='blue')
legend("topright", legend = c("Prior: Beta(1,1)", "Posterior: Beta(58,44)"),
       col = c("red", "blue"), lty = c(1,1))
```

Density of log–odds, prior vs posterior
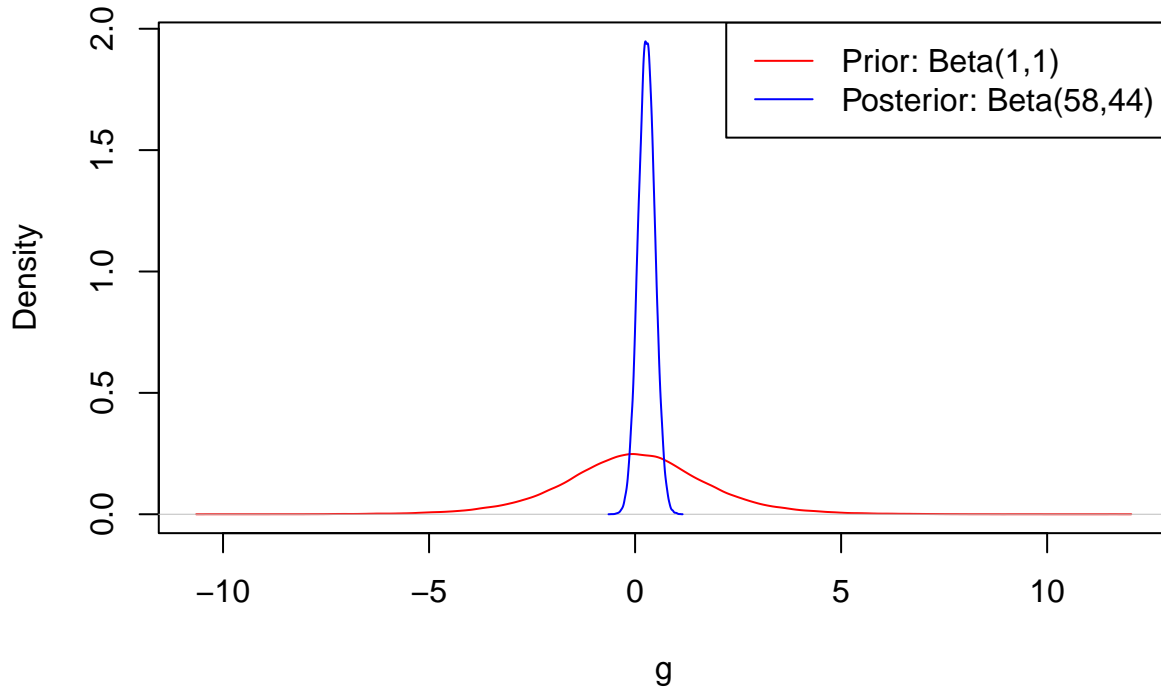
The prior (uniform on θ) becomes a symmetric heavy-tailed distribution for $g = \text{logit}(\theta)$. Because the uniform on $[0, 1]$ places mass near 0 and 1 (substantial probability is over here, like a logistic distribution), the transformed prior for log-odds has heavy tails (logit $\to \pm\infty$ as $\theta \to 0$ or as $\theta \to 1$. The posterior distribution for $g(\theta)$ is concentrated and shifted to the right of 0 (since posterior θ is centered around 0.569). The posterior density for g is narrower and less spread than the prior, reflecting the information gained from the data. The large sample size of $n = 100$ has provided a lot of information, effectively "squeezing" the wide prior distribution into a highly specific posterior distribution, thereby sharply reducing the uncertainty about the log-odds. Essentially, the kernel density plots will show the posterior mass concentrated near $g \approx 0.28$ and the prior much more dispersed and symmetric around 0 (but with heavy tails). The posterior is therefore more informative and shifted to positive values because the observed data favors $\theta > 0.5$, sample proportion being 0.57.

Question 2

A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. The observed tumor counts for these two types of mice are

$$y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6); \quad y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7).$$

(a) 6 points

Assume that type A mice data follows conditionally i.i.d Poisson with parameter θ_A and type B mice data follows conditionally i.i.d Poisson with parameter θ_B. Consider the prior distributions: $\theta_A \sim \text{Gamma}(120, 10)$; $\theta_B \sim \text{Gamma}(12, 1)$; θ_A and θ_B are independent. Plot the posterior distributions for θ_A and θ_B, and compare them.

Answer

Given priors $\theta_A \sim \text{Gamma}(a_A = 120, b_A = 10)\$ $ and $\theta_B \sim \text{Gamma}(a_B = 12, b_B 1)\$$, and $n_A = 10$, $S_A = 117$, $n_B = \text{length}(\text{c}(11,11,10,9,9,8,7,10,6,8,8,9,7))$, $S_B = 113$, I calculate the posterior distributions as follows (in-line code was used to calculate sum and n values):

$$\theta_A \mid y \sim \text{Gamma}(a_A + S_A, b_a + n_a) = \text{Gamma}(120 + 117, 10 + 10) = \text{Gamma}(237, 20)$$

$$\theta_B \mid y \sim \text{Gamma}(a_B + S_B, b_B + n_B) = \text{Gamma}(12 + 113, 1 + 13) = \text{Gamma}(125, 14)$$

I can now plot the distributions.

