LECTURE: CONTINUOUS RANDOM VARIABLES

1. Poisson Properties

Definition:

Y has a **Poisson distribution** with parameter $\lambda > 0$ if

$$p(y) = e^{-\lambda} \frac{\lambda^y}{y!} \qquad\qquad y = 0, 1, 2, \dots$$

We write $Y \sim \text{Poi}(\lambda)$

This roughly measures the number of events which occur during a fixed time interval (think 1 hour) where the average number of events is λ

Note that the Poisson distribution can have value 0, which corresponds to no events happening in the fixed span of time.

First, let's check that the sum of probabilities is 1:

$$\sum_{y=0}^{\infty} p(y) = \sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} = e^{-\lambda} \underbrace{\sum_{y=0}^{\infty} \frac{\lambda^y}{y!}}_{\text{Taylor series for } e^{\lambda}} = e^{-\lambda} e^{\lambda} = 1$$

Facts:

Let $Y \sim$ Poi (λ), then $E(Y) = \lambda$ and $\text{Var}(Y) = \lambda$

Why? For a Poisson Random Variable, we have

$$E(Y) = \sum_{y=0}^{\infty} yp(y) = \sum_{y=0}^{\infty} ye^{-\lambda}\frac{\lambda^y}{y!} = e^{-\lambda}\sum_{y=1}^{\infty}\lambda^y\frac{\cancel{y}}{\cancel{y}(y-1)!} = e^{-\lambda}\sum_{y=1}^{\infty}\frac{\lambda^y}{(y-1)!}$$

Here we used that the term at $y = 0$ is 0.

Now we let $z = y - 1$. Substituting this in, we get

$$E(Y) = e^{-\lambda}\sum_{z=0}^{\infty}\frac{\lambda^{z+1}}{z!} = \lambda e^{-\lambda}\sum_{z=0}^{\infty}\frac{\lambda^z}{z!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

The proof of the variance is more tedious and will be omitted

Example 1:

Customers arrive at your restaurant for dinner at an average rate of 5 customers per 15 minutes.

(a) What is the probability that no one arrives in a 15-minute period?

Let X be the number customers who arrive in a 15-minute period. Let's model X with a Poisson Distribution.

Then $X \sim \text{Poi}(5)$. If no customers arrive in that time, then $X = 0$

$$P(X = 0) = \frac{e^{-5}5^0}{0!} = e^{-5} \approx 0.0067$$

So there is less than 1% chance of this occurring.

(b) What is the probability that at least 3 customers arrive in a 15-minute period?

Here we want $P(X \geq 3)$, but using the complement, we get:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) = 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \frac{e^{-5}5^0}{0!} - \frac{e^{-5}5^1}{1!} - \frac{e^{-5}5^2}{2!} \\ &\approx 1 - 0.0067 - 0.0337 - 0.0842 = 0.08754 \end{aligned}$$

(c) What is the probability that exactly 5 customers arrive in a 15-minute period?

$$P(X = 5) = \frac{e^{-5}5^5}{5!} \approx 0.175$$

(d) What is the average number of customers in the 15-minute period

$$E(X) = \lambda = 5 \qquad \text{(just like in the problem ☺)}$$

2. Continuous Random Variables

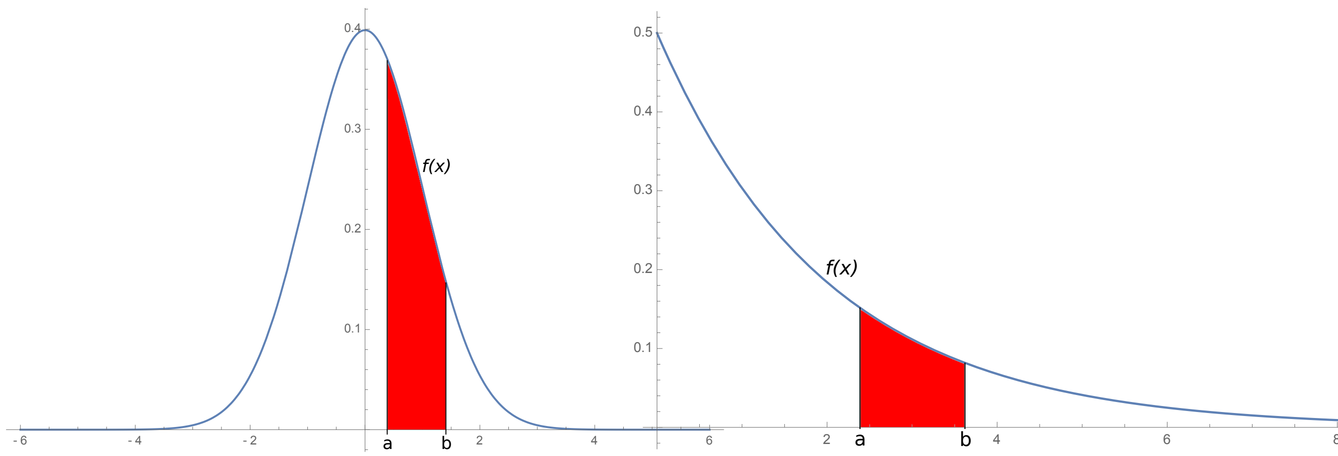
Many quantities in real life are not discrete is nature, like the amount of rainfall in a day or the lifespan of a washing machine.

