Math3810 - Probability
Section 001 - Fall 2025
Notes: Expectation and Densities Through Conditioning

University of Colorado Denver / College of Liberal Arts and Sciences

Department of Mathematics - Dr. Robert Rostermundt

The Problem:

The train leave the station every fifteen minutes starting on the hour. We arrive at the train station at a random time between 7:10am and 7:30am. What is our expected waiting time for the train?

Let W equal the waiting time until our train leaves the station and let T be the time we arrive at the train station. Then $Im(W) = [0, 15)$ with, as of yet, an unknown distribution. If we let $T = 0$ correspond to 7:10am we see that $T \sim Uni(0, 20)$. And so we are looking for $E[W]$. Our definition of expected value is

$$E[W] = \int_{-\infty}^{\infty} w \cdot f_W(w)\, dw.$$

However, at this point we do not know the distribution of W and do not know it's density function f_W.

Theoretical Tools:

We will compute this expectation in two ways by conditioning on the following events:

$A := \{0 < T \leq 5\}$ which is the event that we arrive between 7:10am and 7:15am.

$B := \{5 < T < 20\}$ which is the event that we arrive between 7:15am and 7:30am.

Then since A and B partition the sample space we can compute the expected value as

$$E[W] = \mathbb{P}(A) \cdot E\big[\, W \mid A \,\big] + \mathbb{P}(B) \cdot E\big[\, W \mid B \,\big].$$

Alternatively, we can describe the density for W using conditional densities as

$$f_W(w) = \mathbb{P}(A) \cdot f_{W|A}(w) + \mathbb{P}(B) \cdot f_{W|B}(w).$$

From here we can use integration to directly compute the expectation.

Calculations:

We first make the easy calculations

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(0 < T \le 5) = F_T(5) = \frac{5-0}{20-0} = \frac{1}{4} \\
\mathbb{P}(B) &= 1 - \mathbb{P}(A) = \frac{3}{4}
\end{aligned}$$

Method #1: We need to find $E\big[W \,|\, A\big]$ and $E\big[W \,|\, B\big]$. To do this we will find the relationship between T and W. In particular,

$$W = \begin{cases} 5 - T & : \quad 0 < T \le 5 \\ 20 - T & : \quad 5 < T < 20 \end{cases}$$

Then

$$E\big[W \,|\, A\big] = E\big[5 - T \,\big|\, 0 < T \le 5\big] = 5 - E\big[T \,\big|\, 0 < T \le 5\big] = 5 - \frac{5}{2}.$$

To see this, we have the conditional distribution $T \,\big|\, 0 < T \le 5 \sim Uni(0,5)$ and so $E\big[T \,|\, 0 < T \le 5\big] = 5/2$. That gives us

$$E\big[W \,|\, A\big] = \frac{5}{2}.$$

Similarly,

$$E\big[W \,|\, B\big] = E\big[20 - T \,\big|\, 5 < T < 20\big] = 20 - E\big[T \,\big|\, 5 < T < 20\big] = 20 - \frac{25}{2}.$$

To see this, we have the conditional distribution $T \,\big|\, 5 < T < 20 \sim Uni(5,20)$ and so $E\big[T \,|\, 5 < T < 20\big] = 25/2$. That gives us

$$E\big[W \,|\, A\big] = \frac{15}{2}.$$

Putting the pieces together we have

$$E[W] = \underbrace{\mathbb{P}(A)}_{1/4} \cdot \underbrace{E\big[W \,|\, A\big]}_{5/2} + \underbrace{\mathbb{P}(B)}_{3/4} \cdot \underbrace{E\big[W \,|\, B\big]}_{15/2} = \frac{50}{8}$$

So the expected waiting time is 6.25 minutes.

