Lecture 10

Continuous Random Variables

In this section you will compute probabilities by doing integrals.

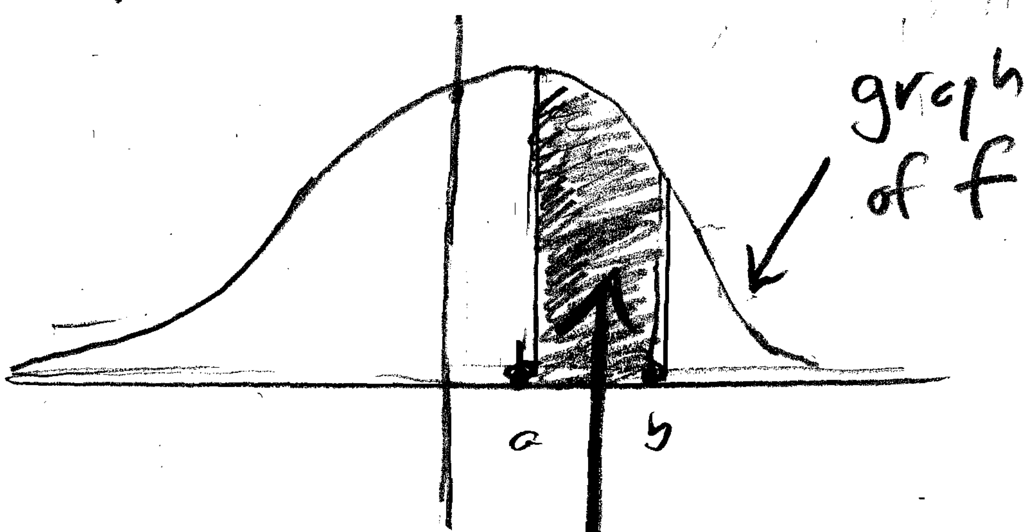
Definition

A random variable X is said to be continous if there exists a nonnegative function $f(x)$ defined on interval $(-\infty, \infty)$ such that for any interval $[a, b]$ we have

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

$f(x)$ is said to be the probability density function of X, abbreviated pdf.

The usual geometric interpretation of the integral $\int_a^b f(x)\,dx$ as the area between a and b under the graph of f will be very important later



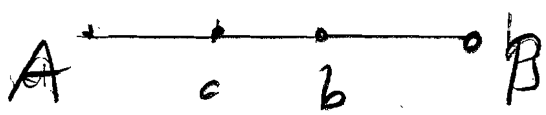
$P(a \leq X \leq b)$ = this area

2 $f(x) \neq P(X=x)$ in fact $f(x)$ is not the probability of anything

f is a <u>density</u> ie something you integrate to get the magnitude of a physical quantity

Think of a wire stretching from a to b with density

$\lambda \ \frac{gm}{cm}$



Then the actually mass of the wire between a and b

is $\int_a^b \lambda(x)\,dx$

So λ is mass per unit length $\lambda(x) = \lim \frac{\Delta m}{\Delta x}$

Similarly $f(x) =$ probability per unit length.

Both λ and f are derivatives of functions whose values have meaning

Properties of $f(x)$

(i) $f(x) \geq 0$ ← no immediate physic interpretation, see later

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$ ← total probability $= 1$

Any function f satisfying (i) and (ii) is a pdf.

Example: The Uniform Distribution on $[0,1]$

Physical Problem

Pick a random number in $[0,1]$

Call the result X.

So X is a random variable.

Questions

What is $P(X = \frac{1}{2})$?

