III. General Random Variables

Continuous random variables

So far, we have only been concerned with random variables that take a **discrete** set of values (that is, the possible outcomes our countable, and can be indexed by a subset of the integers).

Of course, many things in "real life" are more naturally modeled as continuous:

- the velocity of a car
- the amount of time a router waits until it sees another packet
- the location where a dart I'm about to throw will land
- the height of the next person I encounter

All of the concepts and methods we have learned about for discrete random variables (e.g., pmfs, expectation, conditioning) have continuous counterparts. However, the continuous case introduces subtleties that we have to consider carefully.

First and foremost among these is that the idea of a "probability mass function" does not make sense for continuous random variables. Remember that for a discrete random variable X, we defined

$$p_X(k) = \mathrm{P}\left(\{X = k\}\right).$$

If X is continuous, the probability that it takes any particular value exactly is zero; now $\mathrm{P}\left(\{X = x\}\right) = 0$ for all x, and so this quantity is no longer informative.

We characterize continuous-values random variables using a *density function* in place of the mass function.

Probability density functions (pdfs)

The **probability density function** $f_X(x)$ of a continuous random variable is the function which obeys

$$\mathrm{P}\left(X \in B\right) = \int_B f_X(x)dx$$

for "every"[1] subset $B \subseteq \mathbb{R}$.

For example:

$$\mathrm{P}\left(a \leq X \leq b\right) = \int_a^b f_X(x)dx,$$

$$\mathrm{P}\left(X > -1\right) = \int_{-1}^{\infty} f_X(x)dx,$$

and for $a \leq b \leq c \leq d$ we have

$$\mathrm{P}\left(X \in [a,b] \cup [c,d]\right) = \int_a^b f_X(x)dx + \int_c^d f_X(x)dx,$$

etc.

Note that by necessity,

$$\mathrm{P}\left(-\infty < X < \infty\right) = 1,$$

and so

$$\int_{\infty}^{\infty} f_X(x) = 1,$$

[1]Technically, it can't be defined for every subset. The exceptions are important from a philosophical perspective, but are mostly so rare and arcane that they have no practical bearing. Making this completely formal requires an advanced course on measure theory, and is way beyond the scope of this course.

which is the **normalization property** for pdfs. We also need

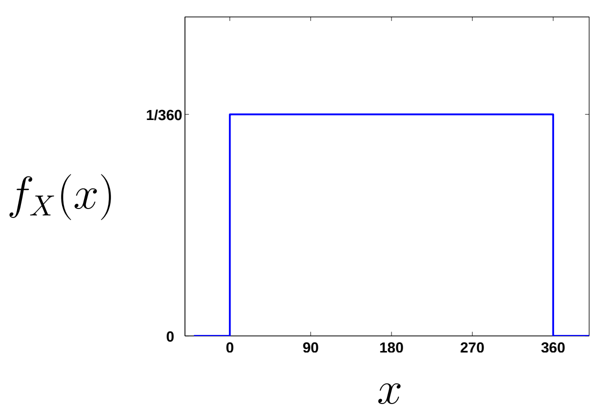
$$f_X(x) \geq 0$$

for all $x \in \mathbb{R}$ to avoid negative probabilities.

Note that $f_X(x)$ does **not** tell you $\mathrm{P}(X = x)$. As we mentioned above, for continuous random variables, $\mathrm{P}(X = x) = 0$ for any x. When dealing with continuous random variables, we can only assign probabilities to intervals of $\mathbb{R}$.

Example. You are blindfolded and then spun around multiple times. Let X be the direction you are facing (in degrees from North). A reasonable model for X would be to have a pdf of

$$f_X(x) = \begin{cases} \frac{1}{360} & 0 \leq x < 360 \\ 0 & \text{otherwise.} \end{cases}$$



What is the probability that you are within ± 10 of North?

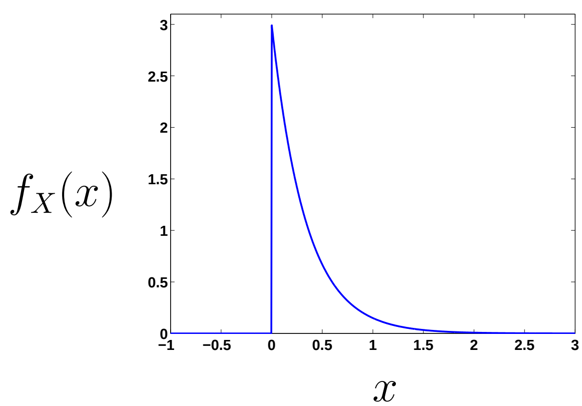
Example. Exponential densities are often used to model continuous waiting times like

- How long until this light bulb burns out?
- How long until the next train comes?