```
yA <- c(12,9,12,14,13,13,15,8,15,6)
yB <- c(11,11,10,9,9,8,7,10,6,8,8,9,7)

nA <- length(yA)
nB <- length(yB)
sumA <- sum(yA)
sumB <- sum(yB)

aA_prior <- 120
bA_prior <- 10
aB_prior <- 12
bB_prior <- 1
aA_post <- aA_prior + sumA
bA_post <- bA_prior + nA
aB_post <- aB_prior + sumB
bB_post <- bB_prior + nB

x <- seq(0,25,length=400)
plot(x, dgamma(x, shape=aA_post, rate=bA_post), type="l", col="blue", lwd=2,
     ylab="Density", xlab=expression(theta),
     main="Posterior densities")
lines(x, dgamma(x, shape=aB_post, rate=bB_post), col="red", lwd=2)
legend("topright", legend=c("A","B"),
       col=c("blue","red"), lwd=2)
```

Posterior densities
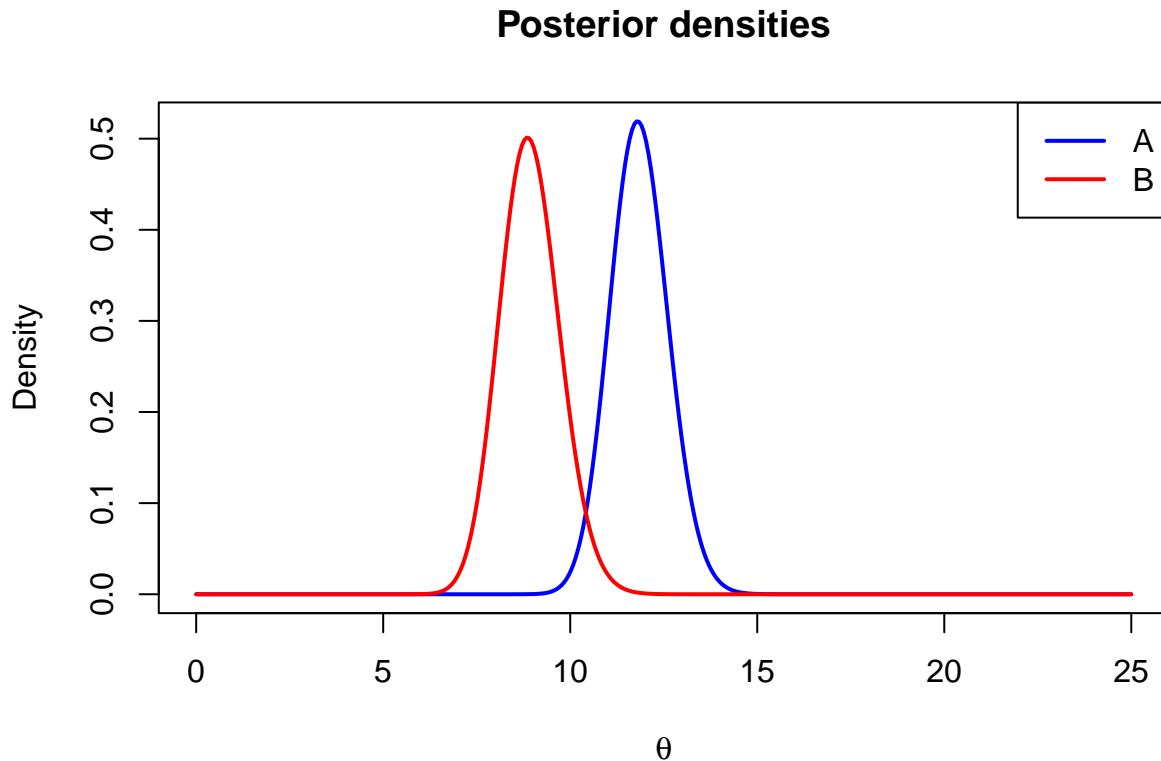
The posterior for θ_A is centered at a higher value of θ than the posterior for θ_B. This matches the raw data: strain A mice had higher tumor counts on average than strain B ($11.7 > 8.6923077$). It looks like B is slightly more spread (verifiable via variances), and both distributions are unimodal and bell-shaped (since shape parameters are large). From the plot, their credibility intervals at 95% confidence may overlap slightly (verifiable by calculating intervals).

The posterior distribution of θ_A is shifted to the right of the posterior distribution of θ_B, reflecting higher tumor counts in strain A. Both distributions are fairly concentrated, with standard deviations less than 1, meaning that there is little overlap between them. This comparison may suggest strong evidence that the expected tumor rate in strain A exceeds that in strain B.

(b) 6 points

Calculate the posterior means, modes and 95% quantile-based credible intervals for θ_A and θ_B, and compare these results for the two types of mice.

Answer

The mean of a gamma distribution is given by shape over rate, or $\frac{a}{b}$. The mode is given by $\frac{a-1}{b}$ for $a > 1$.

