Advanced Macroeconometrics – Assignment 2

Elisabeth Fidrmuc   Miriam Frauenlob   Tim Koenders

May, 2023

Contents

Question 1.

```r
knitr::opts_chunk$set(fig.width=12, fig.height=8, fig_path='figures/',echo=TRUE, warning=FALSE, message

set.seed(123)

# Define parameters
mu <- 5
sigma <- 9
n <- 100

# Generate data
x <- rnorm(n, mean = mu, sd = sigma)

# Compute estimates of the mean with the first 1, ..., n draws
estimates <- cumsum(x) / seq_along(x)

# Plot estimates
plot(estimates, type = "l", col = "blue", xlab = "Sample size", ylab = "Estimate of the mean")

# Add true mean as a horizontal line
abline(h = mu, col = "red")
```
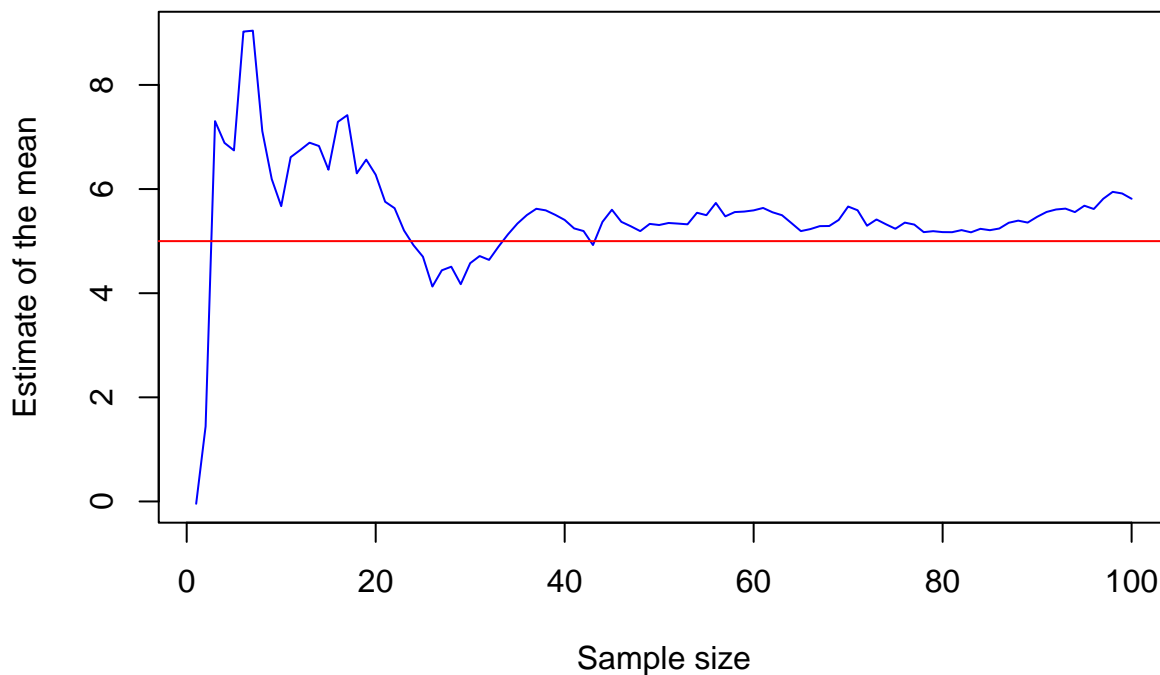
The plot highlights the law of large numbers in action, where increasing the sample size results in estimates of the mean that converge to the true mean. While initial estimates may deviate from the true mean, as the sample size increases, the estimates become more stable and approach the true mean with greater precision.

```r
knitr::opts_chunk$set(fig.width=12, fig.height=8, fig_path='figures/',echo=TRUE, warning=FALSE, message

set.seed(123)

# Define parameters
mu <- 0
sigma <- 1
n <- 1000

