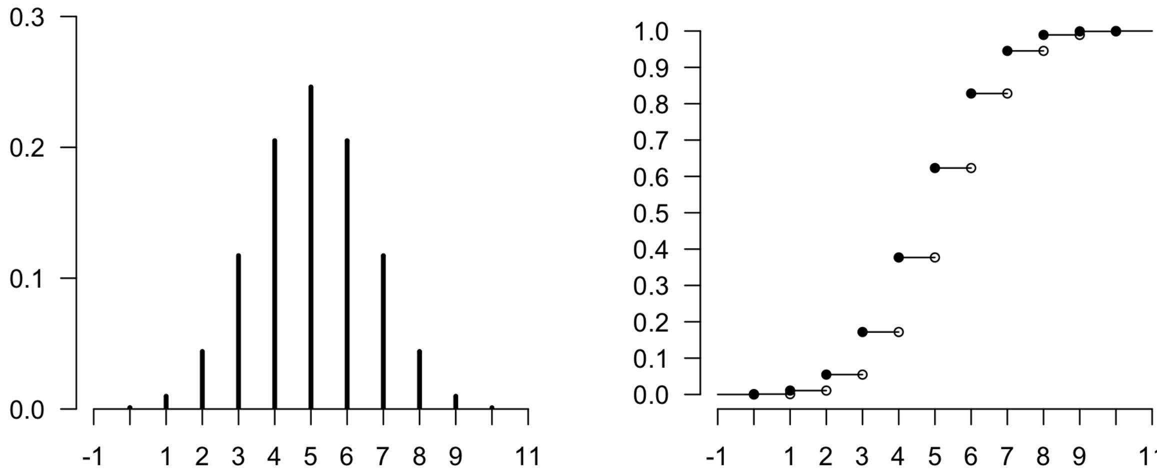

Figure 3.12

3.4 The cumulative distribution function

Thus far we've studied discrete random variables: random quantities X where the set of possible values is finite or countably infinite. In order to specify the distribution of a discrete random variable, we have two options:

- (**pmf**) for each value $x \in \text{Range}(X)$, specify its individual probability $P(X = x)$. If the range is small enough, these can simply be listed out in a table. Otherwise, this information must be summarized in a *function* called the **probability mass function**:

$$p_X(x) = \begin{cases} P(X = x) & x \in \text{Range}(X) \\ 0 & \text{else.} \end{cases}$$

- (**cdf**) define a function $F_X : \mathbb{R} \to [0, 1]$ that returns $F_X(x) = P(X \leq x)$ for any $x \in \mathbb{R}$.

Figure 3.12 visualizes these options, and as we saw in Section 3.2, these are fundamentally two sides of the same coin: given one, you can always recover the other, either summing up the pmf to get the cdf, or differencing the cdf to get the pmf. Having said that, we have yet to define a random variable primarily in terms of its cdf. We have worked exclusively with the pmf, and only occasionally visualized the cdf as a curious byproduct that did not see much action. But what was once a bit player will now become the star as we turn our attention to other types of random variables beyond the discrete case.

In order to fully specify any random variable, regardless its type, we must ultimately write down two objects:

- (**range**) the set of all possible values $\text{Range}(X) \subseteq \mathbb{R}$;

- (**distribution**) a probability measure $A \mapsto P(X \in A)$.