```
mean_A <- aA_post / bA_post
mode_A <- (aA_post - 1) / bA_post
ci_A   <- qgamma(c(0.025,0.975), shape=aA_post, rate=bA_post)

mean_B <- aB_post / bB_post
mode_B <- (aB_post - 1) / bB_post
ci_B   <- qgamma(c(0.025,0.975), shape=aB_post, rate=bB_post)
```

7

For strain A, the posterior mean is 11.85, mode is 11.8, and the 95% credible interval is [10.3892382, 13.4054483].

For strain B, the posterior mean is 8.9285714, mode is 8.8571429, and the 95% credible interval is [7.4320642, 10.5603081].

The posterior means and modes show that θ_A is larger than θ_B, with relatively tight credible intervals that overlap only slightly. There is good evidence that generally, strain A has a higher tumor rate.

(c) 6 points

We are now interested in the rate ratio θ_A/θ_B. Using Monte Carlo methods, plot smooth kernel density approximations to the prior and posterior distributions of the ratio. Explain what you observe.

Answer

```
set.seed(465)
M <- 100000

prior_A <- rgamma(M, shape=aA_prior, rate=bA_prior)
prior_B <- rgamma(M, shape=aB_prior, rate=bB_prior)
ratio_prior <- prior_A/prior_B

post_A <- rgamma(M, shape=aA_post, rate=bA_post)
post_B <- rgamma(M, shape=aB_post, rate=bB_post)
ratio_post <- post_A/post_B

