CONTINUOUS RANDOM VARIABLES (Chapter 4)

Cumulative distribution functions

df

The *(cumulative) distribution function (cdf)* of a random variable Y is
$$F(y) = P(Y \le y), \quad y \in \mathbb{R} \quad \text{ie } -\infty < y < \infty$$

Example 1 Let Y be the number of heads that come up on 2 tosses of a coin. Find Y's cdf.

$Y \sim Bin(2, \tfrac{1}{2})$ $\qquad f(y) = \binom{2}{y}\left(\tfrac{1}{2}\right)^y\left(1-\tfrac{1}{2}\right)^{2-y}, \quad y = 0, 1, 2$ Ch 3

Y's pdf is $p(y) = \begin{cases} 1/4, & y = 0 \\ 1/2, & y = 1 \\ 1/4, & y = 2 \end{cases}$

Observe that: $F(0) = P(Y \le 0) = p(0) = 1/4$

$F(0.3) = P(Y \le 0.3) = p(0) = 1/4$ (same), etc.

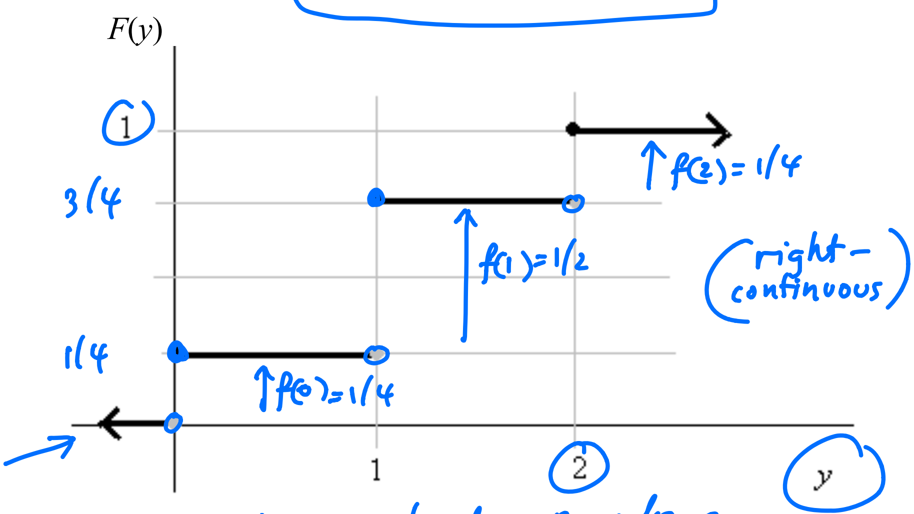
Also: $F(1) = P(Y \le 1) = p(0) + p(1) = 1/4 + 1/2 = 3/4$

$F(1.9) = P(Y \le 1.9) = p(0) + p(1) = 1/4 + 1/2 = 3/4$ (same)

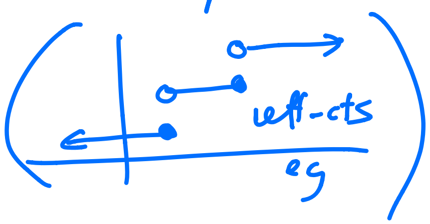
$F(-3) = P(Y \le -3) = 0$

$F(2.1) = P(Y \le 2.1) = 1$ etc.

Therefore Y's cdf is $\boxed{F(y) = \begin{cases} 0, & y < 0 \\ 1/4, & 0 \le y < 1 \\ 3/4, & 1 \le y < 2 \\ 1, & y \ge 2 \end{cases}}$



piecewise constant everywhere



This is a step function, where each 'jump' corresponds to a probability.
Eg, the jump at 2 is 1/4, which is the probability that Y equals 2.
Note that $F(0)$ equals 0.25, not 0.

Three properties of a cumulative distribution function

If $F(y)$ is a cdf then:

1. $F(y) \to 0$ as $y \to -\infty$
2. $F(y) \to 1$ as $y \to +\infty$
3. $F(y)$ is nondecreasing.