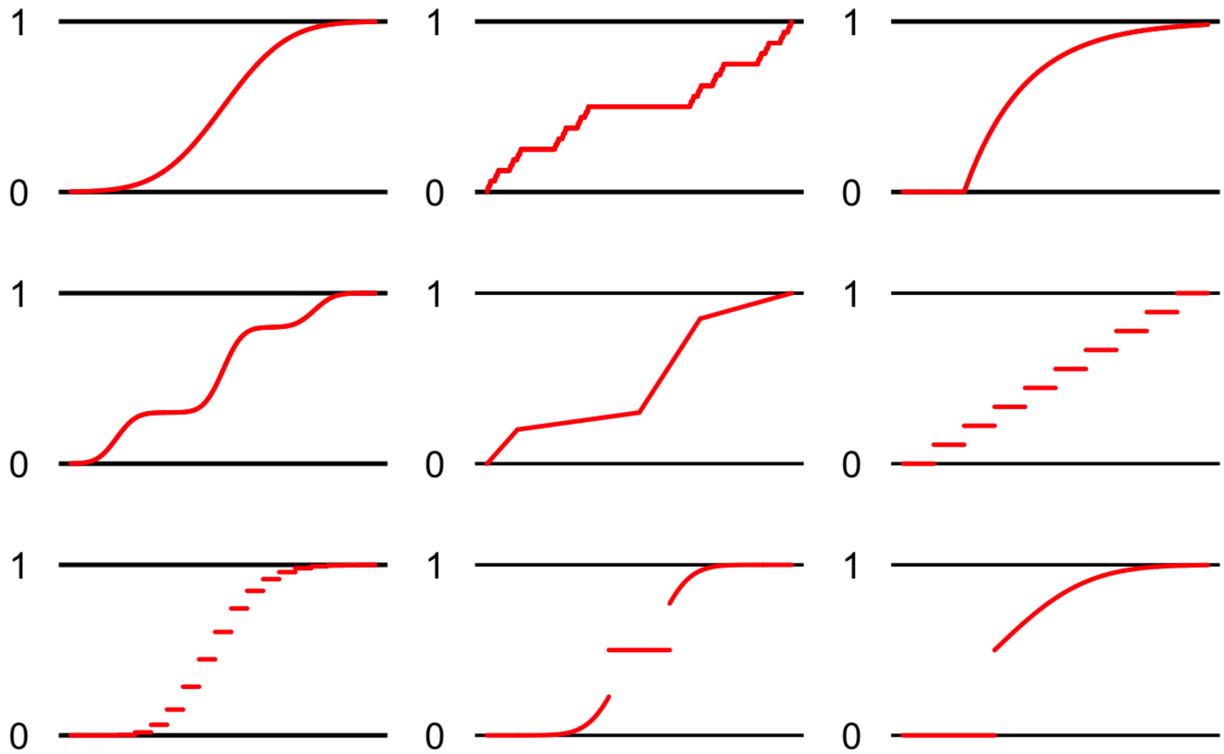

Figure 3.13: Valid cdfs come in all shapes and sizes

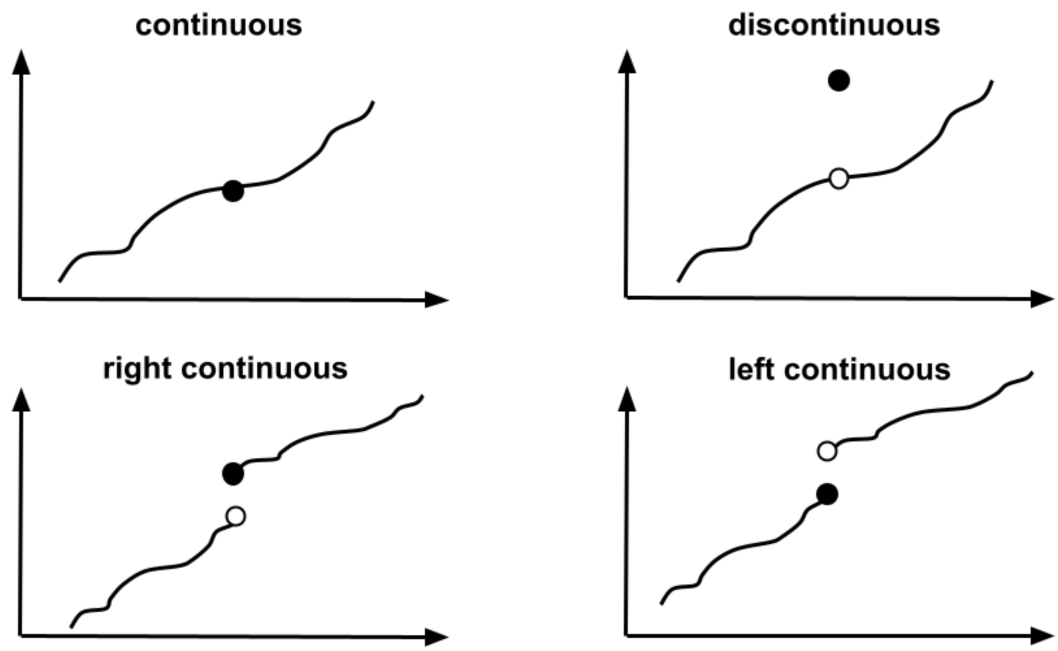

Figure 3.14: Types of continuity for a function $f : \mathbb{R} \to \mathbb{R}$

The question then becomes: *how*? Merely as a technical matter, how do we pull this off? Unlike in a first calculus course where we can write down functions $f(x) = x^2$ that take real numbers and return real numbers, it is not straightforward to write down a simple "formula" that summarizes how *any* set $A \subseteq \mathbb{R}$ gets assigned its probability $P(X \in A)$. And even if it were, how do we ensure that the resulting map is a valid probability measure that satisfies all of our axioms? It turns out that the universal answer to the question "how do we write down the distribution of *any* random variable" is "write down its cdf." All random variables, no matter how pathological, possess a cdf, and the cdf is sufficient for computing the entire distribution, which we will now see.

