Lecture 10 : Continuous Random Variables

In this section you will compute probabilities by doing integrals.

Definition

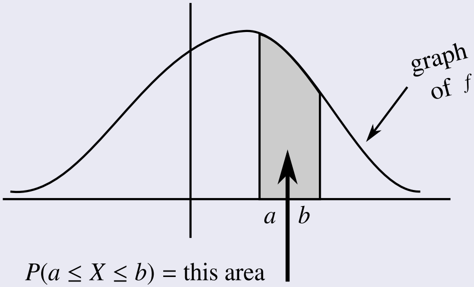
A random variable X is said to be continuous if there exists a nonnegative function $f(x)$ definition interval $(-\infty, \infty)$ such that for any interval $[a, b]$ we have,

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

Definition (Cont.)

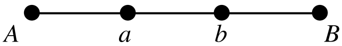
$f(x)$ is said to be the probability density function of X, abbreviated pdf.

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



Z $f(x) \neq P(X = x)$ in fact $f(x)$ is not the probability of anything f is a *density* i.e., something you integrate to get the magnitude of a physical quantity.

Think of a wire stretching from a to b with density $\lambda \dfrac{gm}{cm}$



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

$$\lambda(x) = \lim \frac{\triangle m}{\triangle x}$$

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

So both $\lambda(x)$ resp. $f(x)$ are *densities* which must be integrated to get the actual length resp. probability.

Properties of $f(x)$

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

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

Any function $f(x)$ satisfying (i) and (ii) is a probability density function.

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

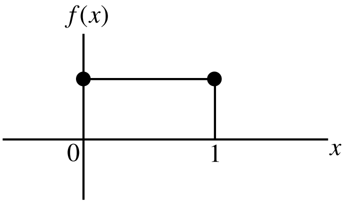
What is $P\left(X = \frac{1}{2}\right)$?
What is $P\left(0 \leq X \leq \frac{1}{2}\right)$
What is $P\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right)$

So for any interval $[a, b]$ which is a subinterval of $[0, 1]$ we have the formula

$$P(X \in [a, b]) = P(a \leq X \leq b) = \int_a^b 1 dx = b - a = \text{ length } ([a, b])$$

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

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



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

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

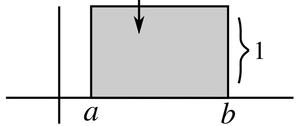
More generally suppose we replace $[0, 1]$ by the interval $[a, b]$

Z *We can't have*

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

Why?

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



$$= b - a \neq 1$$

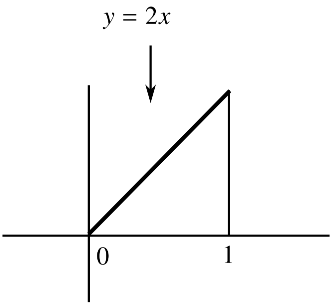
So we have to define

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

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

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

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

Since $f(x) \geq 0$ and

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

$f(x)$ is indeed a pdf.

Problem

For the linear density compute

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

Solution

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

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

No decimals please.

Here are some usual properties of *continuous* random variables.
They are all consequences of the fact that if X is continuous and c is any number then

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

So if X is a continuous random variable, all point probabilities are zero.

Theorem

(i) $P(a \leq X \leq b) = P(a \leq X < b)$ *(because $P(X = b) = 0$)*
(ii) $P(a \leq X \leq b) = P(a < X \leq b)$ *(because $P(X = a) = 0$)*
(iii) $P(a \leq X \leq b) = P(a < X < b)$

end points don't matter.

Good Citizen Computations

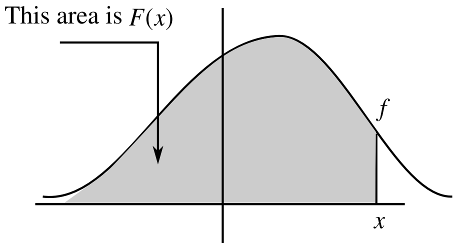
The Cumulative Distribution Function

Definition

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

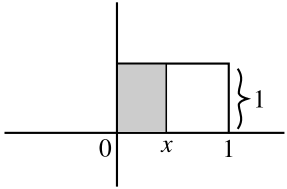
$$F(x) = \int_{-\infty}^{x} f(x)dx$$

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



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

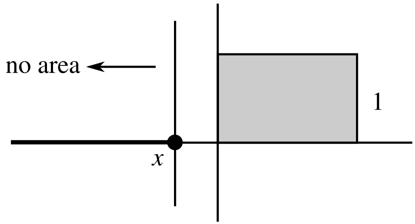
$X \sim U(0,1)$



There will be *three* formulas corresponding to the *two* discontinuities in $f(x)$.