# Generate two random samples from a standard normal distribution
x <- rnorm(n, mean = mu, sd = sigma)
y <- rnorm(n, mean = mu, sd = sigma)

# Convert the standard normal samples to a Cauchy distribution with scale one
z <- x / y

# Compute estimates of the mean with the first 1, ..., n draws
estimates <- cumsum(z) / seq_along(z)

# Plot estimates
plot(estimates, type = "l", col = "blue", xlab = "Sample size", ylab = "Estimate of the mean")

# Add true mean as a horizontal line
abline(h = mu, col = "red")
```
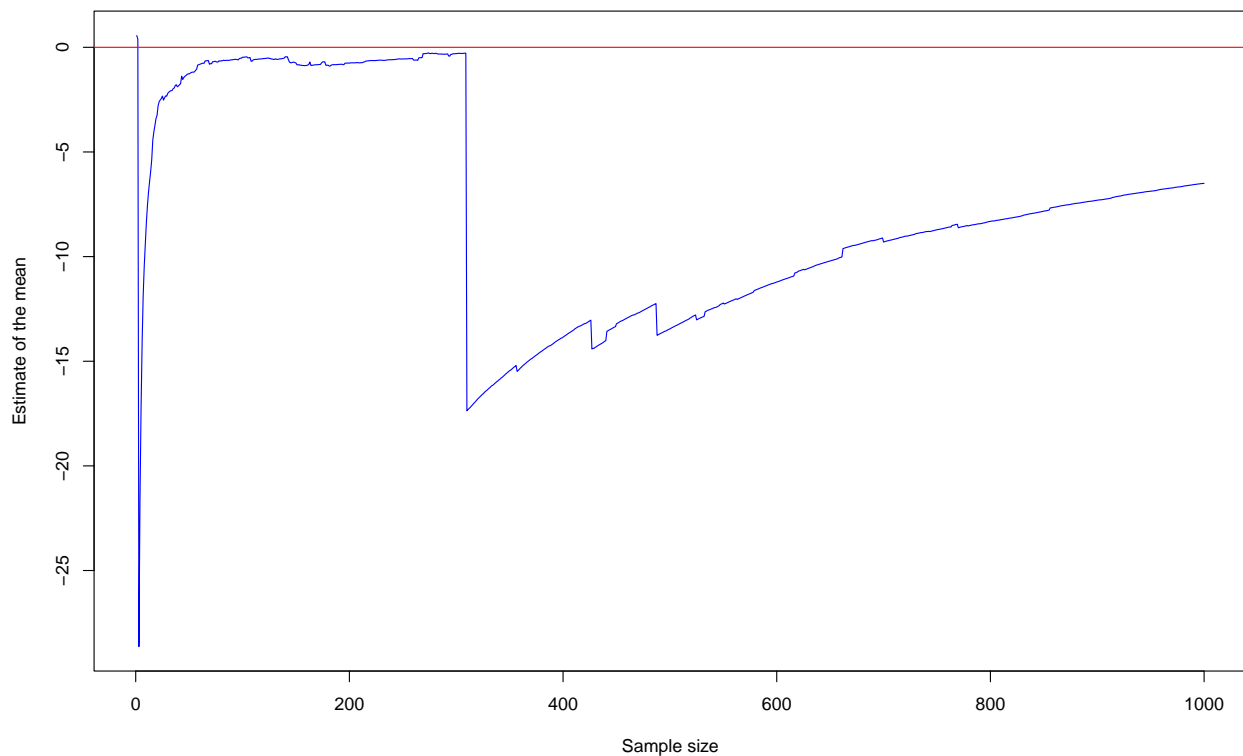
The plot generated by the code shows the estimates of the mean of a Cauchy distribution with increasing sample sizes. The blue line represents the estimated mean, computed as the cumulative sum of the ratio of two random samples from a standard normal distribution (converted to a Cauchy distribution), divided by the sample size. The red line represents the true mean, which is zero in this case. The Cauchy distribution does not have a well-defined mean and variance, as its probability density function has "heavy tails" that extend to infinity in both directions. This means that the distribution does not converge to a particular value as the sample size increases.