Recall the four properties of a cumulative distribution function:

- **nondecreasing**: for any $x \leq y$, it must be that $F_X(x) \leq F_X(y)$;
- **right continuous**: F_X has all its limits as you approach *from the right*:

$$\lim_{x \to a^+} F_X(x) = F_X(a).$$

  This is visualized in Figure 3.14.

- **goes to zero**: $\lim_{x \to -\infty} F_X(x) = 0$;
- **goes to one**: $\lim_{x \to \infty} F_X(x) = 1$.

Every cdf you will ever encounter for the rest of your life satisfies these four properties...*at least*. Figure 3.13 displays cartoons of what a valid cdf might look like. How is the cdf sufficient for computing the entire distribution of a random variable? That is, given *any* set $A \subseteq \mathbb{R}$, how does the cdf allow us to compute the number $P(X \in A)$? To show this, we start with intervals of the form $A = (a, b]$ for $a < b$:

Theorem 3.7. Let X by any random variable with cdf $F_X(x) = P(X \leq x)$, and let $A = (a, b]$ for any real numbers $a < b$. Then

$$P(X \in A) = P(a < X \leq b) = F_X(b) - F_X(a). \tag{3.14}$$

Proof. Let hilarity ensure:

$$\begin{aligned}
P(X \in A) &= P(a < X \leq b) \\
&= P(a < X \text{ and } X \leq b) \\
&= P(a < X) + P(X \leq b) - P(a < X \text{ or } X \leq b) && \text{(inclusion/exclusion)} \\
&= P(a < X) + P(X \leq b) - P(X \in \mathbb{R}) \\
&= 1 - P(X \leq a) + P(X \leq b) - P(X \in \mathbb{R}) && \text{(complement rule)} \\
&= 1 - P(X \leq a) + P(X \leq b) - 1 && \text{(axiom: total measure 1)} \\
&= P(X \leq b) - P(X \leq a) \\
&= F_X(b) - F_X(a).
\end{aligned}$$

□

So, computing the distribution of X for a generic interval $A = (a, b]$ can always be collapsed to a cdf calculation. But what about sets of other forms: (a, b), $(0, \infty)$, $\{c\}$, etc? We will not offer a formal proof, but any "ordinary-looking" subsets of $\mathbb{R}$ like these can ultimately be expressed as limits, unions, intersections, or complements of intervals like $(a, b]$, and so the overall distribution $P(X \in \mathbb{R})$ for any set A can be computed using Theorem 3.7 together with our basic rules from Section 2.3.

In this way then, we see that "the cdf is queen." It is a universal object. All random variables have one. It fully characterizes the distribution of the random variable. The cdf is sufficient in principle for computing $P(X \in A)$ for any set A. But if it's so important, why did we hardly ever mention it in Section 3.2? While the cdf is crucially important theoretically, its practical use can be limited. In practice, we would like to have *alternative* characterizations of a distribution that are easier to work with and more interpretable. The precise nature of that alternative characterization depends on the *type* of the random variable. As we saw, when the random variable is of discrete type for instance, we can characterize the distribution with the pmf, and we got a lot of mileage out of this option.

So what other types of random variables are there? We know that the cdf is queen, and it possesses its four universal properties: nondecreasing, *right* continuous, goes to 0 as $x \to -\infty$, goes to 1 as $x \to \infty$. Those apply to all random variables, period end of story I don't care what. As such, we can start to classify random variables according to any *extra* properties their cdf may possess. For example...

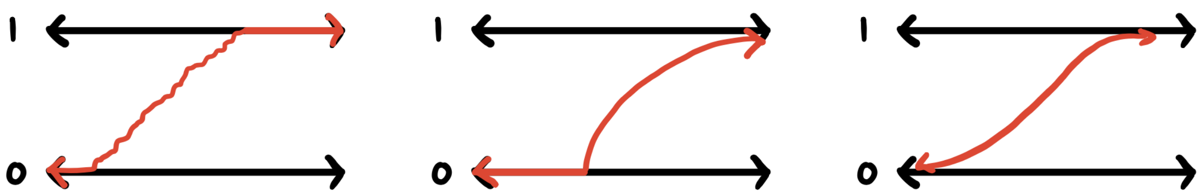

Figure 3.15: Cartoons of continuous cdfs

3.5 Continuous random variables

Definition 3.12. A random variable X is **continuous** if its cdf $F_X(x) = P(X \leq x)$ is a continuous function for all $x \in \mathbb{R}$.

