SDA Group Submission Assignment Assign3

Group Gr18

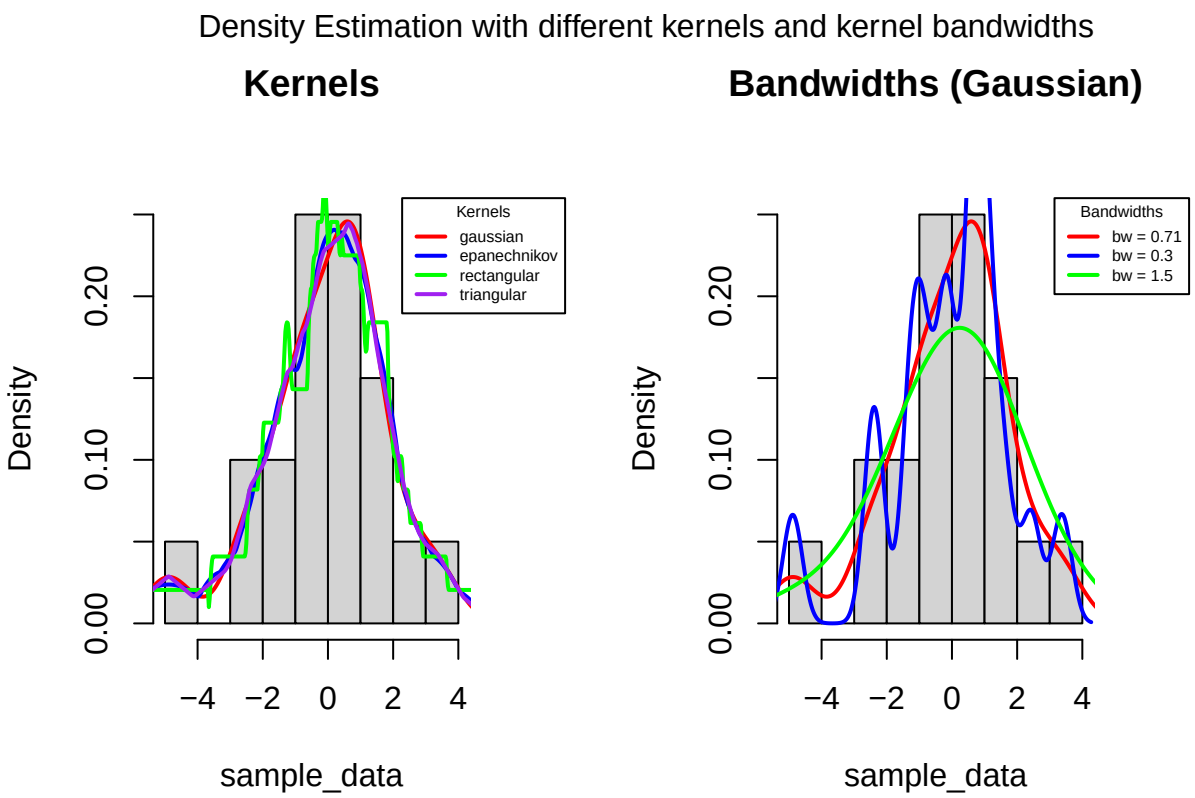
MengliFeng (2720589) and PepijnVanOostveen (2801582)

Exercise 1

a.

```r
# Set seed for reproducibility
set.seed(123)

# Generate random sample from t-distribution with 3 degrees of freedom
n <- 20
sample_data <- rt(n, df = 3)

# Define different kernel types and colors
kernels <- c("gaussian", "epanechnikov", "rectangular", "triangular")
colors_kernels <- c("red", "blue", "green", "purple")

# Define different bandwidth choices and colors
bandwidths <- c(density(sample_data)$bw, 0.3, 1.5)
colors_bandwidths <- c("red", "blue", "green")

# Adjust plot margins to make space for legends
par(mfrow = c(1, 2), mar = c(5, 4, 6, 4))  # Extra right margin for the legend

# Plot histogram with different kernel choices
hist(sample_data, probability = TRUE, main = "Kernels", col = "lightgray", border =
↪  "black")

for (i in seq_along(kernels)) {
  lines(density(sample_data, kernel = kernels[i]), col = colors_kernels[i], lwd = 2)
}

# Add legend outside the plot
legend("topright", inset = c(-0.3, 0), legend = kernels, col = colors_kernels, lwd = 2,
↪  cex = 0.5, title = "Kernels", xpd = TRUE)

# Plot histogram with different bandwidth choices
hist(sample_data, probability = TRUE, main = "Bandwidths (Gaussian)", col = "lightgray",
↪  border = "black")

for (i in seq_along(bandwidths)) {
  lines(density(sample_data, bw = bandwidths[i]), col = colors_bandwidths[i], lwd = 2)
}
```

```r
# Add legend outside the plot
legend("topright", inset = c(-0.3, 0), legend = paste("bw =", round(bandwidths, 2)), col
↪  = colors_bandwidths, lwd = 2, cex = 0.5,
       title = "Bandwidths", xpd = TRUE)

# --- Add an Overall Title ---
mtext("Density Estimation with different kernels and kernel bandwidths",line = 4, cex =
↪  1, font = 1, adj = 1)
```



b.

From the generated plots, we can observe: • Effect of Kernel Choice: Different kernels produce similar overall shapes, but their smoothness varies slightly. The Gaussian kernel is the smoothest, while the rectangular kernel has more abrupt changes. • Effect of Bandwidth Choice: The bandwidth has a much larger influence than the kernel. A smaller bandwidth (0.3) captures more fluctuations in the data, while a larger bandwidth (1.5) smooths out more features. • Key Influence: Bandwidth choice has a bigger impact on the estimator compared to kernel choice. ## c.

```r
h_opt <- function(x) {
  sigma_hat <- min(sd(x), IQR(x) / 1.34)  # Compute standard deviation and interquartile
↪  range
  h_optimal <- 1.06 * sigma_hat * length(x)^(-1/5)  # Optimal bandwidth formula
  return(h_optimal)
}

# Compute optimal bandwidth for the sample
h_opt_value <- h_opt(sample_data)
```

```r
# Compare with R's default bandwidth
default_bw <- density(sample_data)$bw

# Print results
cat("Optimal Bandwidth (h_opt):", h_opt_value, "\n")
```

```
## Optimal Bandwidth (h_opt): 0.831087
```

```r
cat("Default R Bandwidth:", default_bw, "\n")
```

```
## Default R Bandwidth: 0.7056399
```

Exercise 2

a.

b.

c.