The pdf is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

for some parameter λ. Here is a sketch for $\lambda = 3$:



Note that

$$\int_{\infty}^{\infty} f_X(x)dx = \lambda \int_{\infty}^{\infty} e^{-\lambda x}dx = \frac{\lambda}{-\lambda}\left[e^{-\lambda x}\right]_0^{\infty} = -1[0-1] = 1.$$

What is the probability that $X \geq 3$?

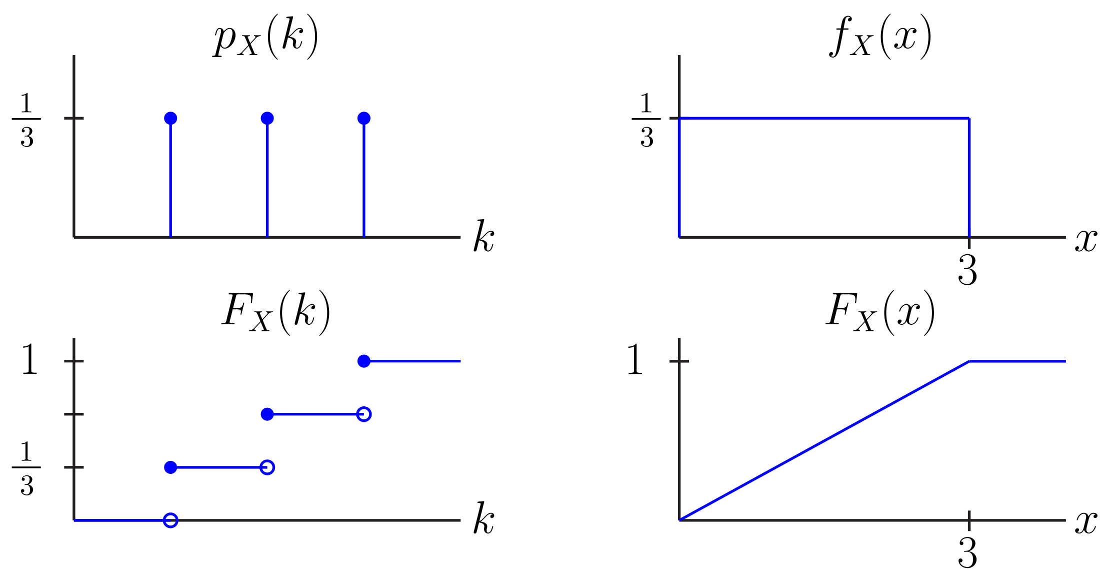
Cumulative distribution functions (cdfs)

The cumulative distribution function (cdf) is another way to characterize a random variable. The cdf of a random variable X is defined as:

$$F_X(x) = \mathrm{P}\left(X \le x\right).$$

This quantity is well-defined for both continuous and discrete random variables, and can be related to the mass/density function as follows:

$$F_X(x) = \begin{cases} \sum_{k \le x} p_X(k) & \text{for } X \text{ discrete} \\ \int_{-\infty}^{x} f_x(u)du & \text{for } X \text{ continuous.} \end{cases}$$



For continuous X, $F_X(x)$ is a continuous function of x, and by the Fundamental Theorem of Calculus,

$$f_X(x) = \frac{dF_X(x)}{dx},$$

i.e., the pdf is just the derivative of the cdf.

For a discrete random variable, the relationship between the cdf and pmf is

$$p_X(k) = F_x(k) - F_X(k-1).$$

Some important properties of a cdf F_X include:

- $F_X(x)$ is monotonically nondecreasing (as x increases, $F_X(x)$ will never decrease).
- $F_X(x) \to 0$ as $x \to -\infty$.
- $F_X(x) \to 1$ as $x \to \infty$.

Like the pdf, the cdf completely characterizes a continuous-valued random variable by giving us a way to calculate the probability of the random variable falling into a specified range. For example:

$$\mathrm{P}\left(a < X \leq b\right) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx,$$

