Chapter 9

Continuous Distributions

References: Pruim 3.1, Larsen & Marx 3.4

In this chapter, we will turn out attention to continuous random variables. A continuous random variable is a random variable which (in theory) can take on any real value in some interval.

So how do we define probabilities for continuous random variables. The fact that a continuous random variable takes uncountably many values eliminates the possibility of assigning a probability to each possible value, as we did in the discrete case.

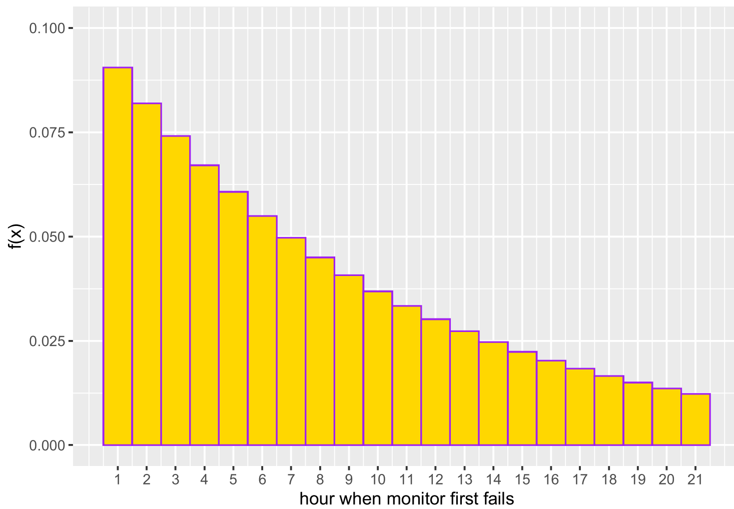
We return to the key idea behind probability histograms:

$$area = probability.$$

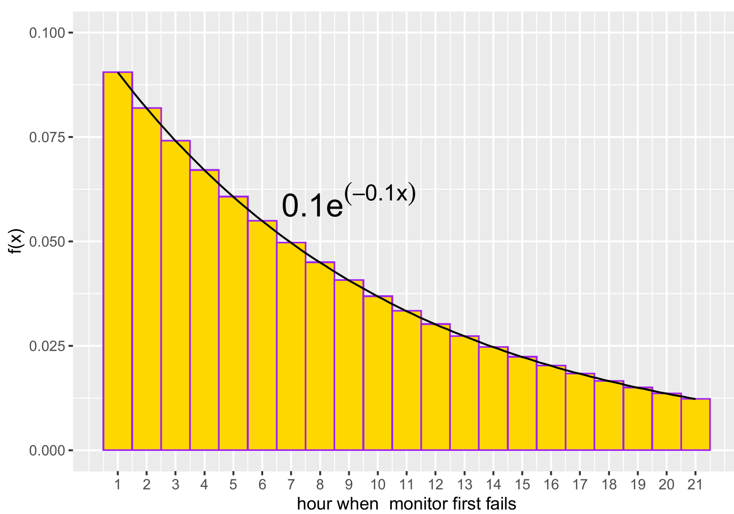
Let's consider an example. Suppose an electronic surveillance monitor is turned on briefly at the beginning of every hour and has a 0.905 probability of working properly, regardless of how long it has remained in service. If we let the random variable X denote the hour at which the monitor first fails, then:

$$\begin{aligned} f(x) &= P(X = x) = \text{P(Monitor works for x-1 hours and then fails)} \\ &= (0.905)^{x-1} \cdot (0.095), \qquad x = 1, 2, 3, ... \end{aligned}$$

The figure below shows the probability histogram of X. Here the height of the bar is $f(x)$, and since the width of the bar is one, the area of the bar is also equal to $f(x)$.



Now take a look at the graph below, where the exponential curve $0.1e^{-0.1x}$ is superimposed on the graph. Notice how closely the area under the curve approximates the area of the blocks. It follows that the probability that X lies in some given interval is numerically similar to integral of the exponential curve above that same interval.



Implicit in the similarity here between $f(x)$ and the exponential curve $y = 0.1 \cdot e^{-0.1x}$ is our sought-after alternative to probability assignments for continuous sample spaces. Instead of defining probabilities for individual points, we will define probabilities for intervals of points, and those probabilities will be areas under the graph of some function (for example: $y = 0.1 \cdot e^{-0.1x}$), where the shape of the function will reflect the desired probability "measure" to be associated with the sample space

Definition 9.1. A **Probability Density Function** (PDF) is any function f such that

$$f(x) \geq 0, \forall\ x \in \mathbb{R}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

The continuous random variable X defined by the PDF $f(x)$ satisfies

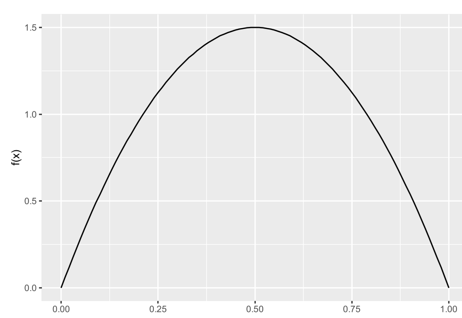
$$P(a \leq X \leq b) = \int_a^b f(x)\, dx.$$

The set of real numbers where the PDF is strictly positive - the range of the random variable - is called the <u>support</u> of the distribution.

. . . .

Example 9.1. Suppose we would like a continuous random variable X to "select" a number between 0 and 1 in such a way that intervals near the middle of the range would be more likely to be represented than intervals near either 0 or 1. One function having that property is:

$$f(x) = 6x(1-x), 0 \leq x < 1.$$



Verify that this function is a "legitimate" PDF and use it to calculate

- $P(0 \leq X \leq \frac{1}{4})$
- $P(\frac{1}{2} \leq X \leq \frac{3}{4})$.