So stronger than being merely right-continuous, the cdf is now just plain *continuous*, meaning that the value of F_X at a point a is equal to its limit at that point regardless which direction we approach from:

$$\lim_{x \to a} F_X(x) = F_X(a).$$

Figure 3.15 displays cartoons of continuous cdfs, which do not have any holes or jumps. One of the consequences of this is that a continuous random variable will now have a range which is an *uncountably* infinite set, like $(0,\ 1)$, $[0,\ \infty)$, or all of $\mathbb{R}$. But because of this, we must learn to accept a jarringly uninuintive property:

Theorem 3.8. If X is a continuous random variable, then $P(X = c) = 0$ for any $c \in \mathbb{R}$.

Proof.

$$\begin{aligned}
P(X = b) &= \lim_{a \to b} P(a < X \leq b) \\
&= \lim_{a \to b} [F_X(b) - F_X(a)] \\
&= \lim_{a \to b} F_X(b) - \lim_{a \to b} F_X(a) \\
&= F_X(b) - F_X(b) && \text{(cdf is continuous at } b\text{)} \\
&= 0.
\end{aligned}$$

□

Think about a random variable X with $\text{Range}(X) = \mathbb{R}$. There are simply "too many" real numbers for us to assign non-zero probability to each one and have these "add up" to one. So if we wish to place a valid probability measure on all of $\mathbb{R}$, we have to accept Theorem 3.8.

One quick, practical consequence of Theorem 3.8 is that it does not matter whether or not you include the endpoints when you compute the probability that a continuous random variable is in some interval:

Corollary 3.9. If X is a continuous random variable, then

$$P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = P(a < X < b).$$

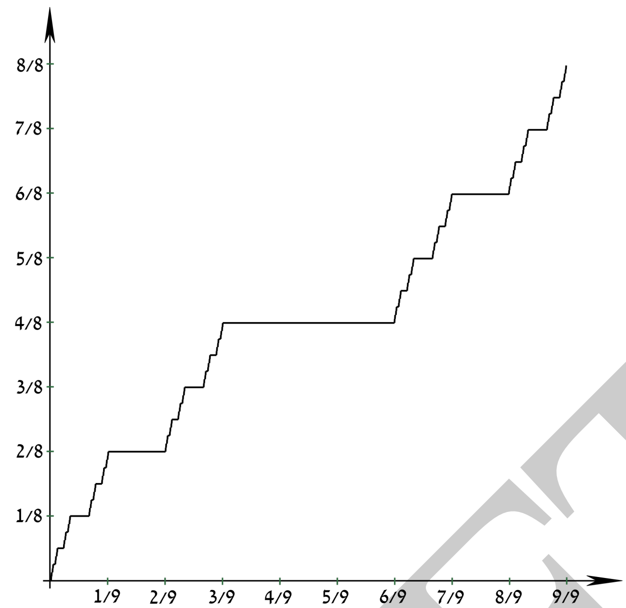

Figure 3.16: The Cantor distribution is an example of a *singular* distribution. This cdf is continuous everywhere, but it is not *absolutely* continuous, and the derivative is equal to zero (almost) everywhere, so it does not have a probability density function.

Partial proof. We won't itemize all the cases, but as an example, consider $[a, b]$. This can be rewritten as a disjoint union: $[a, b] = \{a\} \cup (a, b) \cup \{b\}$. So by countable additivity $P(a \leq X \leq b) = P(X = a) + P(a < X < b) + P(X = b)$. We know from Theorem 3.8 that $P(X = a) = P(X = b) = 0$, so $P(a \leq X \leq b) = P(a < X < b)$. □

Continuity is a nice property to require because it enforces a certain amount of smoothness, but it is ultimately a fairly weak condition. It does not rule out functions like the Cantor distribution displayed in Figure 3.16, which is just plain weird. In practice, when working with continuous RVs, we require an even stronger form of smoothness...