Definition:

A random variable that takes on any real value is called a **continuous random variable**.

Note: This differs fundamentally with discrete random variables in the sense that we cannot calculate $P(Y = y)$ any more. In fact, that probability will be 0 (see below)

A continuous random variable is best described by a probability density function (pdf). Here are some examples of pdfs:



Definition:

The function $f(y)$ is a **probability density function (pdf)** for a continuous random variable Y if $f(y) \geq 0$ for all y and

$$\int_{-\infty}^{\infty} f(y)dy = 1$$

Note: This is the continuous analog of the sum of probabilities is 1, where sum is replaced by an intergal.

Definition:

In that case, the probability that Y is between a and b is the area under the density curve between a and b, i.e.

$$P(a \leq Y \leq b) = \int_a^b f(y)dy$$

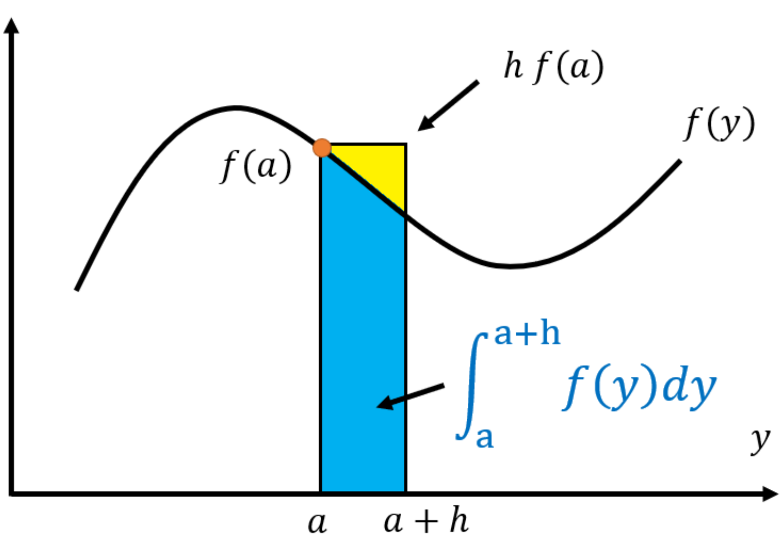
In the pictures above, this is illustrated in red.

Note: For a continuous distribution we have

$$P(Y = a) = \int_a^a f(y)dy = 0$$

This is different from discrete random variables where $P(Y = 3) = \frac{1}{2}$ for example

Interpretation of f:



If h is small then

$$P(a \leq Y \leq a+h) = \int_a^{a+h} f(y)dy \approx \text{ Area of Rectangle } = hf(a)$$

$$\text{Hence } f(a) \approx \frac{P(a \leq Y \leq a+h)}{h}$$

So $f(a)$ gives you the (small) probability that Y is in the tiny interval $[a, a+h]$ but scaled by the size h of that interval

Another difference is that for discrete random variables $P(8 \leq X \leq 12)$ and $P(8 < X \leq 12)$ are very different but for continuous ones, they are the same, since $P(X = 8) = 0$

Example 2:

$$\text{Let } f(y) = \begin{cases} cy^2 & 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the value of c for which $f(y)$ is a pdf

Notice $f(y) \geq 0$ and moreover

$$1 = \int_{-\infty}^{\infty} f(y)dy = \int_0^2 cy^2 dy = c\left[\frac{y^3}{3}\right]_0^2 = \frac{8}{3}c \Rightarrow c = \frac{3}{8}$$

(b) What is $P(1 \leq Y \leq 2)$?

$$P(1 \leq Y \leq 2) = \int_1^2 f(y)dy = \int_1^2 \frac{3}{8}y^2 dy = \frac{3}{8}\left[\frac{y^3}{3}\right]_1^2 = \frac{7}{8}$$

3. Cumulative Distribution Functions

Another way to describe a continuous random variable is with its cumulative distribution function:

Definition:

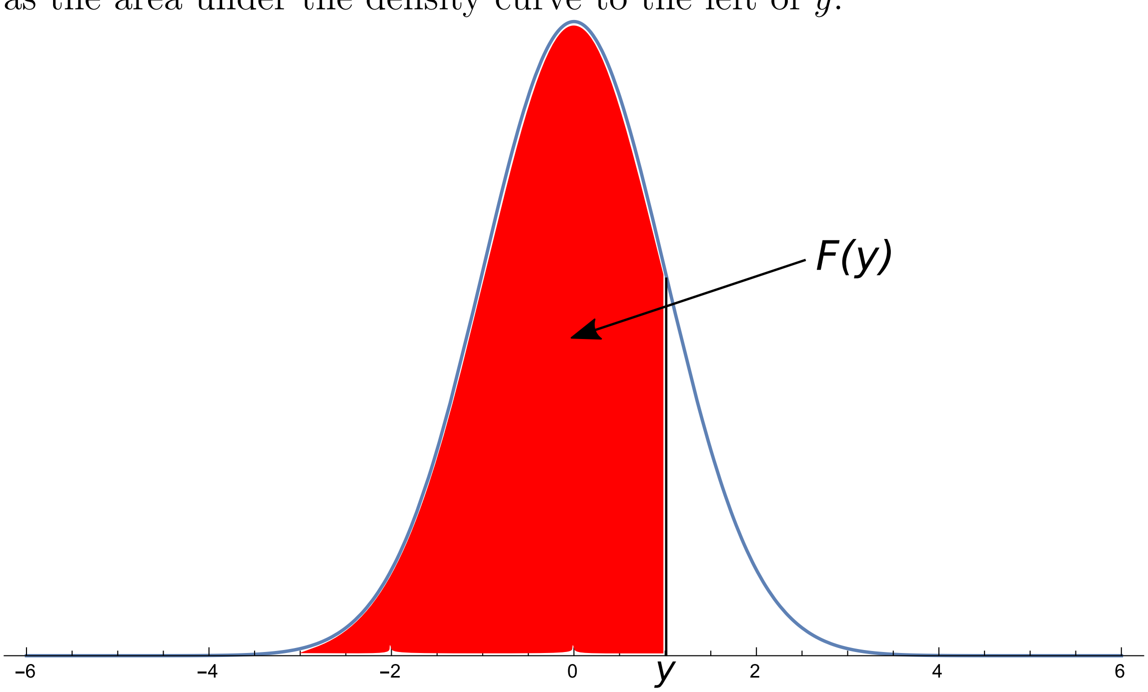
The **cumulative distribution function (cdf)** of Y is

$$F(y) = P(Y \leq y)$$

If Y has density function $f(y)$, then

$$F(y) = \int_{-\infty}^{y} f(z)dz$$

For a continuous random variable, this can be visualized graphically as the area under the density curve to the left of y.



Facts:

(1) $F(y)$ is a non-decreasing (think increasing)

(2) $\lim_{y\to-\infty} F(y) = 0$

(3) $\lim_{y\to\infty} F(y) = 1$

(4) $F(y)$ is continuous

For a continuous random variable, the cdf and pdf are related via the fundamental theorem of calculus.

Relationship between the cdf and pdf:

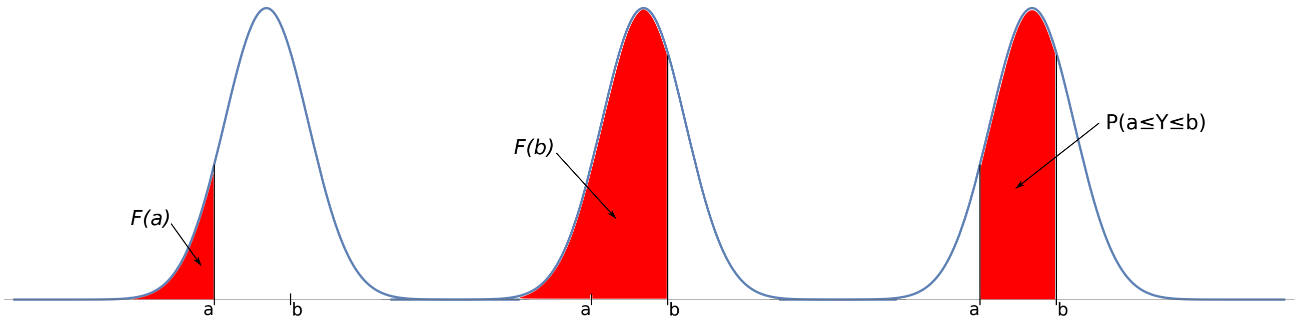
Let Y be a continuous random variable with pdf $f(y)$ and cdf $F(y)$. Then

(1)
$$F(y) = \int_{-\infty}^{y} f(z)dz$$

(2)
$$f(y) = F'(y)$$

(3)
$$P(a \le Y \le b) = \int_a^b f(y)dy = F(b) - F(a)$$

We can illustrate $P(a \le Y \le b) = F(b) - F(a)$, using the graphs below. You can see that subtracting the first area from the second area yields the third area.



In practice, the cdf is easier to deal with than the pdf, the normal distribution being the prime example (pun intended ☺). Moreover cdf is used to define the **median** and **quartiles** of a probability distribution.