A PDF is not a probability. For instance, it would be incorrect to say

$$f(x) = P(X = x).$$

However, so long as f is continuous at x, you can think of $f(x)\,dx$ as approximately equal to $P(X = x)$ for some very small dx.

The following lemma demonstrates one way in which continuous random variables are very different from discrete random variables.

Lemma 9.1. *Let X be a continuous random variable with PDF f. Then for any real number a:*

- $P(X = a) = 0$
- $P(X < a) = P(X \leq a)$
- $P(X > a) = P(X \geq a)$

Associated with every random variable, discrete or continuous, is a cumulative distribution function (CDF). For discrete random variables, the CDF is a non-decreasing step function, where the "jumps" occur at the values of x for which the PMF is non-zero. For continuous random variables, the CDF is defined just as it was for a discrete random variable, but we use an integral rather than a sum to get the CDF from the PDF.

Definition 9.2. Let X be a continuous random variable with PDF f. Then the **Cumulative Distribution Function** (CDF) for X is

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

for any $x \in \mathbb{R}$.

. . . .

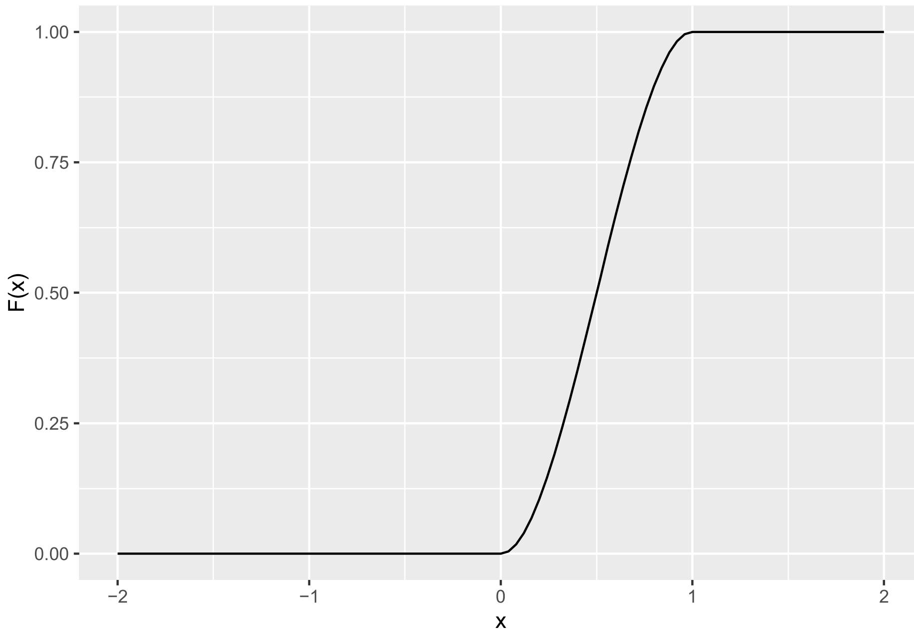
Example 9.2. Determine the CDF for the random variable from example 9.1.

$$f(x) = 6x(1-x), \quad 0 \leq x < 1.$$

A plot of the CDF we just derived is shown below. It is S-shaped which is actually a rather common shape for the CDF of a continuous random variable to have.

```
f <- function(x){
          ifelse(x < 0, 0,
                  ifelse(x<= 1, x^2*(3-2*x), 1) )
     }

ggplot() +
  geom_function(fun = f, xlim = c(-2,2) ) +
              labs(x = "x",y = "F(x)")
```



. . . .

Definition 9.2 should remind us of the Fundamental Theorem of Calculus. When F is defined as the integral of f, then f is the derivative of F. There is a slight difference however, because the Fundamental Theorem is usually presented only for continuous integrands f, but densities need not be continuous. However, the worst that can happen is the derivative of F may not exist at points where f is discontinuous, but we can still say the following.

Lemma 9.2. *Suppose $F(x)$ is the CDF of a continuous random variable X with PDF f. Then*

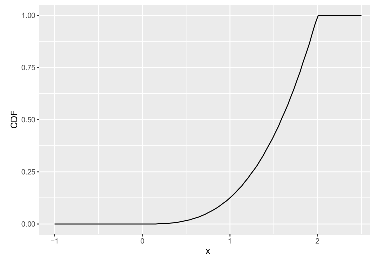
$$f(x) = \frac{d}{dx}F(x)$$

is a PDF for X.

. . . .

Example 9.3. Let

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3/8 & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$



Find a PDF for X.

. . . .

Lemma 9.2 implies that we can define random variables by specifying a CDF rather than a PDF. In fact, for continuous distributions, the CDF is the more important of the two functions, since probabilities are completely determined by it.

Definition 9.3. Two random variables X and Y are said to be **identical in distribution**, denoted

$$X \stackrel{d}{=} Y$$

if their CDFs are the same. That is,

$$F_X(x) = F_Y(x) \quad \forall x.$$

It is possible for two different PDFs to give rise to the same CDF. For example

$$f(x) = \frac{3x^2}{8}, 0 \le x \le 2$$

is a different PDF from the one we derived in example 9.3 which would give rise to the same CDF.

9.1 Practice Problems

1. Find the value of c that makes $f(x)$ below a valid PDF.

$$f(x) = \begin{cases} c\sin(x) & 0 < x < \pi/2 \\ 0 & otherwise \end{cases}$$

2. Suppose X, the lifetime in hundreds of hours of a randomly chosen bulb made in a certain factory, has PDF

$$f(x) = \frac{1}{10}e^{-x/10}x > 0.$$

   Find the following probabilities:

   (a) $P(X \geq 4)$

   (b) $P(X \geq 4|X \geq 2)$