STAT 312 Week 6 Class

Cumulative Distribution Function (CDF)

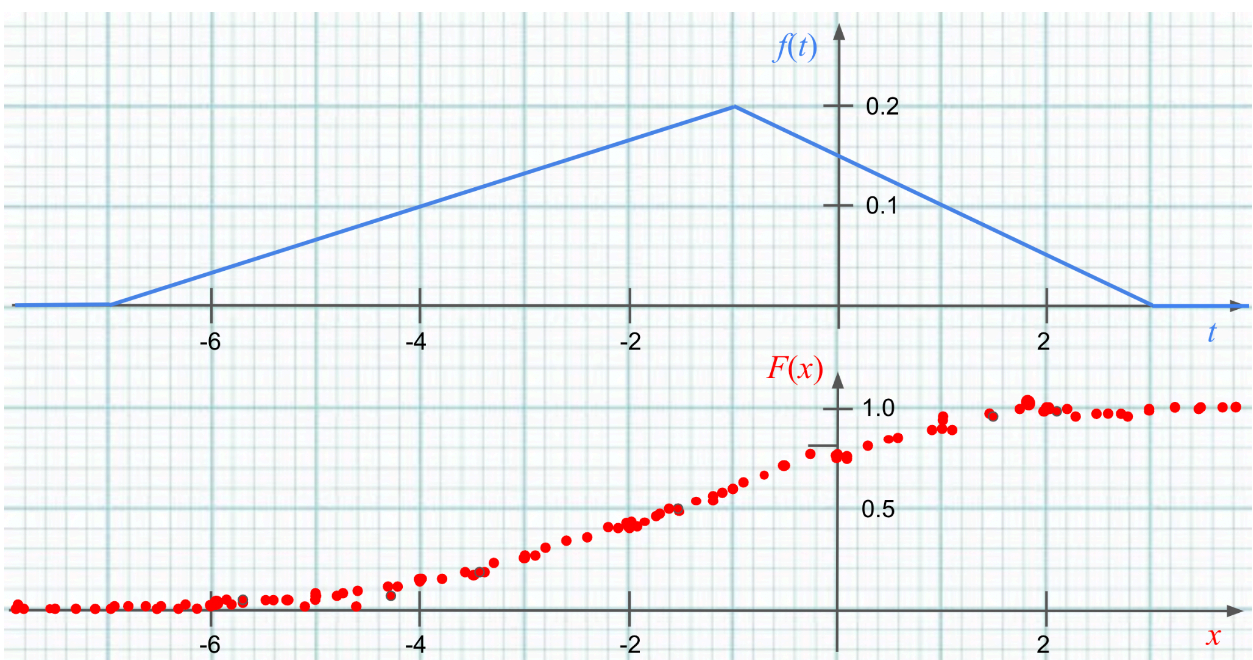
Recall that the **cumulative distribution function** $F(x)$ gives the probability that a random variable is less than or equal to x:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

In preparation for today, you examined the random variable T, the time a student arrives at class (in minutes after class starts). T was assumed to have PDF

$$f(t) = \begin{cases} \frac{t+7}{30} & -7 < t \leq -1 \\ \frac{3-t}{20} & -1 < t < 3 \\ 0 & \text{otherwise} \end{cases}$$

and you calculated the CDF $F(x)$ for different values of x and graphed them.



What are some properties of the CDF?

- The CDF can only increase as you move from left to right. That's because as you increase x, you can only accumulate more probability; you can't lose probability.
- As $x \to -\infty$, $F(x) \to 0$.
- As $x \to \infty$, $F(x) \to 1$.

Let's come up with a formula for $F(x)$.

$$F(x) = \begin{cases} 0 & x \leq -7 \\ \frac{(x+7)^2}{60} & -7 < x \leq -1 \\ \frac{-x^2+6x+31}{40} & -1 < x < 3 \\ 1 & x \geq 3 \end{cases}$$

What happens if you take the derivative of the CDF?

$$F'(x) = \begin{cases} 0 & x \leq -7 \\ \frac{x+7}{30} & -7 < x \leq -1 \\ \frac{-x+3}{20} & -1 < x < 3 \\ 0 & x \geq 3 \end{cases}$$

This works because of the following: $F'(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt = f(x)$. This is called the **Fundamental Theorem of Calculus**.

Applications of the CDFs

Finding Quantiles

How would we use the CDF to find the *median* time that this student arrives at class?

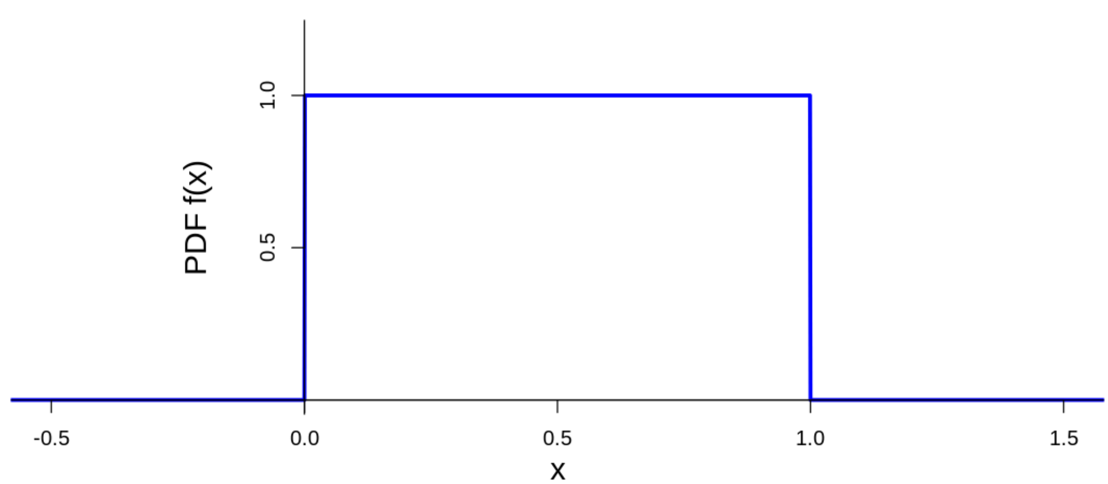
Set $F(x) = 0.5$ and solve for x.