Method #2: Alternatively we compute the density function as

$$f_W(w) = \underbrace{\mathbb{P}(A)}_{1/4} \cdot f_{W|A}(w) + \underbrace{\mathbb{P}(B)}_{3/4} \cdot f_{W|B}(w)$$

and then integrate. Letting $w(t) = 5 - t$, with the inverse relationship $t(w) = 5 - w$, we have from our change of variables formula that

$$f_{W|A}(w) = f_{T|A}\big(t(w)\big) \cdot \left|\frac{dt}{dw}\right| = f_{T|A}(5 - w) \cdot |-1| = \frac{1}{5}, \text{ when } 0 < w \le 5.$$

Similarly, letting $w(t) = 20 - t$, with the inverse relationship $t(w) = 20 - w$, we again have from our change of variables formula that

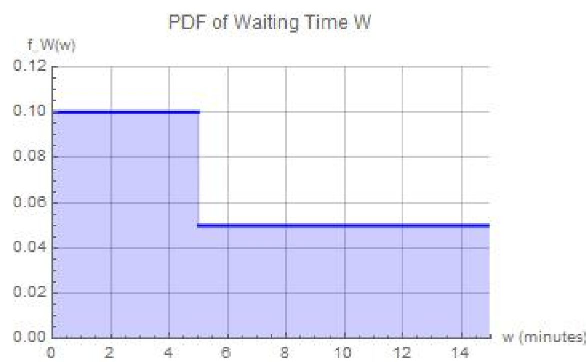
$$f_{W|B}(w) = f_{T|B}\big(t(w)\big) \cdot \left|\frac{dt}{dw}\right| = f_{T|B}(20 - w) \cdot |-1| = \frac{1}{15}, \text{ when } 0 < w \leq 15.$$

Putting these pieces together, and seeing that $f_{W|A}$ only has a non-zero contribution to $f_W(w)$ when $0 \leq w \leq 5$, we have

$$f_W(w) = \underbrace{\mathbb{P}(A)}_{1/4} \cdot \underbrace{f_{W|A}(w)}_{\substack{1/5 \text{ when} \\ 0 \leq w \leq 5}} + \underbrace{\mathbb{P}(B)}_{3/4} \cdot \underbrace{f_{W|B}(w)}_{\substack{1/15 \text{ when} \\ 0 \leq w \leq 15}}$$

Finally we have the density function f_W.

$$f_W(w) = \begin{cases} \dfrac{1}{10} & : \quad 0 \leq w \leq 5 \\[2ex] \dfrac{1}{20} & : \quad 5 < w < 15 \\[2ex] 0 & : \quad \text{Otherwise.} \end{cases}$$



A quick check verifies that

$$\int_{-\infty}^{\infty} f_W(w)\,dw = 1$$

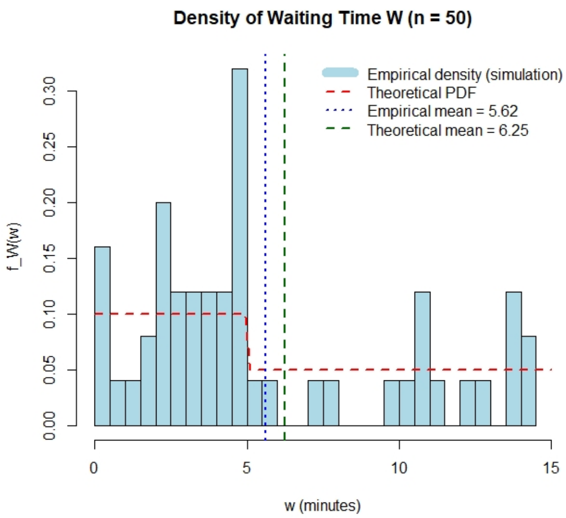
and f_W is a valid density function. We can now compute

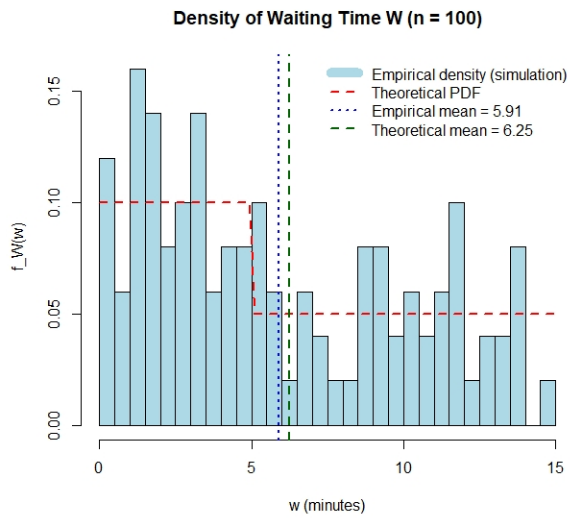
$$E[W] = \int_{-\infty}^{\infty} w \cdot f_W(w)\,dw = \int_{w=0}^{w=5} \frac{w}{10}\,dw + \int_{w=5}^{w=15} \frac{w}{20}\,dw = \frac{250}{40} = 6.25$$

as we saw in our previous calculations.