Question 3

```r
knitr::opts_chunk$set(fig.width=12, fig.height=8, fig_path='figures/',echo=TRUE, warning=FALSE, message

set.seed(123)

# Creating function
sim_linear_model_1 <- function(n, k, alpha, beta, sigma) {
  X <- matrix(rnorm(n * k, mean = 0, sd = 1), n, k)
  x <- X[,1]
  epsilon <- rnorm(n, mean = 0, sd = sigma)
  y <- alpha + X %*% beta + epsilon
  return(list(x = x, y = y))

}

# Create scatterplot with regression line
sim_data_1 <- sim_linear_model_1(n = 100, k = 1, alpha = 2, beta = 3, sigma = 1)
plot(sim_data_1$x, sim_data_1$y, xlab = "x", ylab = "y")
abline(lm(y ~ x, data = sim_data_1), col = "red")
```
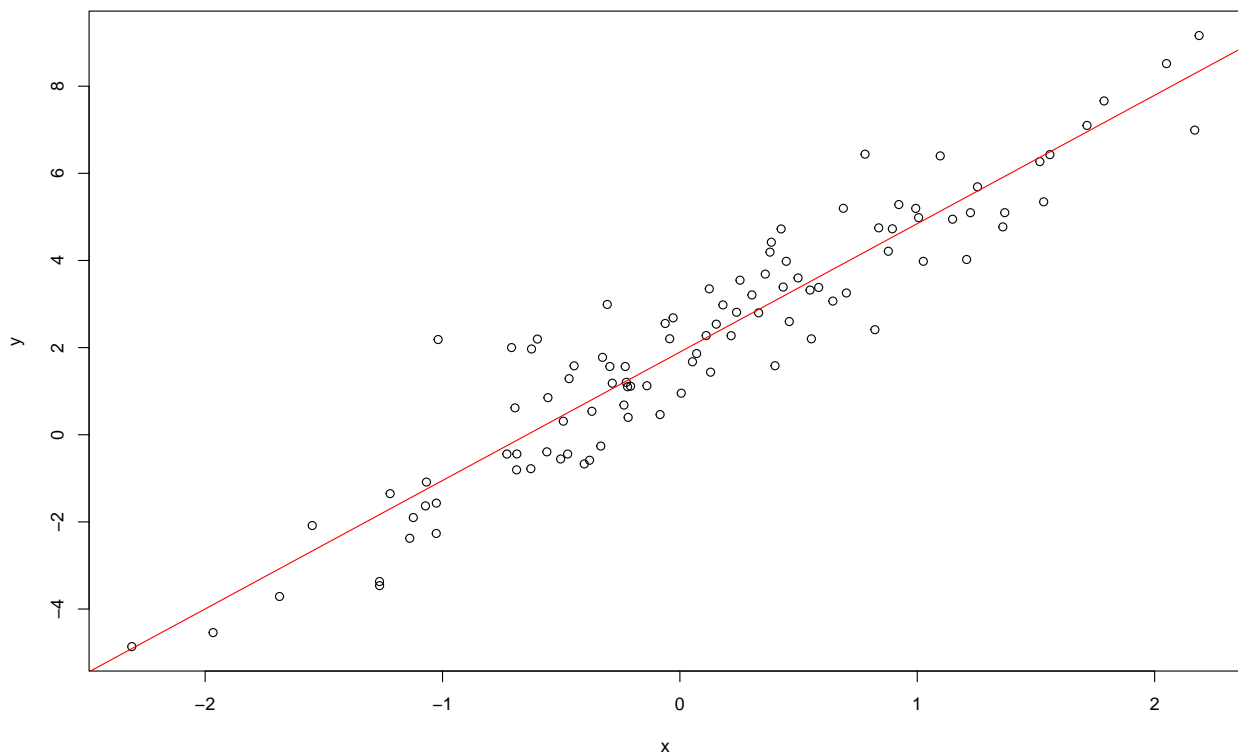
```r
# Define numbers of simulations

n_sims <- 1000

# Create an empty vector to store the estimated slope coefficients
beta_lse <- numeric(n_sims)

# Loop through the simulations
for (i in 1:n_sims) {
  # Simulate the data
  sim_data_1 <- sim_linear_model_1(n = 100, k = 1, alpha = 2, beta = 3, sigma = 1)

  # Estimate the slope coefficient using linear regression
  fit <- lm(sim_data_1$y ~ sim_data_1$x)
  beta_lse[i] <- coef(fit)[2]
}

# Creating histogram
hist(beta_lse, main = "Histogram of Beta Estimates", xlab = "Beta Estimate")
```