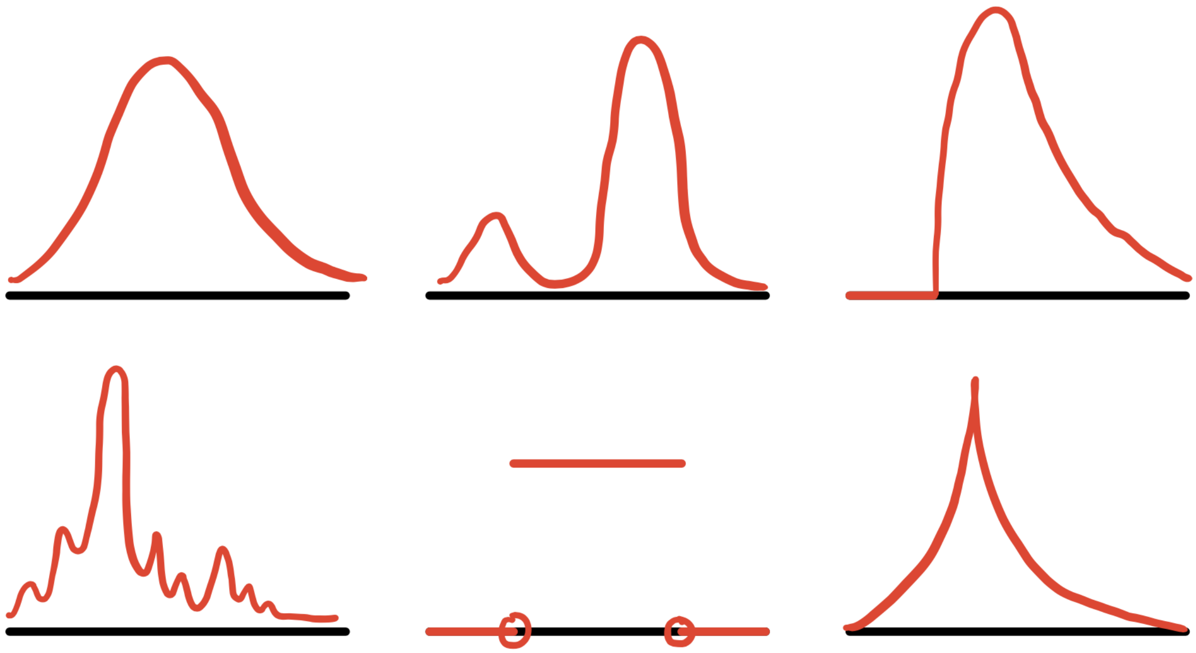

Figure 3.17: Cartoons of what a probability density function might look like. Area under the curve corresponds to probabilities. The total area underneath the pdf must be one.

3.6 Absolutely continuous random variables

Definition 3.13. A random variable X is **absolutely continuous** if its cdf $F_X(x) = P(X \leq x)$ is an absolutely continuous function for all $x \in \mathbb{R}$.

The concept of absolute continuity is beyond the scope of our course. You can look it up if you wish, but all I want you to know is that absolute continuity is a special kind of continuity that is stronger, and it ensures that the cdf is sufficiently smooth as to rule out the pathological cases

By ruling out such cases and requiring the cdf to be extra smooth, we guarantee the existence of a special object, which is the continuous analog of the pmf

Theorem 3.10. If X is an absolutely continuous random variable, then there exists a function $f_X : \mathbb{R} \to \mathbb{R}$ called the **probability density function** (**pdf**) with the following properties:

- $f_X(x) \geq 0$ for all $x \in \mathbb{R}$;
- $F_X'(x) = f_X(x)$;
- $F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$;
- $P(a < X < b) = \int_a^b f_X(x)\,\mathrm{d}x = F_X(b) - F_X(a)$ (fundamental theorem of calculus!);
- $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = P(-\infty < X < \infty) = 1$;