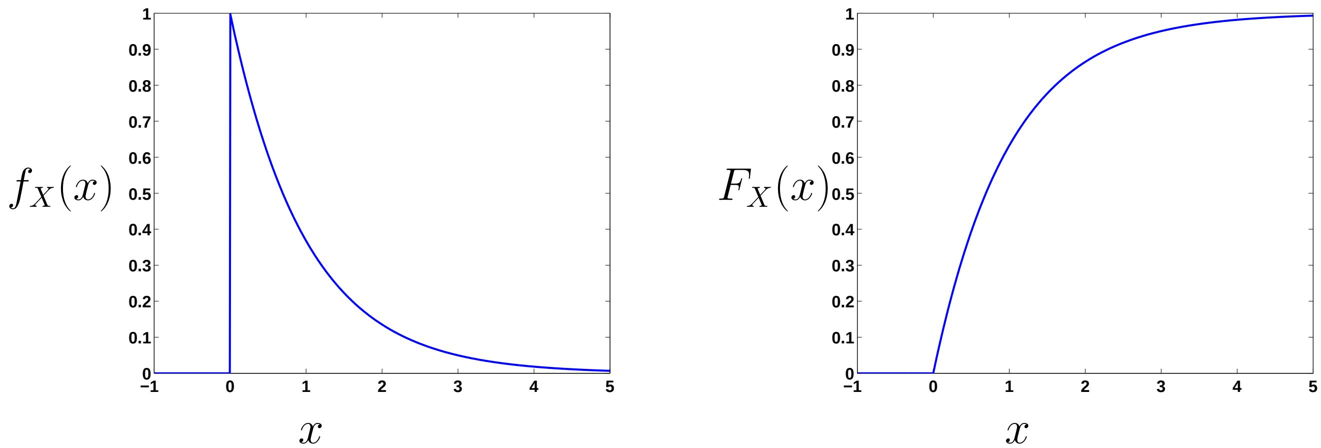
$$\mathrm{P}\left(X > a\right) = 1 - F_X(a) = \int_a^\infty f_X(x)dx.$$

and for $a \leq b \leq c \leq d$ we have

$$\mathrm{P}\left(X \in [a,b] \cup [c,d]\right) = \int_a^b f_X(x)dx + \int_c^d f_X(x)dx.$$

Example: Suppose X is an exponential random variable with $\lambda = 1$. Then

$$\text{pdf:}\quad f_X(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0 & x < 0. \end{cases}, \qquad \text{cdf:}\quad F_x(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$



What are

$$\mathrm{P}\left(X \leq 1\right) =$$

$$\mathrm{P}\left(1 < X \leq 5\right) =$$

$$\mathrm{P}\left(X > 2\right) =$$

Example: Suppose that X_1, X_2, and X_3 are independent random variables that are all uniform on $[0,1]$, and so

$$f_{X_i}(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases} \qquad F_{X_i}(x) = \begin{cases} 0, & x < 0 \\ x & 0 \le x \le 1, \\ 1 & x > 1, \end{cases} \qquad i = 1,2,3.$$

Let $Y = \max(X_1, X_2, X_3)$ be the **maximum** of these three random variables. What is the pdf of Y?

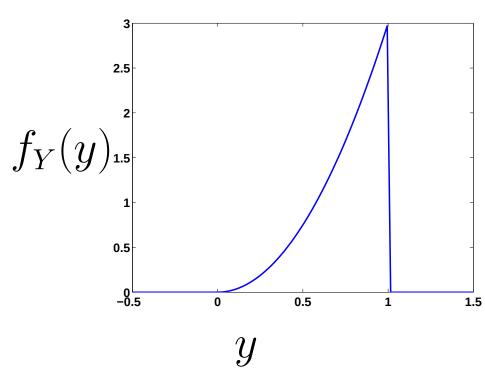
The answer to this question is not clear (at least to me). Clearly, $0 \le Y \le 1$, just like the X_i, but it is also clear that larger values of Y should be more likely than smaller values. The trick is to derive the pdf first finding the cdf, then taking the derivative.

The cdf is

$$\begin{aligned} F_Y(y) &= \mathrm{P}(Y \le y) \\ &= \mathrm{P}(\{X_1 \le y\} \cap \{X_2 \le y\} \cap \{X_3 \le y\}) \\ &= \mathrm{P}(X_1 \le y) \cdot \mathrm{P}(X_2 \le y) \cdot \mathrm{P}(X_3 \le y) \quad \text{(independence)} \\ &= F_{X_1}(y) \cdot F_{X_2}(y) \cdot F_{X_3}(y) \\ &= y^3, \quad \text{for } 0 \le y \le 1. \end{aligned}$$

Thus the pdf is

$$f_Y(y) = \begin{cases} 3y^2, & 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$



Expectation and variance

The expectation of a continuous random variable X is defined as

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Similarly,

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

and

$$\begin{aligned}\mathrm{var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \\ &= \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{\infty} x f_X(x) dx \right)^2\end{aligned}$$

The expectation and variance can still be viewed as a "weighted average" of the values that X (or $(X - \mathrm{E}[X])^2$) can take, but now since X (or $(X - \mathrm{E}[X])^2$) can take a continuum of possible values, our weighted sum simply becomes a weighted integral.

As before, if $Y = aX + b$ for fixed $a, b \in \mathbb{R}$, then

$$\mathrm{E}[Y] = a\,\mathrm{E}[Y] + b$$

and

$$\mathrm{var}(Y) = a^2\,\mathrm{var}(X).$$