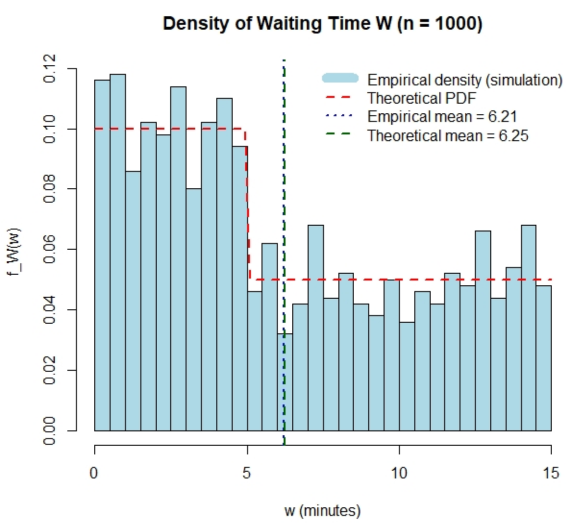
Simulations:

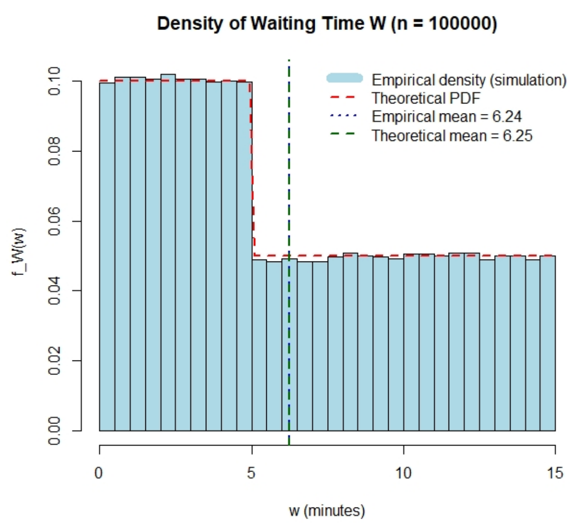
I have simulated this process in R. Below are graphics from four different simulations with number trials $n = 50, 100, 1000, 100000$. The R code is provided after the graphics.





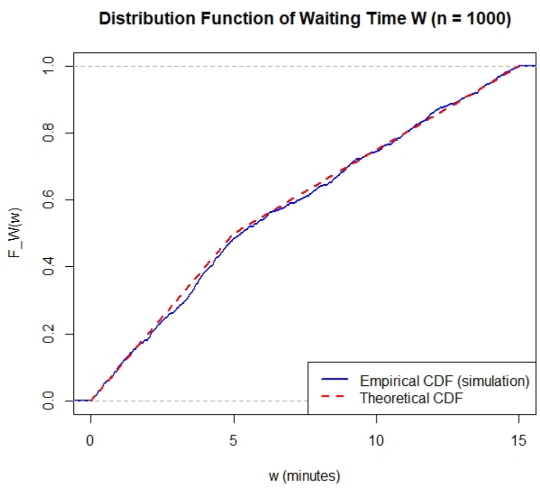
Notice how the empirical mean differs noticeably from the theoretical mean with relatively small sample sizes $n = 50$ and $n = 100$.

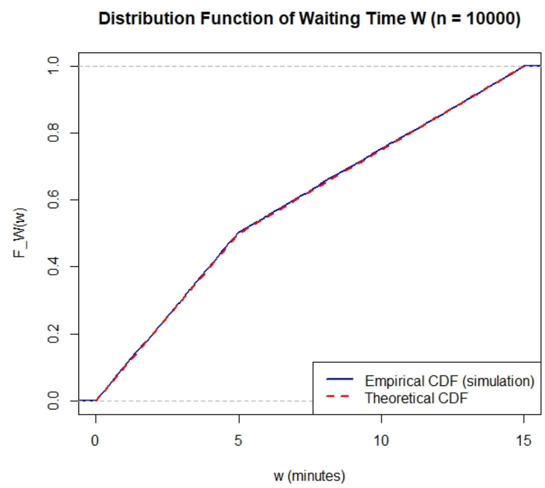




Notice that the empirical mean differs very little from the theoretical mean with larger sample sizes $n = 1000$ and $n = 100000$. Also, the empirical density fits the theoretical density quite nicely for "large" sample sizes.

