INFO510FA24-AB-2

Aditya Bandimatt

2024-11-27

Monte_carlo estimation

Setup and load libraries You want to estimate the average height of adults in a particular city. You have some prior knowledge that the average height of adults globally is around 170 cm, with a standard deviation of 10 cm (prior belief). You then gather a small random sample of 5 adults from the city, with their heights being: 168, 172, 165, 174, 169 (observed data). Using Bayesian Monte Carlo, we will update our belief about the average height in this city. we will estimate the average height of a population using Bayesian Monte Carlo methods.We assume that the average height is normally distributed with a mean of 170 cm and a standard deviation of 10 cm.

```r
# Prior belief about the population's average height
mean_prior <- 170  # Prior mean height
sd_prior <- 10     # Prior standard deviation (spread)

# Visualize the prior distribution
x <- seq(140, 200, length.out = 100)
prior_density <- dnorm(x, mean_prior, sd_prior)

plot(x, prior_density, type = "l", col = "blue", lwd = 2,
     main = "Prior Distribution of Height",
     xlab = "Height (cm)", ylab = "Density")
abline(v = mean_prior, col = "red", lty = 2)
```

Prior Distribution of Height
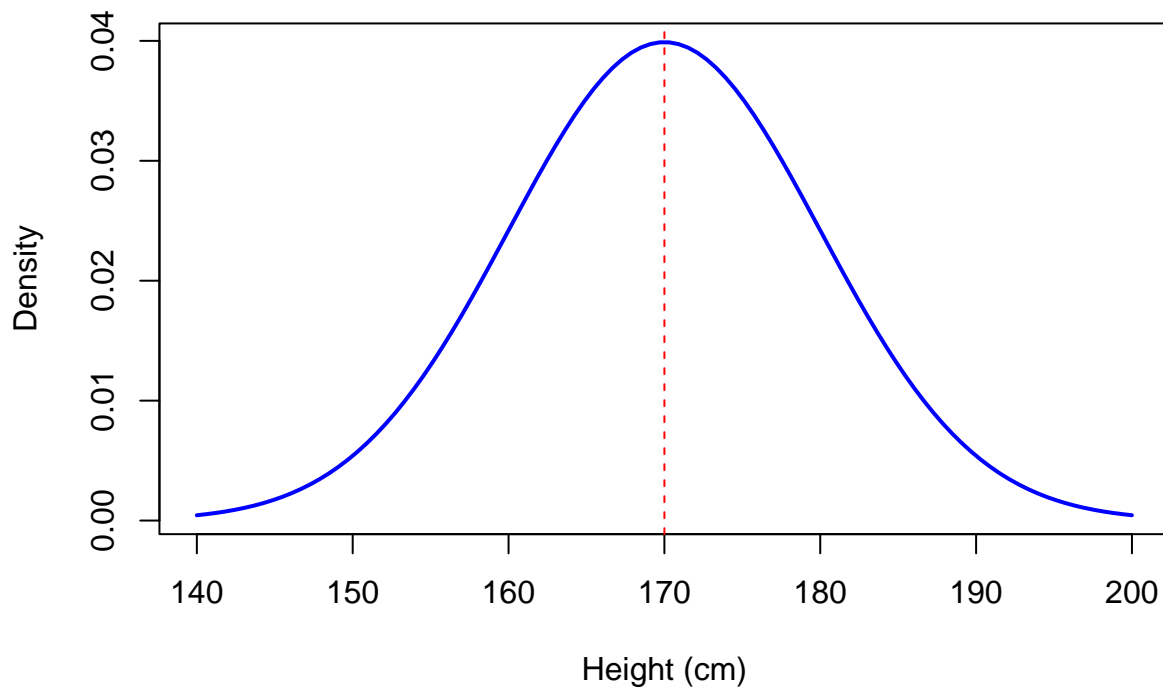
Collect a sample of 5 adults' heights from the city: 168, 172, 165, 174, 169 CM

```r
# New data from a sample
sample_heights <- c(168, 172, 165, 174, 169)

# Calculate sample mean and standard deviation
sample_mean <- mean(sample_heights)
sample_sd <- sd(sample_heights)

# Print the sample summary
sample_mean
```

```
## [1] 169.6
```

```r
sample_sd
```

```
## [1] 3.507136
```

We will use Bayesian Monte Carlo to update our prior belief based on this new data. We will simulate many possible scenarios and see how the new data updates our belief about the average height.