Histogram of Beta Estimates
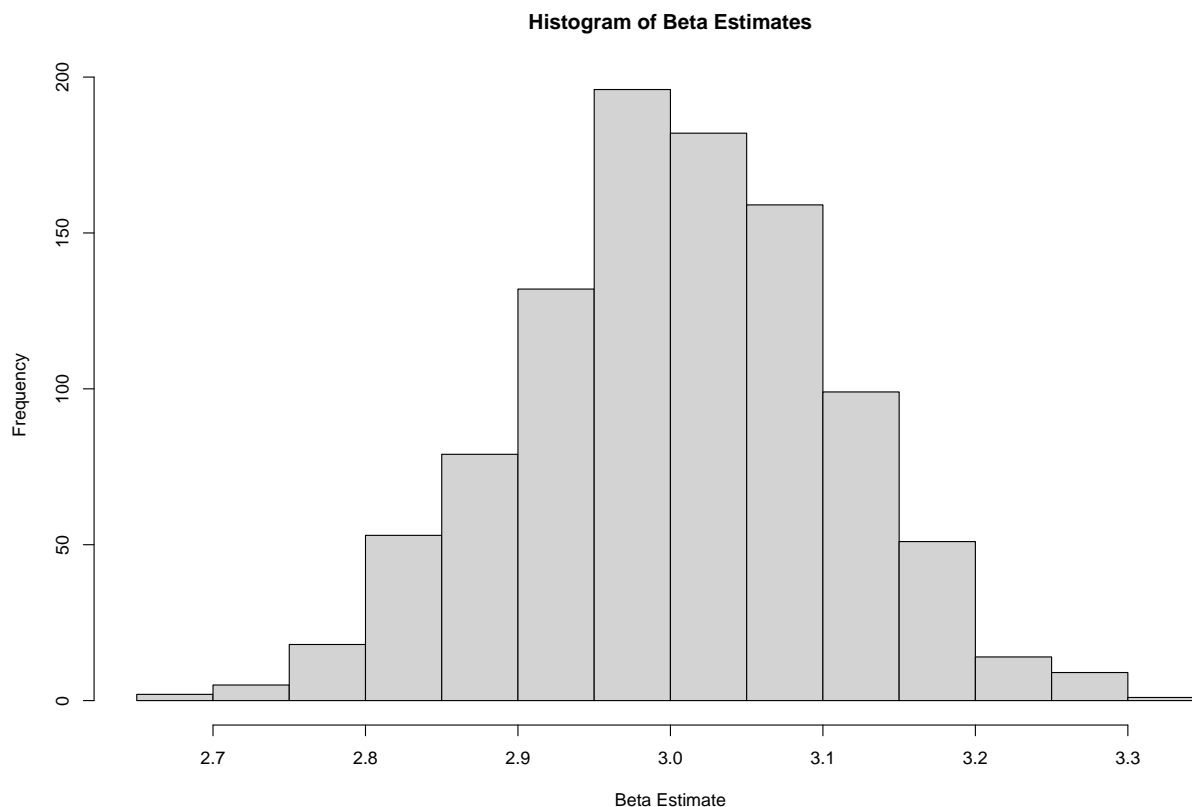
The approximately normal distribution of the beta_lse estimates and its central tendency around the true value of the slope coefficient of 3 suggests that the beta estimate is consistent. This means that as the number of simulations approaches infinity, the estimates produced by the estimator will converge to the true population value of the slope coefficient, and the distribution of estimates will become increasingly concentrated around the true value.