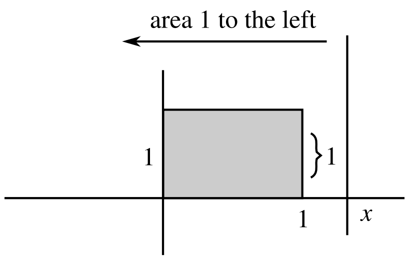
$F(x) = 0,\ x < 0$

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



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

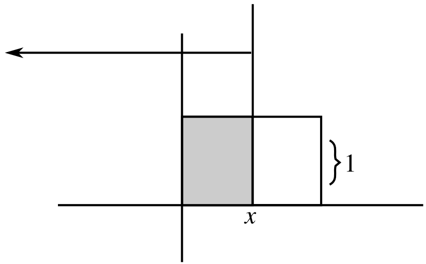
This is not quite so clean



We have area 1 to the left of x and that's all we are going to get no matter how far we push the vertical line to the right.

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

This is where the action is.



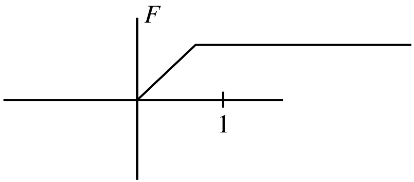
How much area have we accumulated to the left of x. It is the area of a rectangle with base x and height 1 hence area $x \cdot 1 = x$. Thus

$$F(x) = x,\ 0 \leq x \leq 1$$

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

We have obtained

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



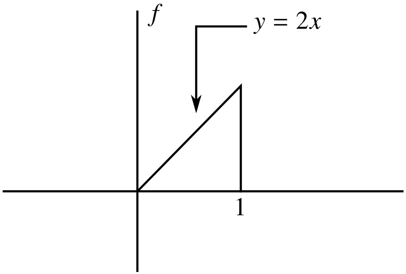
Lesson

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

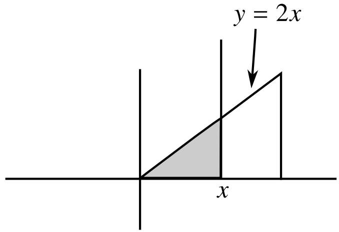


We will go faster. Clearly again

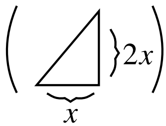
$$F(x) = 0, \quad x < 0$$

$$\text{and} \qquad F(x) = 1, \quad x > 1$$

We have to compute $F(x)$ for $0 \leq x \leq 1$.



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

$$\text{area} \;= \frac{1}{2}bh = \frac{1}{2}x(2x) = x^2$$

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

Theorem

$P(a \leq X \leq b) = F(b) - F(c)$.

Proof.

$P(a \leq X \leq b) = P(X \leq b) - P(X < a)$
But because X is continuous

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

So

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X \leq 0) \\ &= F(b) - F(a) \end{aligned}$$

□

Remark

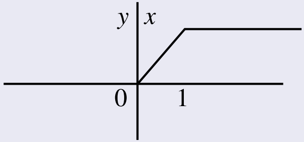
The previous theorem is critical. It is the basis of using tables (in a book or in a computer) to compute probabilities. A grid of values of F (up to 10 decimal places say) are tabulated.

Theorem (How to recover the pdf from the *cdf*)

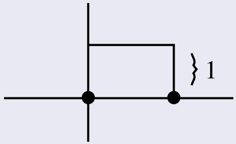
$F'(x) = f(x)$ *at all points where* $f(x)$ *is continuous.*

Example

Suppose $X \sim U(0,1)$ *hence* $F(x)$ *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 of* 0 *and* 1.