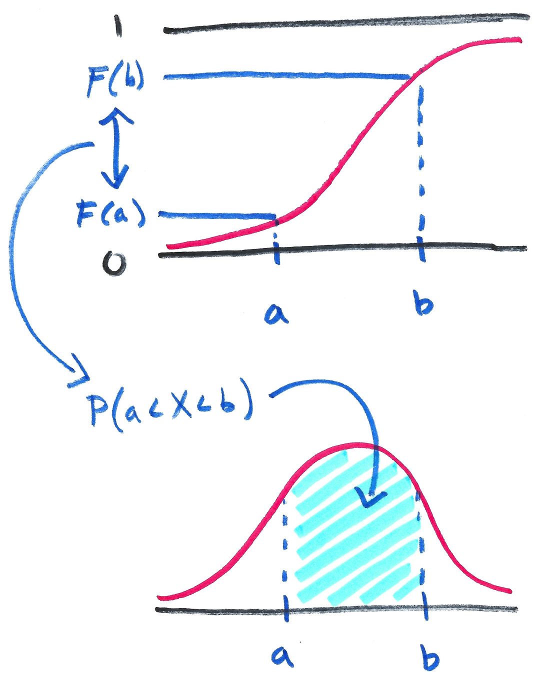

Figure 3.18: $P(a < X < b) = \int_a^b f_X(x)\,\mathrm{d}x = F_X(b) - F_X(a)$

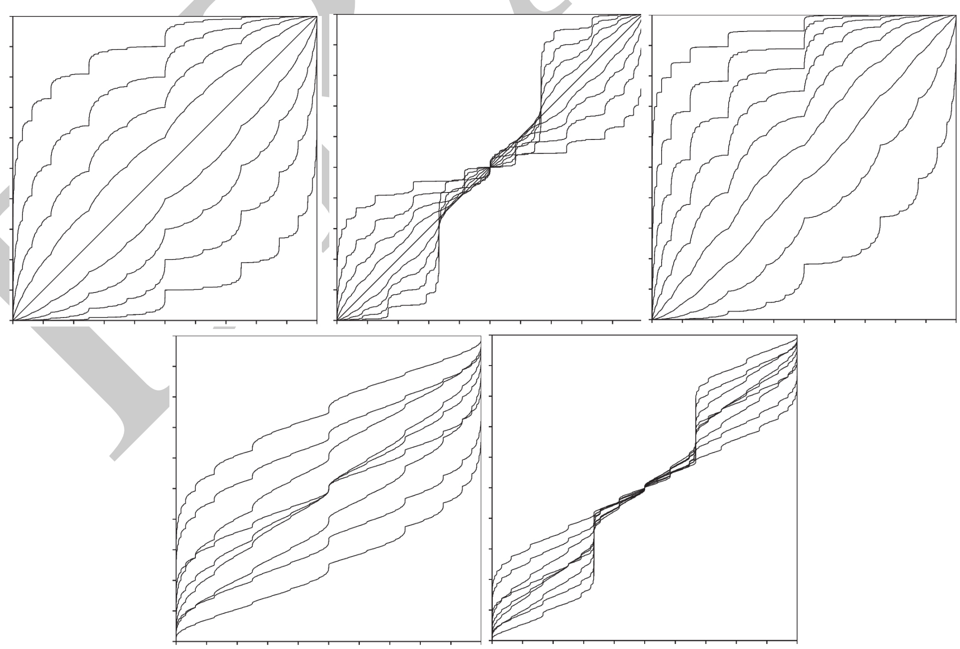

Figure 3.19: Examples of singular distributions. NEED TO CITE WHERE I GOT THESE

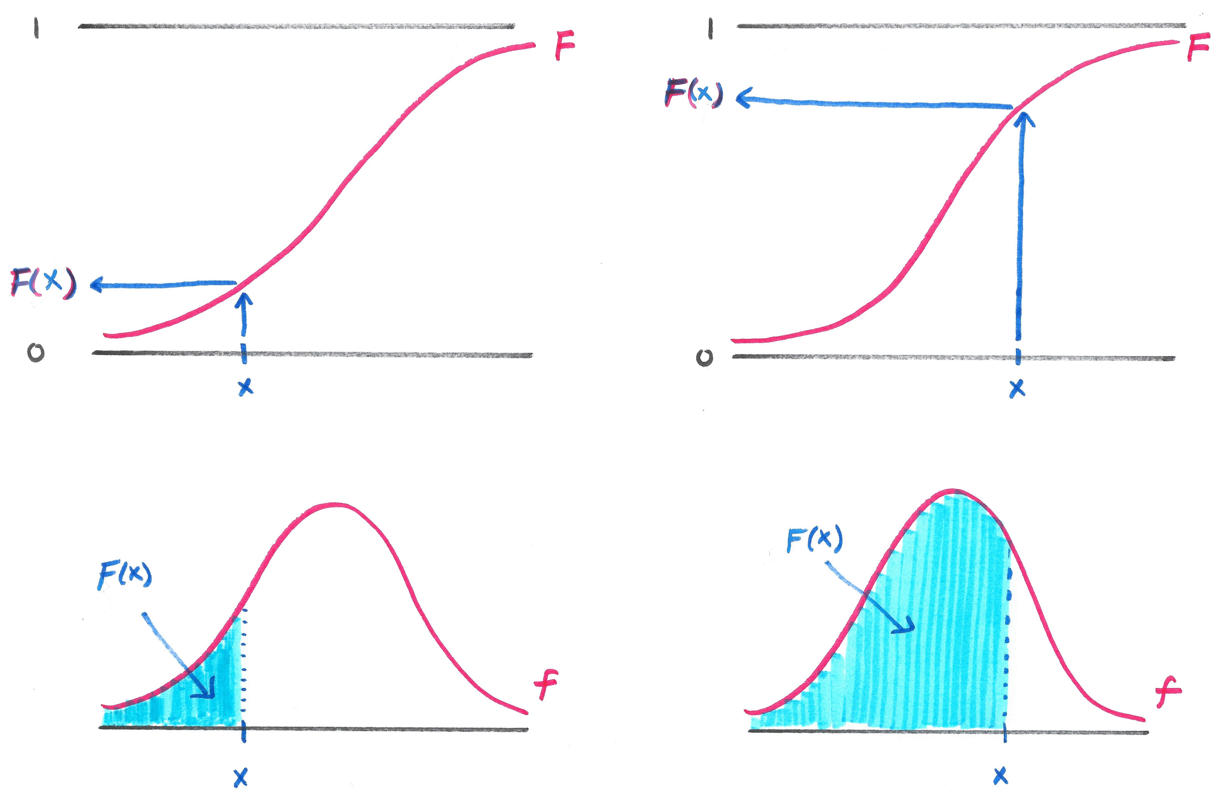

Figure 3.20: the cdf F is tracking the amount of area (probability) under the pdf f that we accumulate as we move from left to right. When we start on the left with "$x = -\infty$," we've accumulated none of it, and as $x \to \infty$, we eventually accumulate all of it (total measure 1).