Latent variables are variables that can only be inferred indirectly from other observable variables that can be directly observed or measured. In the given simulation function sim_linear_model, the true intercept and slope coefficient are set to alpha = 2 and beta = 3 respectively, and we know that the error term has a variance of 1, that is, sigma^2 = 1. Therefore, the latent values of the model in this case are alpha = 2 and beta = 3. These values are the true parameters that generate the data, but they cannot be directly observed or measured. Instead, they are estimated from the observed data.

One potentially interesting regression to run could be to investigate the relationship between a person's level of physical activity and their self-reported levels of happiness. We could use a prior distribution to specify our beliefs about the coefficient of interest, beta, before observing any data. For example, we may specify a prior distribution that assigns higher probability to positive values of beta, based on our a priori expectation that physical activity is likely to have a positive effect on happiness. We can then update our beliefs about beta after observing the data, using Bayes' theorem to obtain the posterior distribution of beta.

The posterior distribution of beta represents our updated beliefs about the coefficient after observing the data, taking into account both the prior distribution and the likelihood of the data given the model. The posterior distribution can be used to make inferences about the relationship between physical activity and happiness, such as the probability that the coefficient of physical activity is positive or negative, or the range of plausible values for the coefficient.

We will now implement a Bayesian analysis of the relationship between physical activity and happiness.

```
knitr::opts_chunk$set(fig.width=12, fig.height=8, fig_path='figures/',echo=TRUE, warning=FALSE, message

library(MASS)  # for the mvrnorm function

# Define function to simulate data
sim_linear_model_prior <- function(n, k, alpha, beta_prior_mean, beta_prior_sd, sigma) {
  X <- matrix(rnorm(n * k, mean = 0, sd = 1), n, k)
  beta <- rnorm(k, mean = beta_prior_mean, sd = beta_prior_sd)
```

```r
  epsilon <- rnorm(n, mean = 0, sd = sigma)
  y <- alpha + X %*% beta + epsilon
  return(list(X = X, y = y))
}

# Set seed for reproducibility
set.seed(123)

# Set prior parameters mu_0 and sigma_0
mu_0 <- 1
sigma_0 <- 2

# Compute posterior for n = 50, 100, and 200
for (n in c(50, 100, 200)) {

  # Simulate data using sim_linear_model_prior()
  k <- 1
  alpha <- 0
  beta_prior_mean <- 0
  beta_prior_sd <- 1
  sigma <- 1

  sim_data <- sim_linear_model_prior(n, k, alpha, beta_prior_mean, beta_prior_sd, sigma)
  X <- sim_data$X
  y <- sim_data$y

  # Compute posterior parameters
  sigma_n <- solve(sigma_0^(-1) + t(X) %*% X)