What is $P(0 \le X \le \frac{1}{2})$

What is $P(\frac{1}{4} \le X \le \frac{3}{4})$

So we arrive at for $[a,b]$ inside $[0,1]$

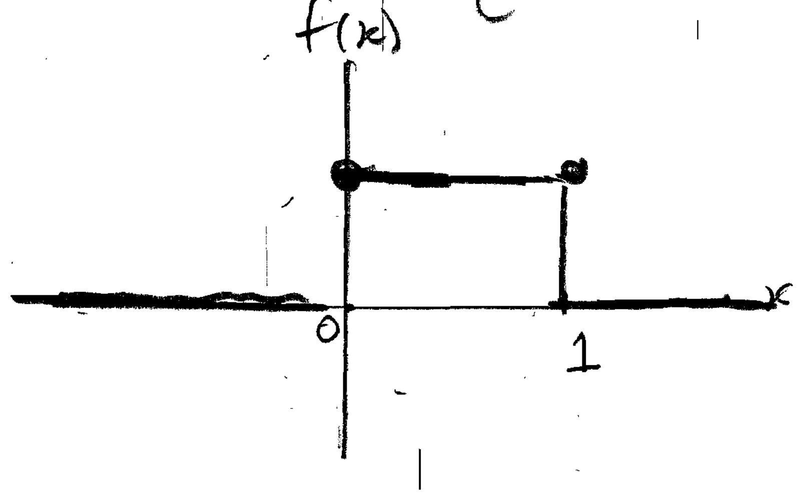
$$P(X \in [a,b]) = P(a \le X \le b)$$

$$= b-a = \text{length}([a,b])$$

This is a continuous random variable. The density function is the "characteristic function of $[0,1]$"

ie.

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$



Definition

A continuous random variable X is said to have uniform distribution on $[0,1]$, abbreviate $X \sim U(0,1)$ if its pdf f is given by

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

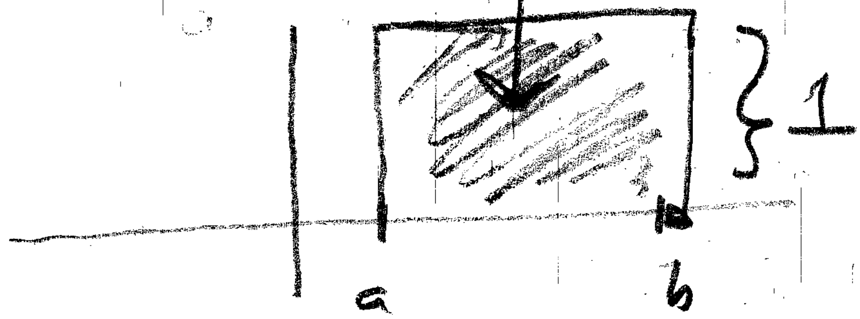
More generally suppose we replace $[0,1]$ by the interval $[a,b]$

We can't have

$$f(x) = \begin{cases} 1, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

Why

$$\int_{-\infty}^{\infty} f(x)\,dx = \text{ this area}$$



$$= b - a \neq 1$$

So we have to define

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

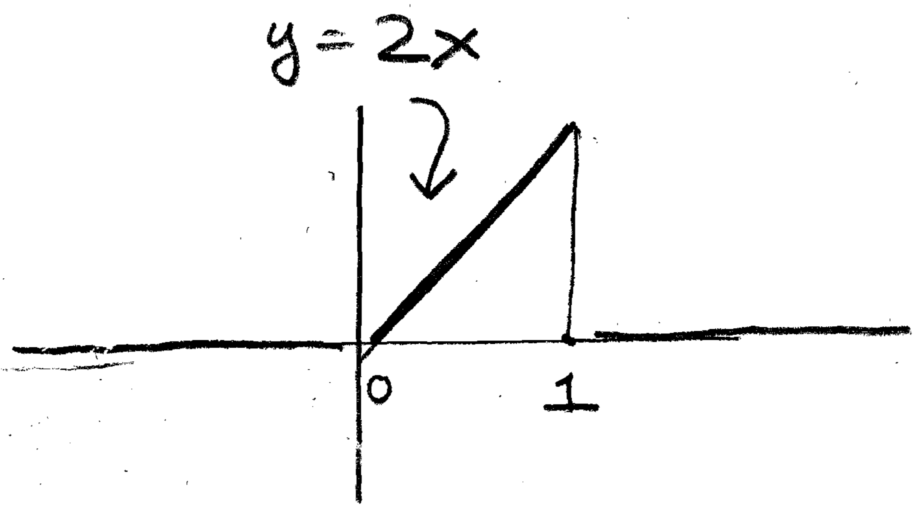
Then $\int_a^b f(x)\,dx = 1$

uniform distribution on $[a,b]$

In this case we write $X \sim U(a,b)$.