plot(density(ratio_prior), xlim=c(0,4), ylim=c(0,3),
     main="Prior vs Posterior density of ratio",
     xlab='ratio', col="blue")
lines(density(ratio_post), col="red")
legend("topright", legend=c("Prior","Posterior"),
       col=c("blue","red"), lty=c(1,1))
```

Prior vs Posterior density of ratio
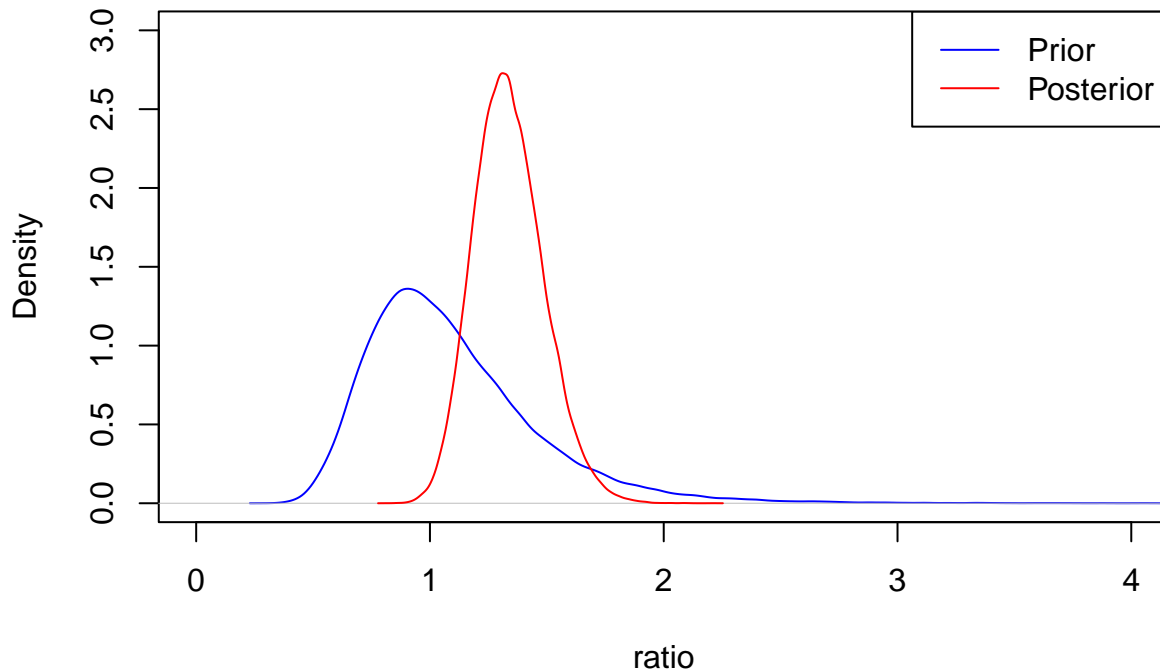
The prior ratio centers around 1, and the posterior ratio is shifted to the right of 1. The prior ratio is broader towards the right, while the posterior ratio is more concentrated. With the ratio being broader after 1 for both prior and posterior, this is further support for the hypothesis that $\theta_A > \theta_B$.

(d) 6 points

Using Monte Carlo methods, calculate the posterior mean and 95% quantile-based credible interval for the ratio θ_A/θ_B. What do you observe?

Answer

We continue with the values from the sampling conducted before.

```
mean_ratio <- mean(ratio_post)
ci_ratio   <- quantile(ratio_post, c(0.025,0.975))
```

The MC posterior mean is 1.3382401 and the 95% quantile-based credible interval for the ratio is [1.0730938, 1.6544168].

The posterior mean indicates that strain A's tumor rate is about 34% higher than strain B on average, and the 95% interval does not include 1, giving substantial evidence that $\theta_A > \theta_B$.

(e) 6 points

Keep the same prior for θ_A and set the prior for θ_B as $\theta_B \sim \text{Gamma}(12 \times n_0, n_0)$. For each value of $n_0 \in \{1, 2, 3, \ldots, 100\}$, use Monte Carlo methods to calculate the posterior probability $P(\theta_A > \theta_B \mid y_A, y_B)$. Describe how sensitive the conclusions about the event $\{\theta_A > \theta_B\}$ are to the prior distribution on θ_B.

Answer

We keep the prior for θ_A fixed and for θ_B, we use $\mathrm{Gamma}(12 \times n_0, n_0)$, with $n_0 \in [100]$. For each n_0, we compute the posterior for θ_B and use MC to estimate $P(\theta_A > \theta_B \mid y)$.