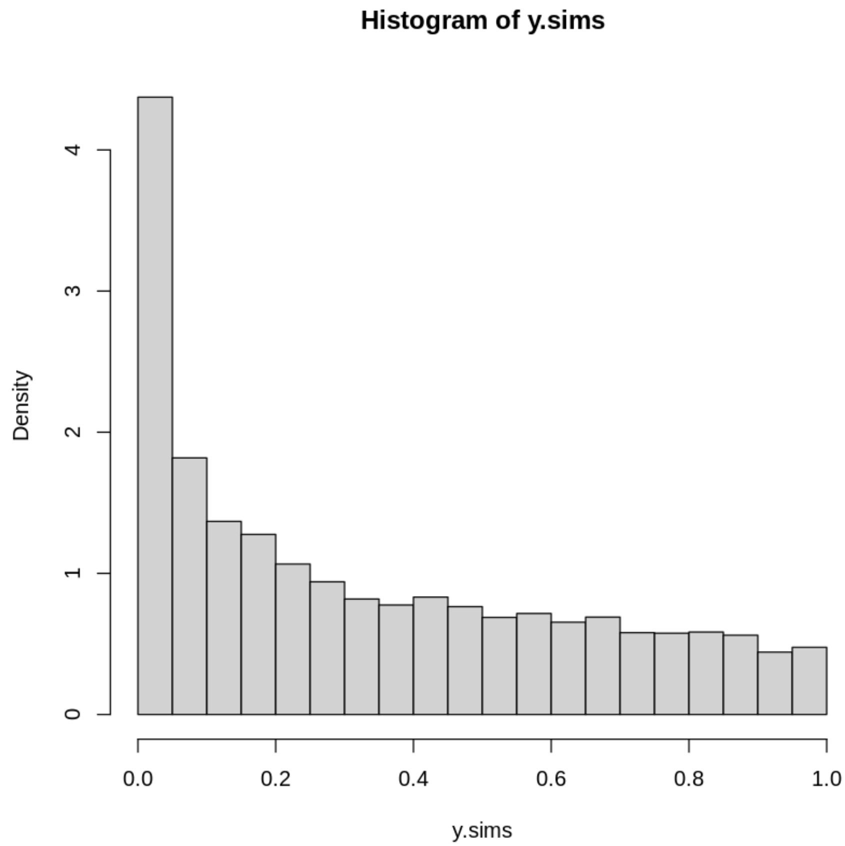
$$\frac{(x+7)^2}{60} = 0.5 \rightarrow x = -12.477, -1.523$$

Only one of these roots is in the range $-7 \leq x \leq 1$. So the median is -1.523. So 50% of the time the student arrives *more than* 1.523 minutes early; the other 50% of the time the student arrives *less than* 1.523 minutes early.

Transformations

In the R assignment for this week, you simulated the distribution of $Y = U^2$, where U was uniformly distributed between 0 and 1.





How would we find the PDF of Y?

- First find its CDF.
- Take the derivative to get its PDF.

The CDF is:

$$F(y) = P(Y \leq y) = P(U^2 \leq y) = P(U \leq \sqrt{y}) = \sqrt{y}$$

The PDF is the derivative:

$$f(y) = F'(y) = \frac{1}{2\sqrt{y}}$$

Let's graph this PDF on top of our simulations to check our answers.

Silent Auction

Suppose n people submit independent bids $U_1, U_2, ..., U_n$ that are uniformly distributed between 0 and 1 (thousand dollars). The winning bid is

$$W = max(U_1, U_2, ..., U_n).$$

In the R assignment for this week, you simulated the distribution of W. You saw that:

- The expected value $E[W]$ gets closer to 1 as n increases.
- In fact, the whole distribution of W shifted to be closer to 1 as n increases.

Now let's work out the math.

- Calculate the PDF of W.

Strategy: Find the CDF and take the derivative.

$$\begin{aligned} F(x) = P(W \le x) = P(max(U_1, U_2, ..., U_n) \le x) \\ &= P(U_1 \le x \text{ and } U_2 \le x \text{ and } ... \text{ and } U_n \le x) \\ &= P(U_1 \le x)P(U_2 \le x)... P(U_n \le x) \\ &= P(U_1 \le x)^n \\ &= x^n \end{aligned}$$

$$f(x) = F'(x) = nx^{n-1}.$$

Let's graph this PDF on top of our simulations to check our answer.

- Calculate $E[W]$.

$$E[W] = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x \cdot nx^{n-1}dx = \int_0^1 nx^n dx = \frac{n}{n+1}.$$

This agrees with what we saw in the simulations. As n increases, $E[W]$ approaches 1.