Also, 4. $F(y)$ is right continuous, meaning that $\lim_{\delta \downarrow 0} F(y+\delta) = F(y)$.

(In Example 1 this corresponds to the fact that $F(0) = 0.25$, not 0.)

Definition of a continuous random variable

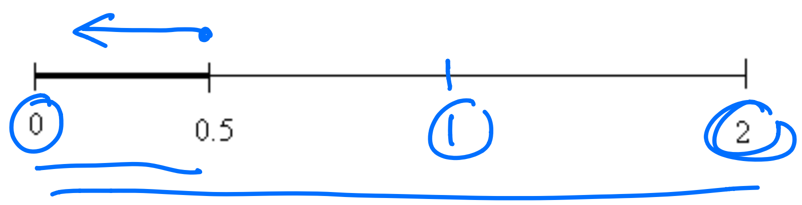
A random variable is said to be *continuous (cts)* if its cdf is continuous (everywhere). (no jumps)

For instance, Y in Example 1 is *not* a continuous rv. ($F(y)$ is discontinuous at 0,1,2.)

Example 2 Let Y be a number chosen randomly between 0 and 2.
Find Y's cdf. Is Y a cts rv?

0.16293..
1.87342..

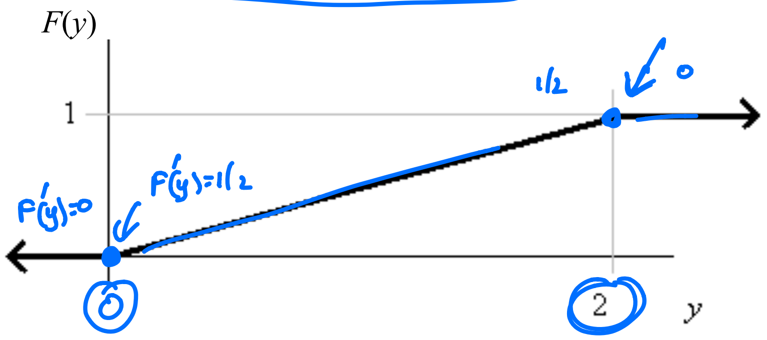
$F(0.5) = P(Y \le 0.5) = 0.5/2 = 0.25$



$F(1) = P(Y \le 1) = 1/2$
$F(1.5) = P(Y \le 1.5) = 1.5/2 = 0.75$, etc.

Also: $F(-1) = P(Y \le -1) = 0$
$F(4) = P(Y \le 4) = 1$, etc.

We see that $F(y) = \begin{cases} 0, & y \le 0 \\ y/2, & 0 < y < 2 \\ 1, & y \ge 2 \end{cases}$



Observe that $F(y)$ is continuous everywhere (ie for all y between $-\infty$ and ∞).
Hence Y is a continuous random variable.

The probability density function of a continuous random variable

$f_Y(1) = P(Y=1)$ in Ch 3

What is the probability that Y equals 1? Answer: $P(Y = 1) = 0$.

(There is an uncountably infinite number of possible values of Y, and they are all equally likely; so each one occurs with probability 0. Also, this follows from there being no 'jump' at $y = 1$ in Y's cdf, $F(y)$.)

In fact, $\boxed{P(Y = y) = 0}$ for all y. $\in \mathbb{R}$

It follows that the earlier definition of a pdf (ie, $p(y) = P(Y = y)$) is now *useless*.

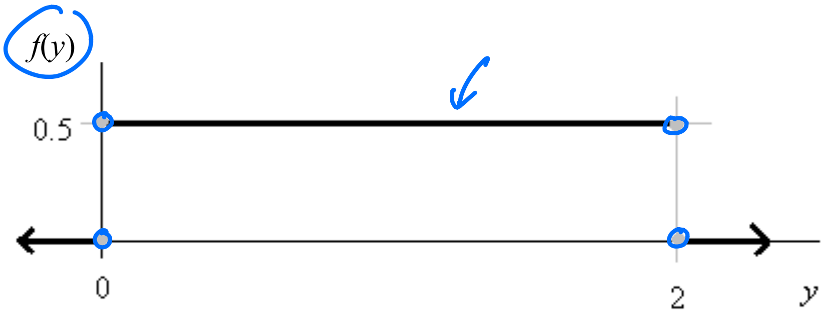
New definition: Suppose that Y is a continuous random variable with cdf $F(y)$. Then Y's *probability density function (pdf)* is $\boxed{f(y) = F'(y)}$ $(= dF(y)/dy =$ the derivative of $F(y))$.