Another Example

Linear density



Consider the function

$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

Then the total probability is

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 2x = \left(x^2\right)\Big|_{x=0}^{x=1} = 1$$

Since $f(x) \geq 0$ and

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

indeed a pdf.

Problem

For the linear density compute

$$P\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right)$$

Solution

$$P\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right) = \int_{-\infty}^{\infty} f(x)\,dx = \int_{\frac{1}{4}}^{\frac{3}{4}} 2x\,dx$$

$$= \left(x^2\right)\Big|_{x=\frac{1}{4}}^{x=\frac{3}{4}} = \frac{9}{16} - \frac{1}{16} = \frac{1}{2}$$

No decimals please.

Here are some useful properties of <u>continuous</u> random variables

They are all consequences of the fact that if X is continuous and c is any number then

$$P(X=c)=0$$

<u>Theorem</u>

(i) $P(a \le X \le b) = P(a \le X < b)$
(because $P(X=b)=0$)

(ii) $P(a \le X \le b) = P(a < X \le b)$
(because $P(X=a)=0$)

(iii) $P(a \le X \le b) = P(a < X < b)$

<u>end points don't matter</u>

Good Citizen Computations

The Cumulative Distribution Function

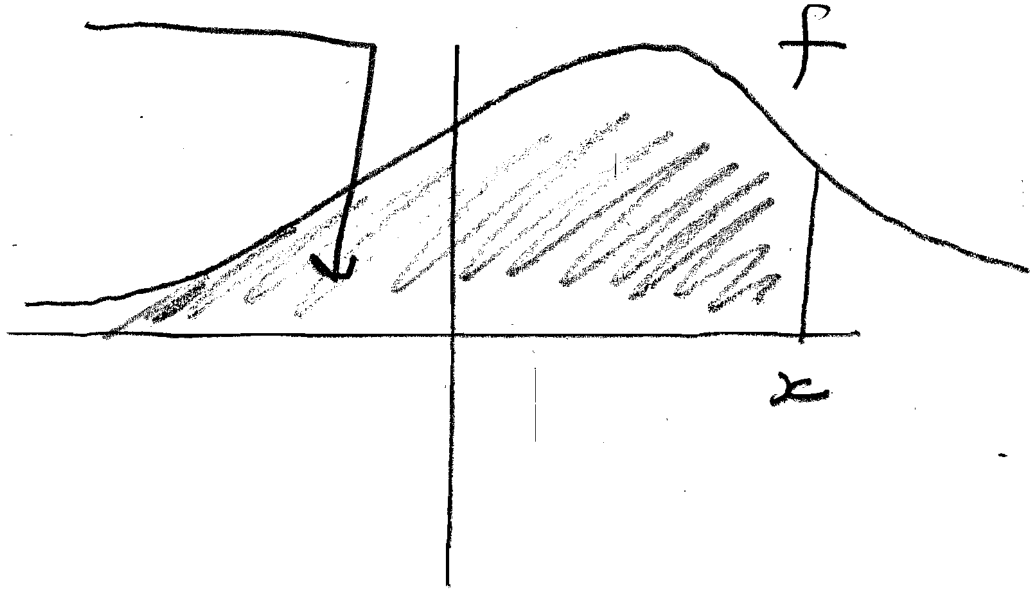
Definition

Let X be a continuous random variable with pdf f. Then the cumulative distribution function F, abbreviate cdf, is defined by