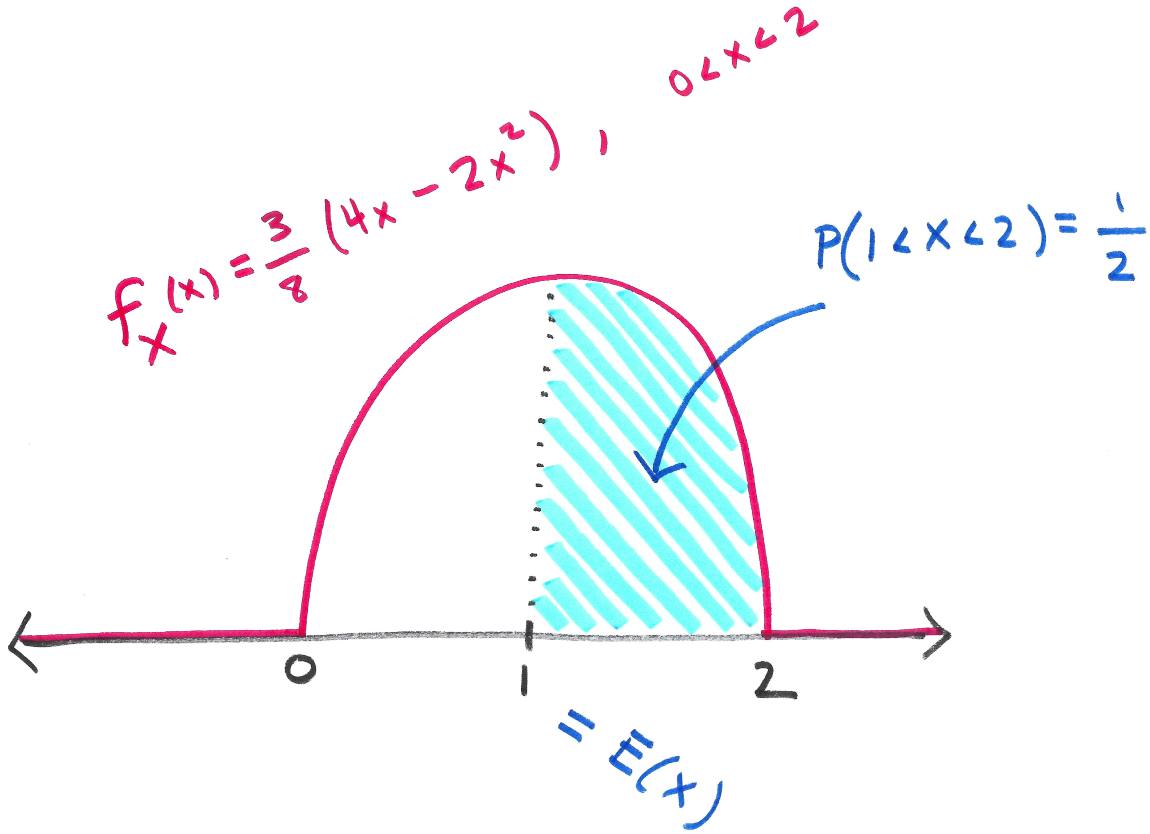

Figure 3.21

Figure 3.17 displays cartoons of what this function might look like, and Figure 3.18 illustrates how probabilities are calculated.

Remark 3.9. Absolutely continuous random variables are the continuous analog to discrete random variables, and the pdf plays the role of the pmf. In discrete world, we *summed* the pmf to compute probabilities. In continuous world, we *integrate* the pdf. In discrete world, we *differenced* the cdf to compute the pmf. In continuous world, we *differentiate* it.

Remark 3.10. The relationship $F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$ means that the cdf is the **area accumulation function** of the pdf. This is illustrated in Figure 3.20.

Remark 3.11. I hasten to emphasize: not all continuous random variables have a pdf. Random variables whose cdf is continuous but not absolutely continuous are called **singular**, and Figure 3.19 displays some examples of what that might look like. The Cantor distribution in Figure 3.16 is the most famous example of this. The derivative of the Cantor distribution is zero (almost) everywhere, so a density with the above properties could not possibly exist. When the pdf does exist however, it's super useful, because it allows us to compute probabilities and moments by *integrating*.