  mu_n <- sigma_n %*% (sigma_0^(-1) * mu_0 + t(X) %*% y)

  # Plot posterior density
  x_grid <- seq(-5, 5, length.out = 1000)
  post_dens <- dnorm(x_grid, mean = mu_n, sd = sqrt(sigma_n))
  plot(x_grid, post_dens, type = "l", lwd = 2, main = paste0("Posterior Density (n = ", n, ")"))

}
```
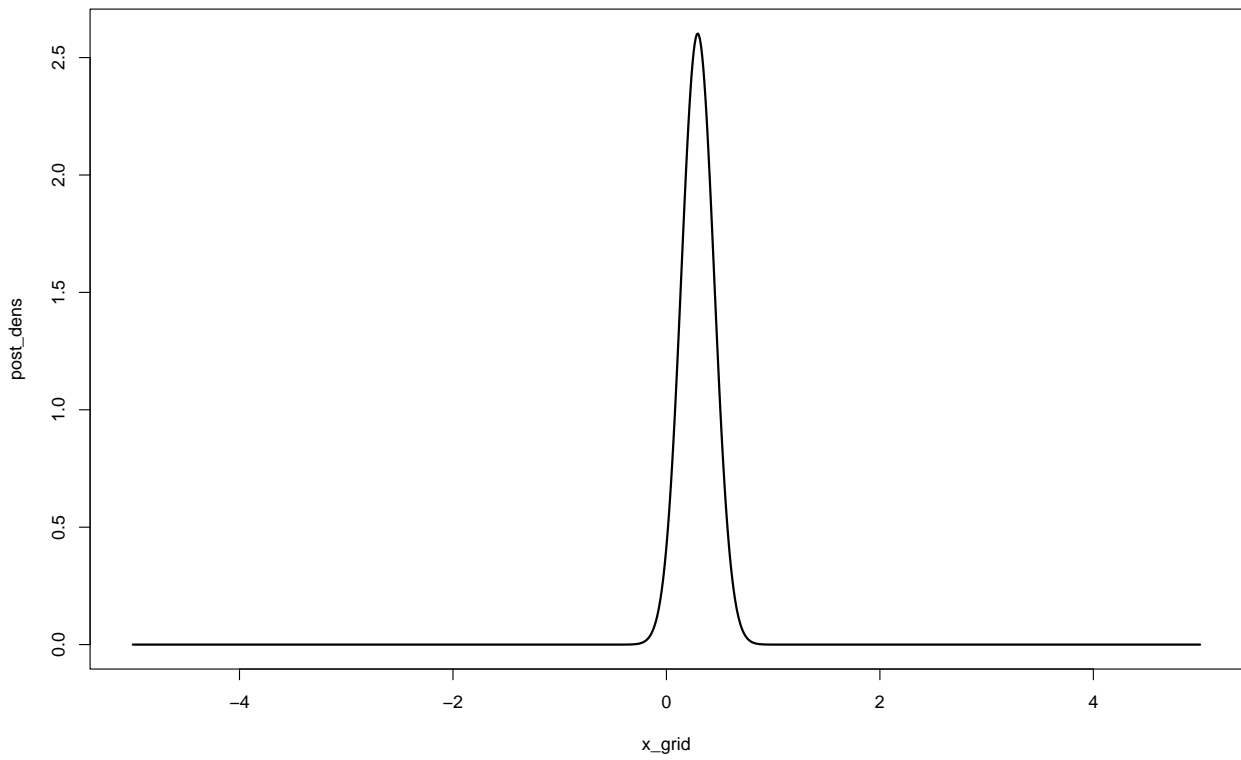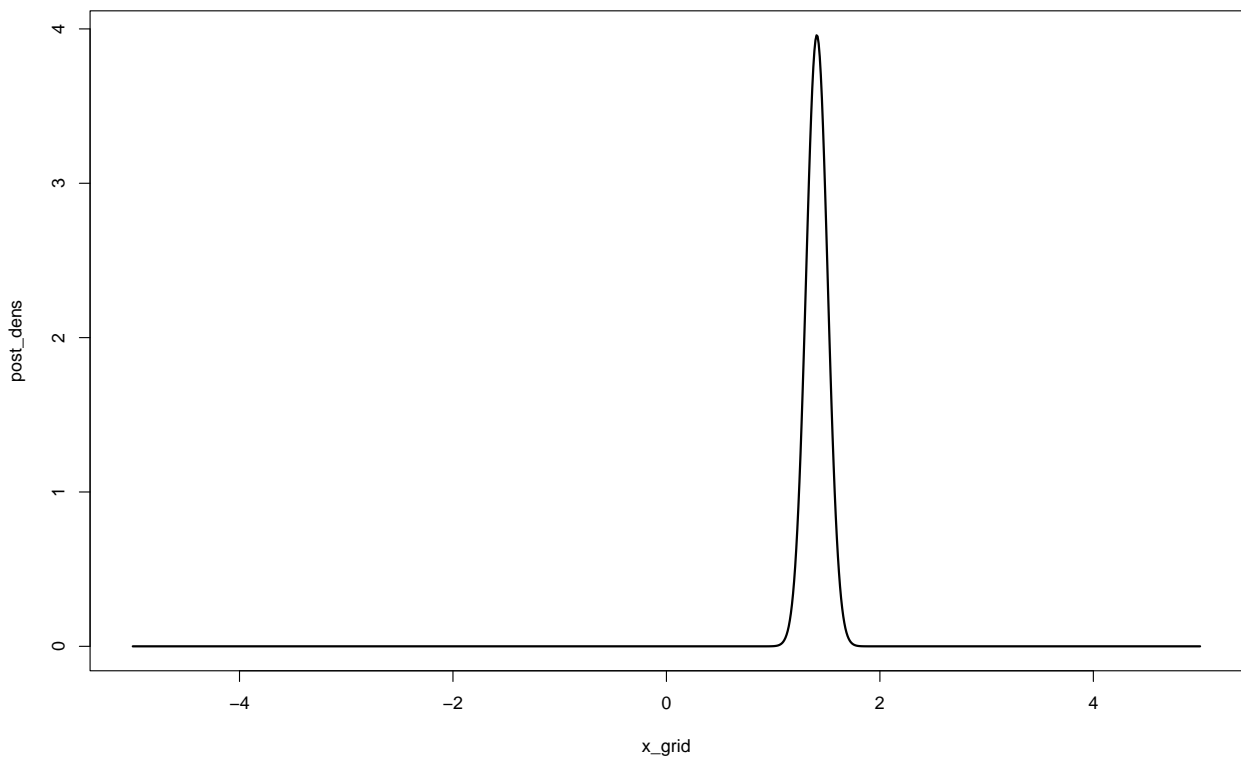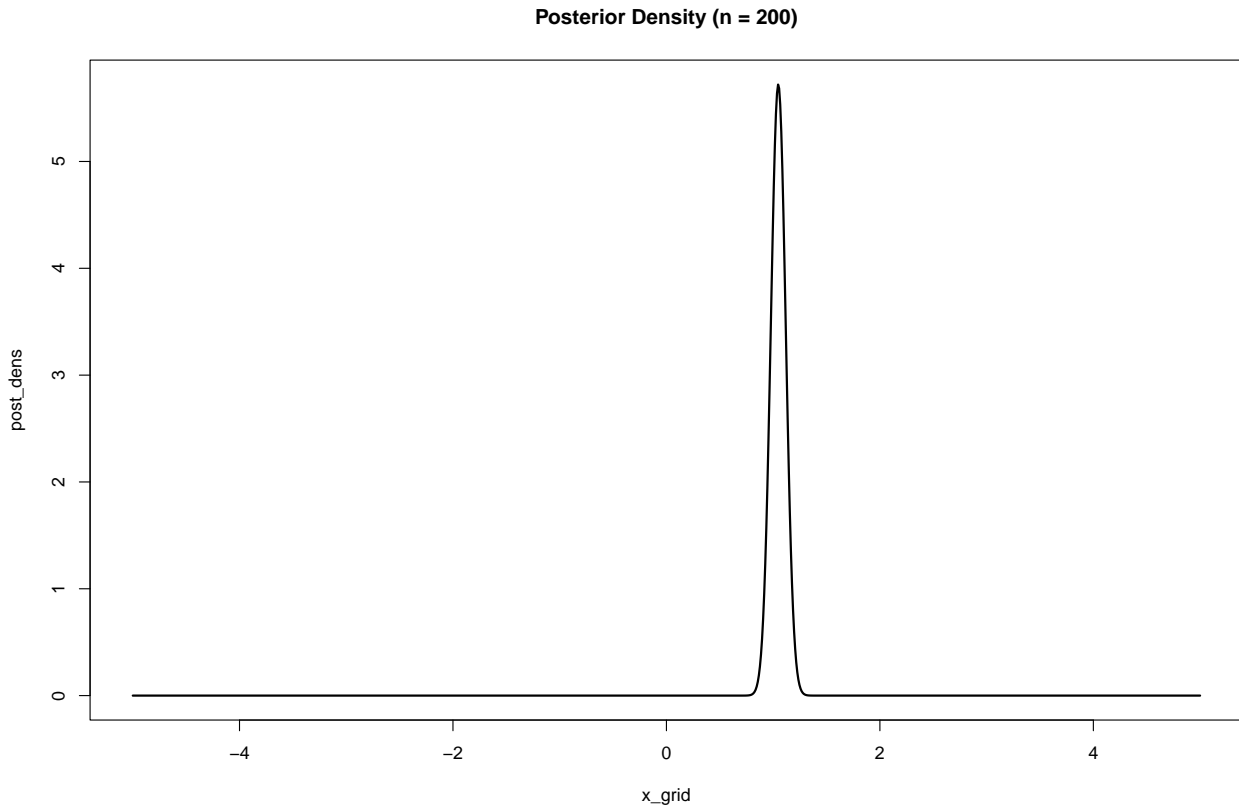
Posterior Density (n = 50)

Posterior Density (n = 100)

Posterior Density (n = 200)

We are interested in the relationship between a person's level of physical activity and their self-reported levels of happiness. To incorporate our prior beliefs about this relationship, we set a positive prior on the slope coefficient beta. The mean of the prior distribution is set to 1, indicating that we believe physical activity is likely to have a positive effect on happiness, while the standard deviation of the prior distribution is set to 2, indicating that we allow for a wide range of plausible values for the slope coefficient beta.

The plots of the posterior density, given different sample sizes, show that, as the sample size increases, the posterior density becomes more concentrated around the true value of the slope parameters. Additionally, as the sample size increases, the posterior density becomes narrower, indicating that the uncertainty in the posterior estimate decreases.