```r
n0_values <- 1:100
p_values <- numeric(length(n0_values))

for (i in seq_along(n0_values)) {
  n0 <- n0_values[i]
  aB_p <- 12 * n0
  bB_p <- n0
  aB_post_n0 <- aB_p + sumB
  bB_post_n0 <- bB_p + nB
  samp_A <- rgamma(5000, shape=aA_post, rate=bA_post)
  samp_B <- rgamma(5000, shape=aB_post_n0, rate=bB_post_n0)
  p_values[i] <- mean(samp_A > samp_B)
}

plot(n0_values, p_values, type="l", lwd=2,
     main="Sensitivity for P(thetaA > thetaB)",
     xlab='n0', ylab="Posterior Probability",
     ylim=c(0,1))
```

Sensitivity for P(thetaA > thetaB)
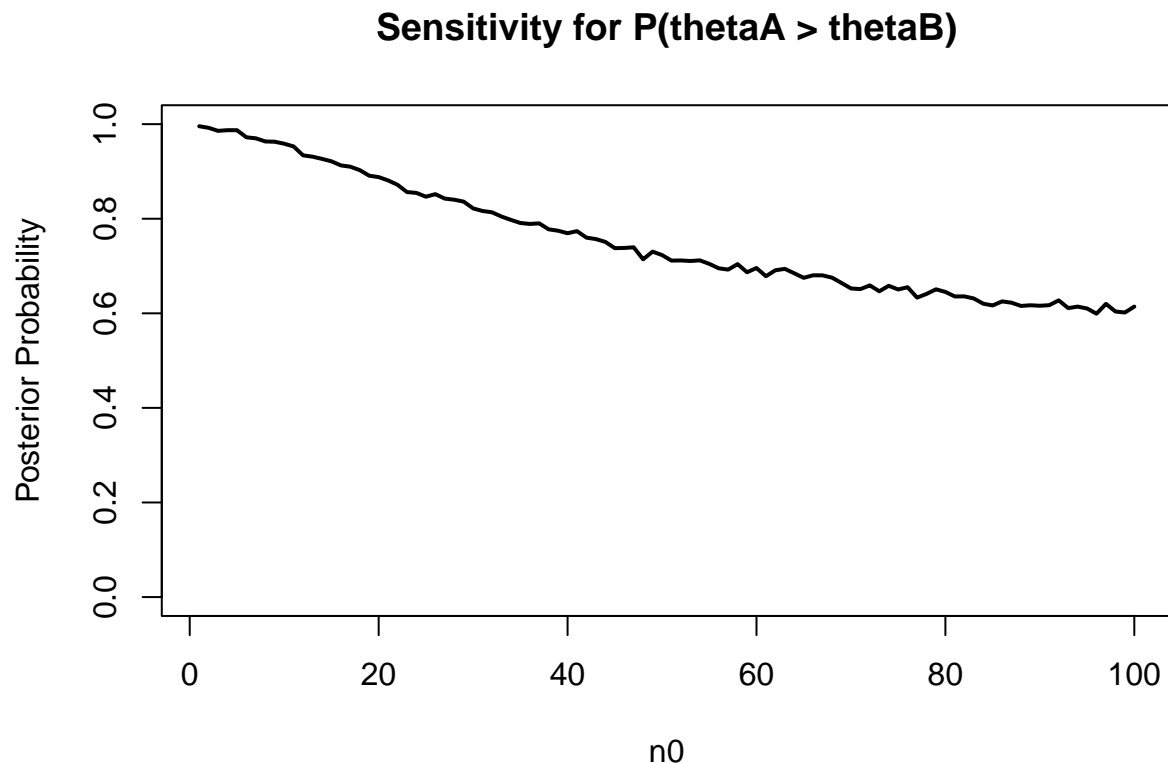
For small n_0, or weak prior, the probability is very high, nearing 99%. As n_0 grows, making the prior for θ_B more concentrated and informative, the posterior probability declines towards roughly 60% for the largest n_0 available to us. The plot shows a roughly monotonic decline; the more concentrated/stronger the prior on θ_B, the more the posterior probability moves towards 50%. But even at larger n_0, the probability still remainds above 50%, so the conclusion is moderately robust, but not completely insensitive to the strength of the prior.

Thus, for weak priors the data dominates and we get very strong evidence for $\theta_A > \theta_B$. As we in crease the effective prior sample size, making the prior stronger, we reduce the posterior probability of this conclusion.