Next are a couple of simulations of the distribution function of W with sample sizes $n = 1000$ and $n = 10,000$.





The R Code: Here is the R code used for the above simulations.

```
# Code to Simulate the Density Function and Expectation of the Waiting Time

set.seed(123)
n <- 100000  # number of simulations

# Step 1: Simulate random arrival times (in minutes after 7:10)
T <- runif(n, 0, 20)


# Step 2: Compute waiting times
W <- ifelse(T < 5, 5 - T, 20 - T)


# Step 3: Define theoretical PDF of W
f_W <- function(w) {
  ifelse(w < 0, 0,
         ifelse(w < 5, 1/10,
                ifelse(w <= 15, 1/20, 0)))
}


# Step 4: Compute theoretical quantities
E_theory <- (1/10)*integrate(function(w) w, 0, 5)$value +
  (1/20)*integrate(function(w) w, 5, 15)$value
E_W2 <- (1/10)*integrate(function(w) w^2, 0, 5)$value +
  (1/20)*integrate(function(w) w^2, 5, 15)$value
Var_theory <- E_W2 - E_theory^2


# Step 5: Empirical stats
E_sim <- mean(W)
Var_sim <- var(W)
```

```
# Step 6: Plot simulated vs. theoretical density
hist(W, breaks = 50, freq = FALSE, col = "lightblue",
     main = sprintf("Density of Waiting Time W (n = %d)", n),
     xlab = "w (minutes)", ylab = "f_W(w)", xlim = c(0, 15))
curve(f_W(x), from = 0, to = 15, add = TRUE, col = "red", lwd = 2, lty = 2)

# Add mean lines
abline(v = E_sim, col = "blue", lwd = 2, lty = 3)
abline(v = E_theory, col = "darkgreen", lwd = 2, lty = 2)

# Add legend
legend("topright",
       legend = c("Empirical density (simulation)",
                  "Theoretical PDF",
                  sprintf("Empirical mean = %.2f", E_sim),
                  sprintf("Theoretical mean = %.2f", E_theory)),
       col = c("lightblue", "red", "blue", "darkgreen"),
       lwd = c(10, 2, 2, 2), lty = c(1, 2, 3, 2), bty = "n")

# Step 7: Print summary results
cat("=== Simulation Summary ===\n")
cat(sprintf("Number of trials: %d\n", n))
cat(sprintf("Empirical mean E[W] = %.4f\n", E_sim))
cat(sprintf("Theoretical mean E[W] = %.4f\n", E_theory))
cat(sprintf("Empirical variance Var(W) = %.4f\n", Var_sim))
cat(sprintf("Theoretical variance Var(W) = %.4f\n", Var_theory))



############################################################################
############################################################################



#Code to Simulate the Distribution Function of the Waiting Time

set.seed(123)
n <- 10000  # number of simulations

# Step 1: Simulate random arrival times
T <- runif(n, 0, 20)

# Step 2: Compute waiting times
W <- ifelse(T < 5, 5 - T, 20 - T)

# Step 3: Theoretical CDF of W
F_W <- function(w) {
  ifelse(w < 0, 0,
         ifelse(w < 5, w/10,
                ifelse(w <= 15, (w + 5)/20, 1)))
}
```

```
# Step 4: Plot empirical and theoretical CDFs
plot(
  ecdf(W),
  col = "blue", lwd = 2,
  pch = ".", cex = 0.5,  # smaller dots
  main = sprintf("Distribution Function of Waiting Time W (n = %d)", n),
  xlab = "w (minutes)", ylab = "F_W(w)", xlim = c(0, 15)
)
curve(F_W(x), from = 0, to = 15, add = TRUE, col = "red", lwd = 2, lty = 2)

legend(
  "bottomright",
  legend = c("Empirical CDF (simulation)", "Theoretical CDF"),
  col = c("blue", "red"),
  lwd = 2, lty = c(1, 2)
)
```

Please let me know if you have any questions, comments, or corrections!