Example 3.16. Let X be absolutely continuous with density

$$f_X(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{else.} \end{cases}$$

A cartoon is displayed in Figure 3.21. The range of X is the set of real numbers where the density is nonzero. This is called the **support** of the density function:

$$\operatorname{supp}(f_X) = \{x \in \mathbb{R} : f_X(x) > 0\}, \tag{3.15}$$

and so $\operatorname{Range}(X) = \operatorname{supp}(f_X)$ for absolutely continuous random variables, and in this case we see that $\operatorname{Range}(X) = (0, 2)$. And recall that because X is continuous, we need not quibble about whether or not we include the endpoints. The constant c is a **normalizing constant** that serves to ensure that the density integrates to 1, so let's find its value. First, note that

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x &= \int_{-\infty}^{0} f_X(x)\,\mathrm{d}x + \int_0^2 f_X(x)\,\mathrm{d}x + \int_2^{\infty} f_X(x)\,\mathrm{d}x \\ &= \cancel{\int_{-\infty}^{0} 0\,\mathrm{d}x} + \int_0^2 c(4x - 2x^2)\,\mathrm{d}x + \cancel{\int_2^{\infty} 0\,\mathrm{d}x} \\ &= c\int_0^2 (4x - 2x^2)\,\mathrm{d}x \end{aligned}$$

So

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1 &\implies c\int_0^2 (4x - 2x^2)\,\mathrm{d}x = 1 \\ &\implies c = \frac{1}{\int_0^2 (4x - 2x^2)\,\mathrm{d}x}. \end{aligned}$$

As such:

$$\int_0^2 (4x - 2x^2)\,\mathrm{d}x = \left[2x^2 - \frac{2}{3}x^3\right]_0^2 = 2\cdot 4 - \frac{2}{3}\cdot 8 = 8/3.$$

So $c = 1/(8/3) = 3/8$. If we wanted to compute $P(X > 1)$, we can do that by integrating the pdf:

$$\begin{aligned} P(X > 1) &= P(X \in (1, \infty)) \\ &= \int_1^{\infty} f_X(x)\,\mathrm{d}x \\ &= \int_1^2 f_X(x)\,\mathrm{d}x + \int_2^{\infty} f_X(x)\,\mathrm{d}x \\ &= \int_1^2 \frac{3}{8}(4x - 2x^2)\,\mathrm{d}x + \cancel{\int_2^{\infty} 0\,\mathrm{d}x} \\ &= \frac{3}{8}\left[2x^2 - \frac{2}{3}x^3\right]_1^2 \\ &= \frac{3}{8}\left[2\cdot 4 - \frac{2}{3}\cdot 8 - 2 + \frac{2}{3}\right] \\ &= \frac{1}{2}. \end{aligned}$$

Definition 3.14. If X is absolutely continuous with density f, then the expected value of X is

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x. \tag{3.16}$$

As in the discrete case, the expected value may not exist or be finite.

Example 3.17. Continuing Example 3.16, We see that

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{0} x f_X(x)\,\mathrm{d}x + \int_{0}^{2} x f_X(x)\,\mathrm{d}x + \int_{2}^{\infty} x f_X(x)\,\mathrm{d}x \\
&= \cancel{\int_{-\infty}^{0} x \cdot 0\,\mathrm{d}x} + \int_{0}^{2} x \frac{3}{8}(4x - 2x^2)\,\mathrm{d}x + \cancel{\int_{2}^{\infty} x \cdot 0\,\mathrm{d}x} \\
&= \frac{3}{8}\int_{0}^{2} (4x^2 - 2x^3)\,\mathrm{d}x \\
&= \frac{3}{8}\left[\frac{4}{3}x^3 - \frac{2}{4}x^4\right]_0^2 \\
&= \frac{3}{8}\left[\frac{4}{3}8 - \frac{2}{4}16\right] \\
&= 1.
\end{aligned}$$

It makes sense that this is the mean because the pdf is symmetric about 1.

Theorem 3.11. (**LOTUS, again**) If X is absolutely continuous with density f_X and $g : \mathrm{Range}(X) \to \mathbb{R}$ is a transformation, then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x. \tag{3.17}$$

The upshot of this result is the same as before: in order to compute $E[g(X)]$, it is not necessary to first derive the entire density of the transformed variable $g(X)$. We can make due with the original density f_X. All of the results in Section 3.3 about the linearity of expectation and the variance apply unmodified to continuous and absolutely continuous random variables. They apply to all random variables regardless of type, in fact:

- $E(aX + b) = aE(X) + b$;
- $E\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i E(X_i)$, no matter how the X_i are related;
- $\mathrm{var}(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$;
- $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$.
- If the X_i are independent, then $\mathrm{var}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 \mathrm{var}(X_i)$.