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

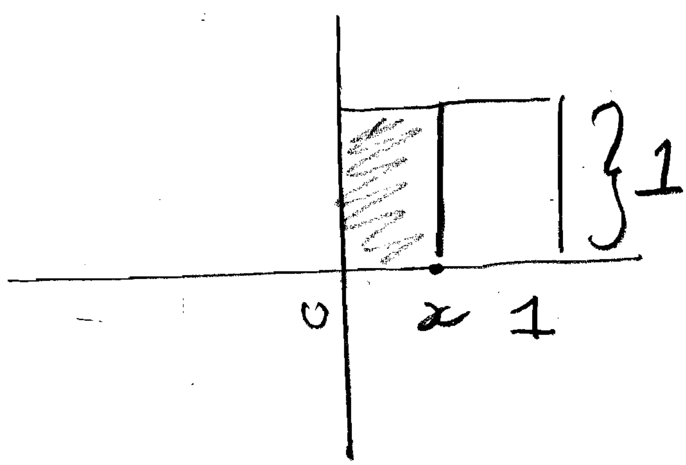
= the area under the graph of f to the left of x.

This area is $F(x)$



We will compute the cdfs for $X \sim U(0,1)$, and $X \sim$ the linear distribution

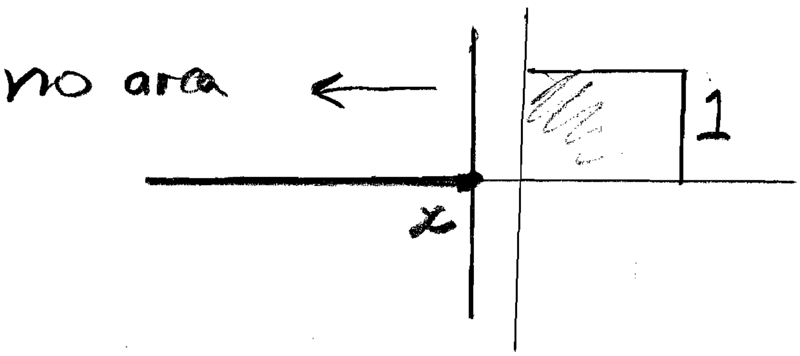
$X \sim U(0,1)$



There will be <u>three</u> formulas corresponding to the <u>two</u> discontinuities in $f(x)$

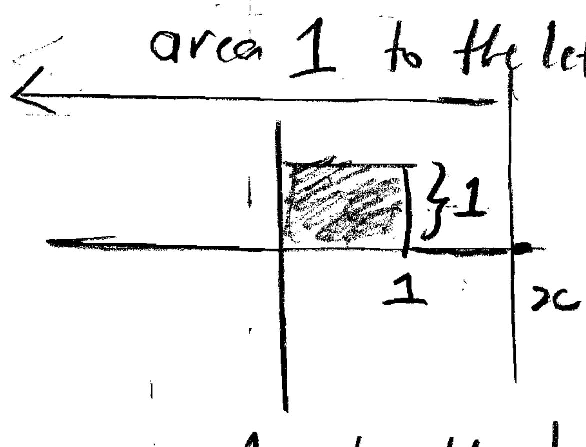
<u>$F(x)=0, \; x<0$</u>

This is clear because we haven't accumulated any probability/area yet



<u>$F(x)=1, \; x>1$</u>

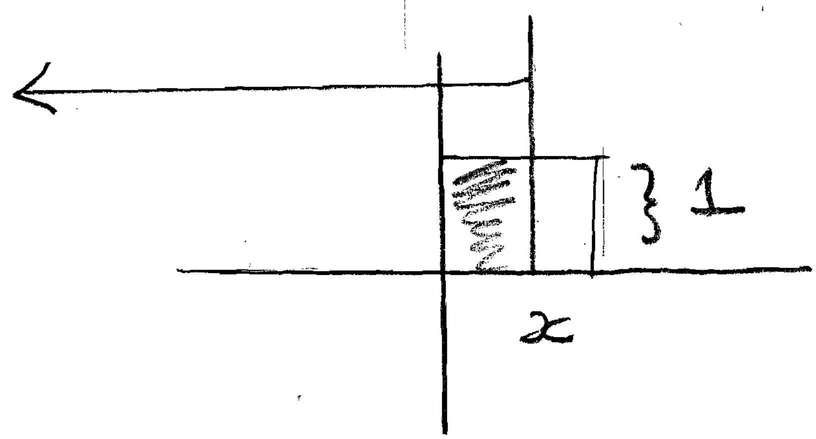
This is not quite so clear



We have area 1 to the left of x and that's all we are going to get

$F(x) = ?\,,\ 0 \le x \le 1$

This is where the action is.

