Cumulative Distribution Functions (cdf)

We saw two different ways to describe random variables:

1) Probability density functions for continuous rvs

2) Probability mass functions for discrete rvs.

Let $X_1 = \text{Uniform}[0,1]$ $\quad X_2 = \text{Uniform}\{1,2,\dots,6\}$

Let $Z = \text{Bernoulli}(\frac{1}{2})$. Let $\{X_1, X_2, Z\}$ be independent.

$$Y := Z X_1 + (1-Z) X_2$$

Y ________ have a pdf or a pmf

3.2

Cumulative distribution function:

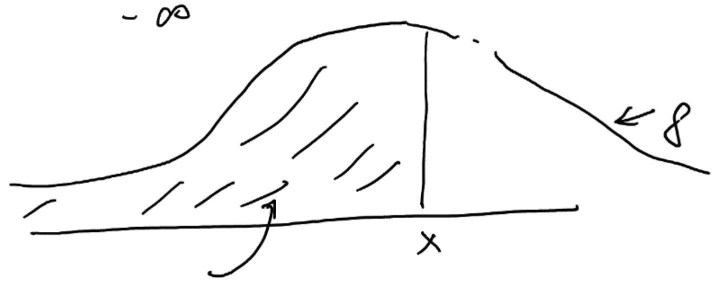
Definition: Given a random variable X, discrete, continuous, whatever, define

$$\boxed{F(t) = P(X \leq t)} \quad \forall t \in \mathbb{R}$$

if X has a pdf f (it is continuous)

$$F(x) = \int_{-\infty}^{x} f(u)\,du = \mathbb{P}(X \leq x)$$

$=$



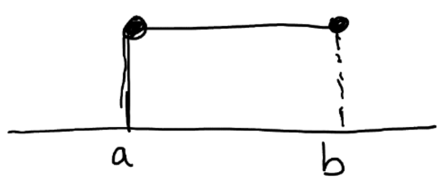
Area under curve.

OTHER NAMES FOR THE CDF:

Cumulative distribution function / distribution function / cdf

Example:

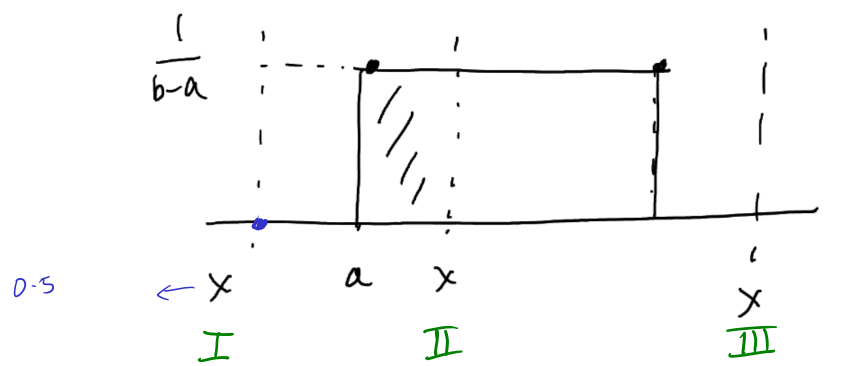
$X \sim \text{Unif}\,[a,b]$. It has the following density / pdf:



$$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \le x \le b \\ 0 & x > b \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(u)\,du .$$

Let us evaluate $F(x)$ at various values of x as seen in the figure below.



$\mathrm{I}:\quad x < a \qquad F(X \le x) = \int_{-\infty}^{x} f(t)\,dt \qquad =$

$\mathrm{II}:\quad a \le x \le b \qquad F(X \le x) = \int_{-\infty}^{a} f(t)\,dt + \int_{a}^{x} f(t)\,dt$

$=$

$\mathrm{III}:\quad x > b \qquad F(X \le x) = \int_{-\infty}^{b} f(t)\,dt \quad + \quad \int_{b}^{x} f(t)\,dt$

Picture of $F(x)$ the cdf:

Example: CDF of a discrete rv

$X \sim \text{Bin}(2, \frac{1}{2})$

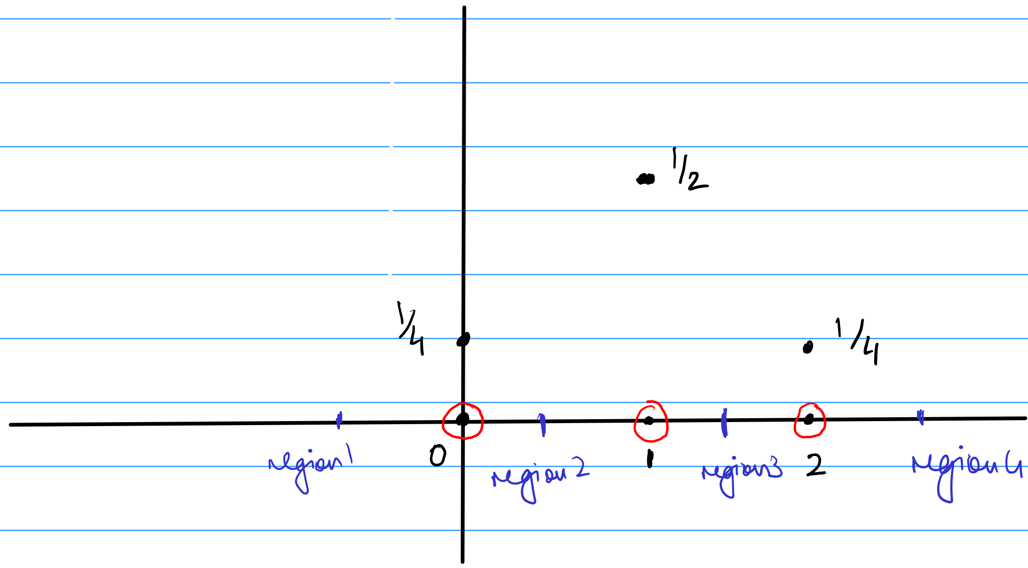
Recall that if $X \sim \text{Bin}(n,p)$ then

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0,1,2,\dots,n$$

$$p_X(0) =$$

$$p_X(1) =$$

$$p_X(2) =$$



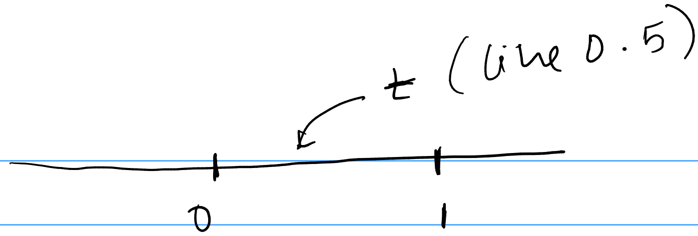
POLL

Evaluate the cdf at $F(0.5)$, $F(1)$

Your answer should be a pair of numbers (a, b).

$F(0.5) = P(X \leq 0.5)$ $\quad$ $F(1) = P(X \leq 1)$

Q & A style poll.