We assume the prior is a normal distribution (mean = 170 cm, sd = 10 cm). We compute the likelihood of observing the new data. We generate many samples from the posterior distribution using Monte Carlo simulation.

```r
# Number of Monte Carlo simulations
n_sim <- 10000

# Generate prior samples
```

```r
prior_samples <- rnorm(n_sim, mean_prior, sd_prior)

# Likelihood: Assume measurement errors follow a normal distribution with
# sample sd. For each sample from the prior, we compute how likely the new
# data is
likelihood <- function(mu) {
  prod(dnorm(sample_heights, mean = mu, sd = sample_sd))
}

# Compute likelihood for each prior sample
likelihood_values <- sapply(prior_samples, likelihood)

# Compute the posterior distribution (unnormalized)
posterior_samples <- likelihood_values * prior_samples

# Normalize the posterior
posterior_samples <- posterior_samples / sum(posterior_samples)

# Resample from the posterior
posterior <- sample(prior_samples, size = n_sim, prob = posterior_samples, replace = TRUE)

# Plot the posterior distribution
posterior_density <- density(posterior)
plot(posterior_density, col = "green", lwd = 2,
     main = "Posterior Distribution of Height",
     xlab = "Height (cm)", ylab = "Density")
abline(v = sample_mean, col = "red", lty = 2)
```
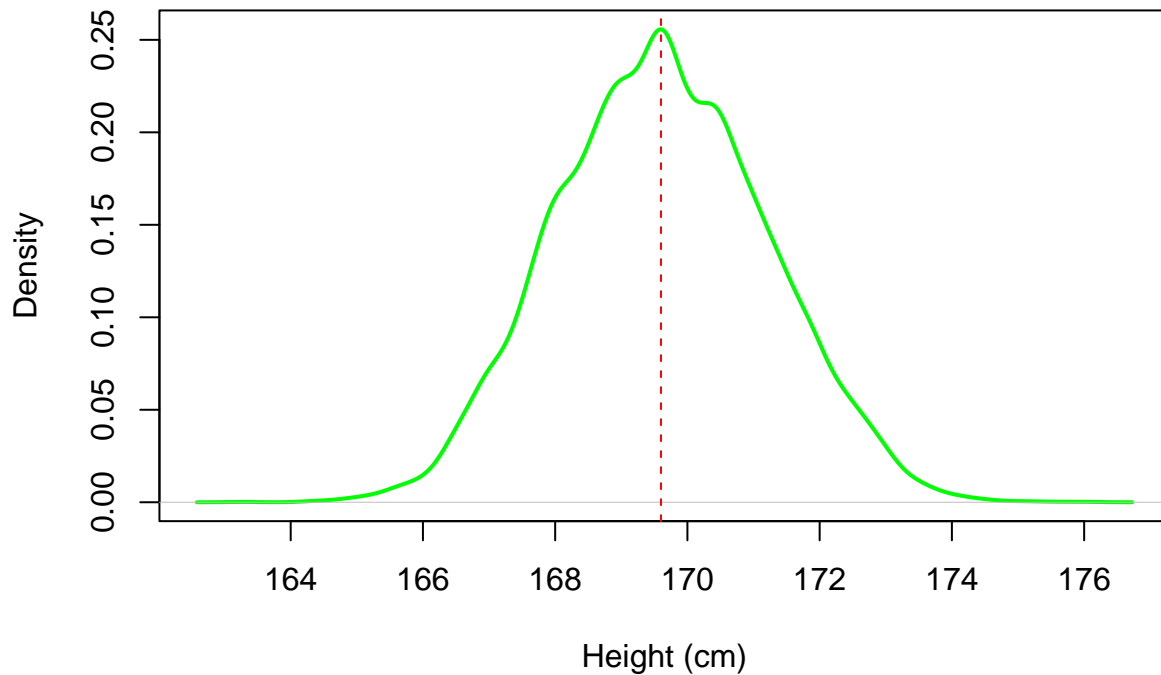
Posterior Distribution of Height


```r
# Updated distribution for the average height of the population,
# after running Monte Carlo simulation

# Posterior mean and credible interval
posterior_mean <- mean(posterior)
credible_interval <- quantile(posterior, c(0.025, 0.975))

# Print results
posterior_mean
```

```
## [1] 169.6055
```

```r
credible_interval
```

```
##     2.5%    97.5%
## 166.5895 172.7158
```

From the simulation, we find that the posterior mean of the population height is approximately r round(posterior_mean, 2) cm, and the 95% credible interval is between r round(credible_interval[1], 2) cm and r round (credible_interval[2], 2) cm