Example 3 Find Y's pdf in Example 2.

$$f(y) = \frac{dF(y)}{dy} = \begin{cases} \dfrac{d0}{dy} = 0, & y < 0 \\ \dfrac{d(y/2)}{dy} = \dfrac{1}{2}, & 0 < y < 2 \\ \dfrac{d1}{dy} = 0, & y > 2 \end{cases}$$

Note that $f(y)$ is undefined at $y = 0, 2$.
(There are two different derivatives at each of these values.)

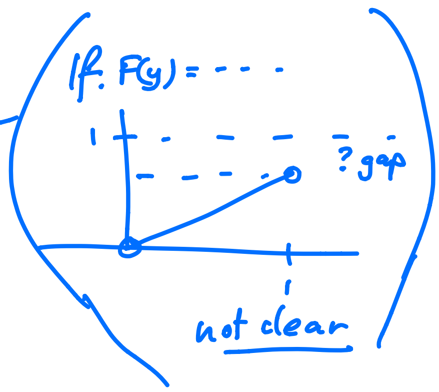
Eg, at $y = 0$, the left derivative is 0 and the right derivative is 1/2.



Observe that $f(y)$ is the slope function of $F(y)$.
Eg, slope of $F(y)$ at 0.1 is $f(0.1) = 1/2$.

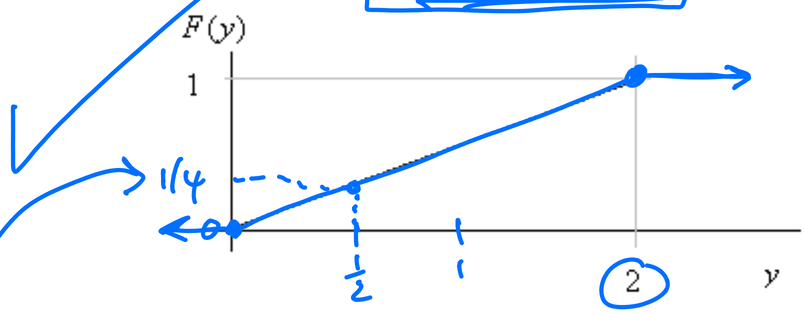
Some simplifications

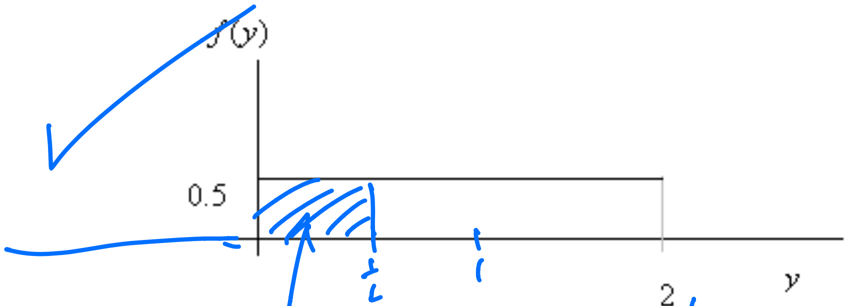
It is conventional to take undefined values as 0.
Also, we will not bother to always indicate exactly where a pdf is zero, nor where a cdf is 0 or 1. (These details have no effect on calculations when considering a purely *continuous* distribution. However, when dealing with a *mixed* distribution, as discussed in the optional Section 4.11, these details are important.)



Graphs will also be simplified.