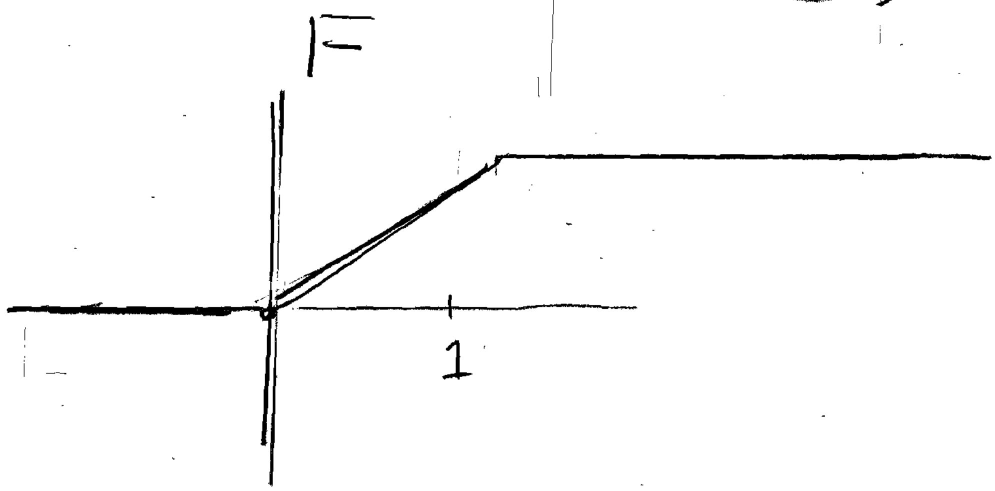


How much area have we accumulated to the left of x. It is the area of a rectangle with base x and height 1 hence area $x \cdot 1 = x$.

Thus $F(x) = x\,,\ 0 \le x \le 1$

We could have done this with integrals instead of pictures but pictures are better.

We have obtained $F(x) = \begin{cases} 0, & x<0 \\ x, & 0 \le x \le 1 \\ 1, & x>1 \end{cases}$