Thus we may write: $\boxed{F(y) = y/2,\ \ 0 < y < 2}$
$\boxed{f(y) = 1/2,\ \ 0 < y < 2.}$

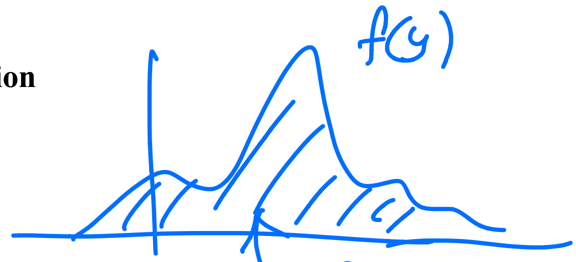




area = 1/4 $\left(=\frac{1/2}{2}\right) = \int_{-\infty}^{1/2} f(y)\,dy = \int_{0}^{1/2} f(y)\,dy + \int_{-\infty}^{0} f(y)\,dy$

0

Two properties of a continuous probability density function



Observe that no value of $f(y)$ above is negative.
Also, the area under $f(y)$ equals 1 (area = 2(1/2) = 1).

area $= \int f(y)dy = 1$

These properties hold for the pdf of every continuous random variable.

If $f(y)$ is the pdf of a cts rv then:

1. $f(y) \geq 0$ for all y (NB: $f(y)$ can be greater than 1.)
2. $\int f(y)dy = 1$. (NB: By default, the integral is over the whole real line; thus it could also be written $\int_{\Re} f(y)dy$ or $\int_{-\infty}^{\infty} f(y)dy$.)

Observe in the last two figures that $F(0.5) = 1/4$ is the same as the area under $f(y)$ to the left of 0.5.

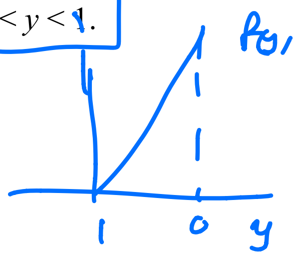
This fact may be written $F(0.5) = \int_{-\infty}^{0.5} f(y)dy$, or equivalently, $F(0.5) = \int_{-\infty}^{0.5} f(t)dt$.

In general, the cdf $F(y)$ of a continuous random variable Y can be obtained from its pdf $f(y)$ via the equation

$$\boxed{F(y) = \int_{-\infty}^{y} f(t)dt\,.}$$

$$\boxed{f(y) = \frac{d}{dy}F(y)}$$

Example 3 Suppose that Y has pdf $\boxed{f(y) = 2y,\ 0 < y < 1}$.
Find Y's cdf.

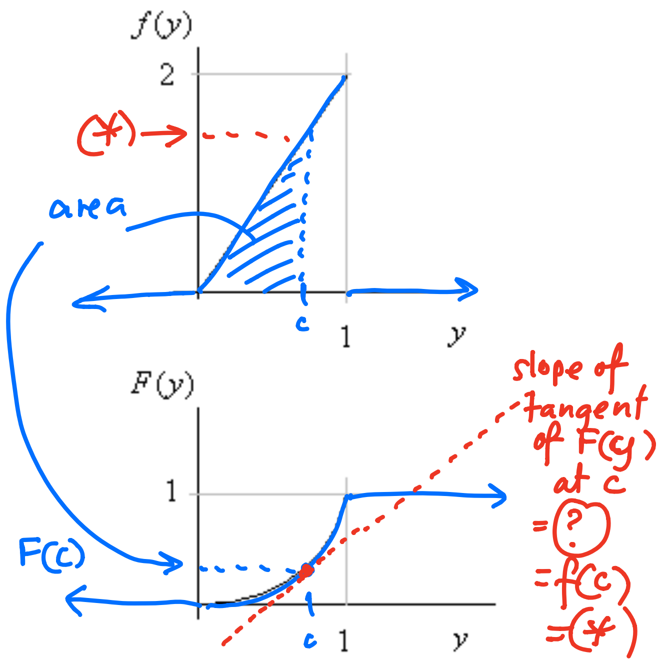


$$F(y) = \int_0^y 2t\,dt = \left[t^2\Big|_{t=0}^{y}\right] = y^2 - 0^2\,.$$

So Y's cdf is $\boxed{F(y) = y^2,\ 0 < y < 1}$.

Note that we could now also 'switch back' to the pdf via differentiation:

$$f(y) = F'(y) = 2y,\ 0 < y < 1\,.$$



Consider any value c of Y ($c \in [0,1]$)

The area under $f(y)$ to the left of c is
$F(c) = c^2$

The slope of $F(y)$ at $y = c$ is
$F'(c) = f(c)$.

Eg: Area under $f(y)$ to left of $y = 1/2$ is
$F(1/2) = (1/2)^2 = 1/4$

Slope of $F(y)$ at $y = 1/2$ is
$f(1/2) = 2 \times 1/2 = 1$.

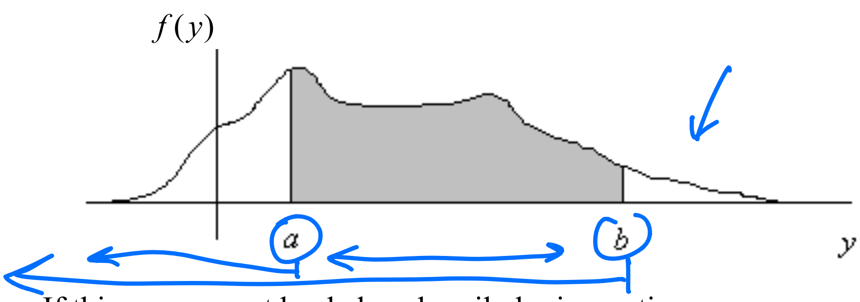
The slopes of $F(y)$ at 0 and 1 are 0 and 2.

Computing probabilities involving cts rv's

($<\ \equiv\ \leq$)

Recall that $\boxed{P(Y = y) = 0}$ for all $y \in \mathbb{R}$.
It follows that $P(Y \leq y) = P(Y < y)$, $P(Y \geq y) = P(Y > y)$, etc.
+P(Y=y)

The probability $P(a < Y < b)$ is the area under Y's pdf between a and b.



If this area cannot be deduced easily by inspection, we may need to do an integral:

$$P(a < Y < b) = \int_a^b f(y)dy. \quad (1)$$

However if we know Y's cdf, then we can instead use the formula:
$$P(a < Y < b) = F(b) - F(a).$$

(This is because $P(a < Y < b) = P(Y < b) - P(Y < a)$,
assuming that Y is a continuous random variable.)

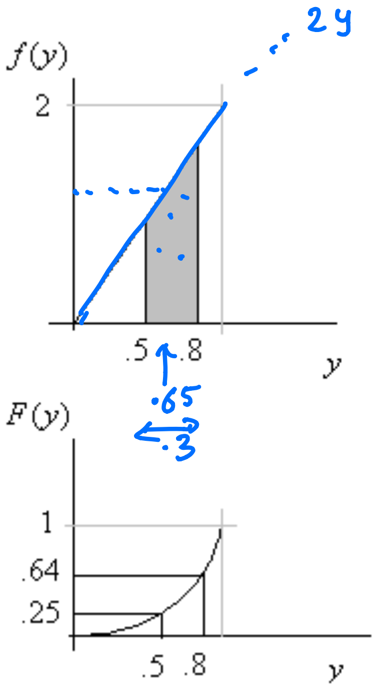
For example, in Eg 3, where $f(y) = 2y$, $0 < y < 1$, what is $P(0.5 < Y < 0.8)$?

Solution 1:

This probability is the area of the shaded region below:

$0.3 \times 2 \times 0.65 = 0.39$.

(0.65 is midway between 0.5 and 0.8, and 2×0.65 is the value of $f(y)$ at that point).



Solution 2:

$$P(0.5 < Y < 0.8) = \int_{0.5}^{0.8} f(y)dy = \int_{0.5}^{0.8} 2y\,dy = \left[y^2 \Big|_{y=0.5}^{0.8} \right]$$

$$= 0.8^2 - 0.5^2 = 0.64 - 0.25 = 0.39$$

Solution 3:

$$P(0.5 < Y < 0.8) = F(0.8) - F(0.5) = 0.8^2 - 0.5^2 = 0.64 - 0.25 = 0.39.$$