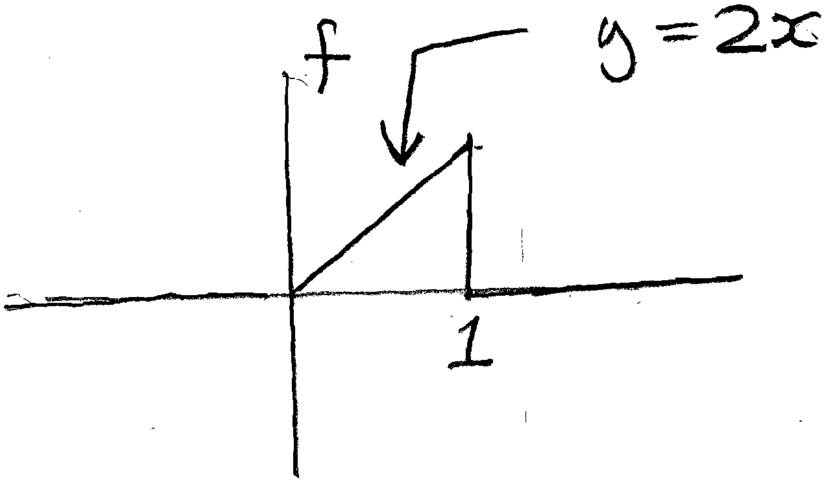
Lesson

cdf's of continuous random variables are continuous and satisfy

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

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

The cdf of the linear distribution



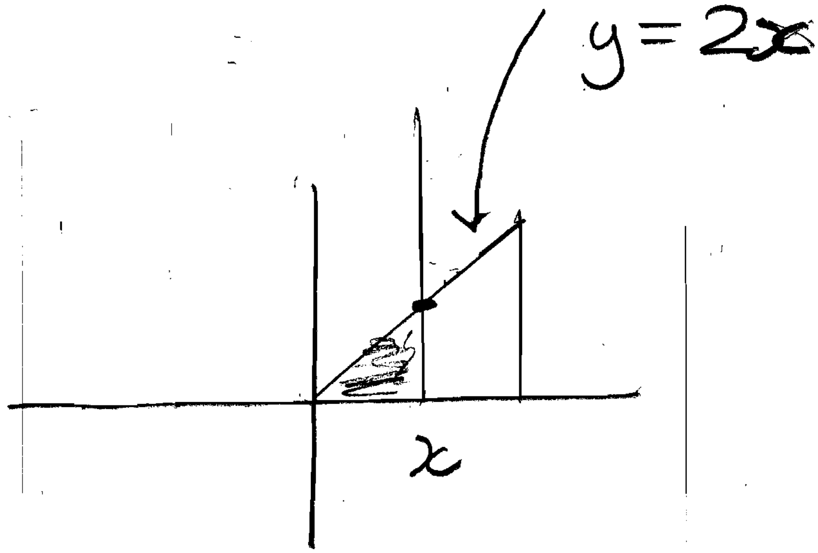
We will go faster. Clearly again

$$F(x)=0, \quad x<0$$

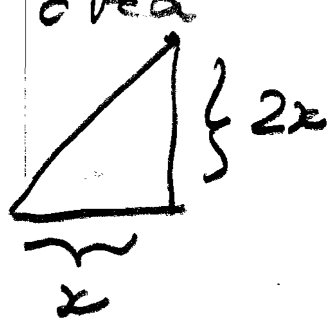
and $$F(x)=1, \quad x>1$$

We have to compute $F(x)$

for $\quad 0 \leq x \leq 1$



So $F(x)$ = shaded area

= Area ()

So we have to compute the area of a triangle with base $b = x$ and height $h = 2x$. But

$$\text{area} = \frac{1}{2} b h = \frac{1}{2} x (2x) = x^2$$

so

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

Do this with integrals.

Importance of the cdf

Coded into the cdf F are all the probabilities $P(a \le X \le b)$

Theorem

$$P(a \le X \le b) = F(b) - F(a)$$

Proof

$$P(a \le X \le b) = P(X \le b) - P(X < a)$$

But because X is continuous

$$P(X < a) = P(X \le a)$$

so

$$P(a \le X \le b) = P(X \le b) - P(X \le a)$$

$$= F(b) - F(a)$$

$\square$

Remark

The previous theorem is critical. It is the basis of using tables (in a book or in a computer) to compute probabilities. A grid of values of F (up to 10 decimal places say) are tabulated

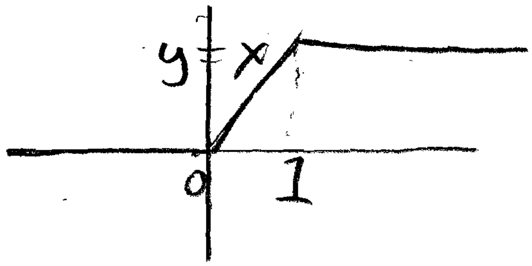
Theorem (How to recover the pdf from the cdf)

$$F'(x) = f(x)$$ at all points where $f(x)$ is continuous

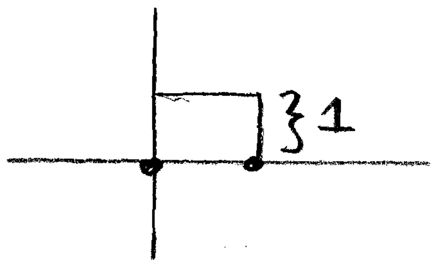
Example

Suppose $X \sim U(0,1)$

$F(x)$ has the graph



So $F(x)$ is differentiable except at 0 and 1 and has derivative



But this is $f(x)$. Note $f(x)$ is discontinuous at 0 and 1.