$$F(t) = F(0.5) = P(X \leq 0.5)$$

$$= P(X = 0) = \frac{1}{4}$$

for other values of $t \in (0,1)$

The same reasoning applies $\Rightarrow$ $F(t) = \frac{1}{4}$

$F(1) = 0.75$

$\hookrightarrow P(X \leq 1) = P(X=0) + P(X=1)$

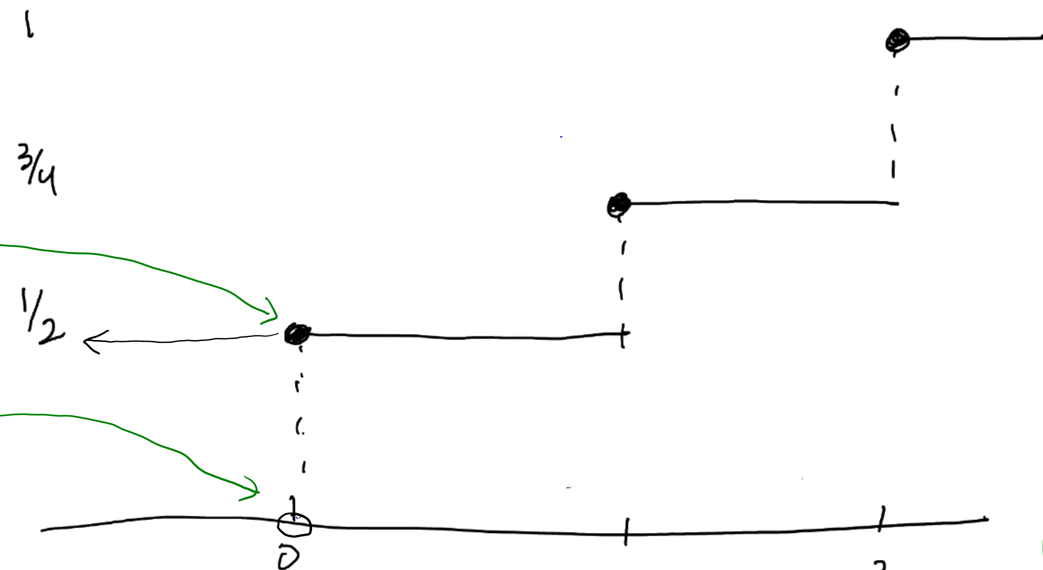
$$= 0.25 + 0.5$$

In general

$$F(t) = \begin{cases} & t < 0 \\ & 0 \leq t < 1 \\ & 1 \leq t < 2 \\ & 2 < t \end{cases}$$

Put this all together

filled-in circle
$f(0) = 1/2$



open circles mean
$F(0) \neq 0$

- <u>Endpoints matter for the CDF!</u>

Contrast with pdf.

Note 3 properties:

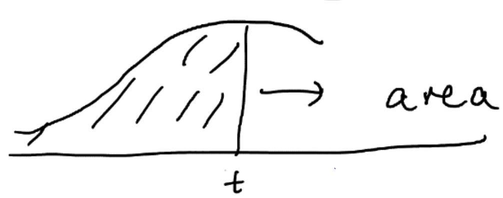
$\lim_{u \to -\infty} F(u)$

1) $F(-\infty) = 0 = \mathbb{P}(X \leq -\infty)$

$\lim_{u \to \infty} F(u)$

$F(+\infty) = 1 = \mathbb{P}(X \leq +\infty)$

2) $F(x)$ is <u>increasing</u>



$\rightarrow$ area under curve is increasing

$F(t) = \mathbb{P}(X \leq t)$

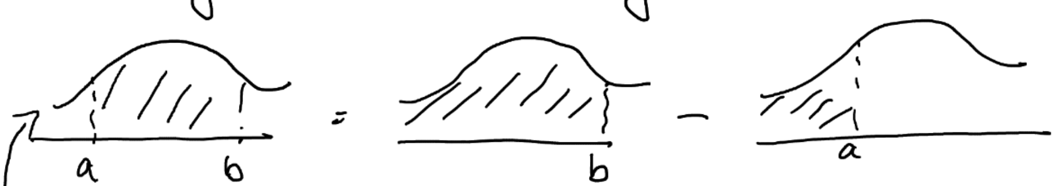
(cadlag)

3) Right continuity (with left limits)

Any function with these 3 properties is a cdf for some rv.

Miscellaneous ideas related to cdfs

Adding and subtracting areas:



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

pdf

BUT, this is Generally true, for both continuous and discrete rvs.

3.13 1) Roll a die first. ($Y \in \{1,2,3,\dots 6\}$

2) Then pick a uniform random variable in $[0,Y]$. $k = 1,2,3,\dots,6$

Find $F(s)$ the cdf of this uniform random variable. In other words, find the cdf of

random parameter

$$X \sim \text{Uniform}([0,Y]) \quad \text{where } Y$$

represents the roll of an independent die.

cdf of X:

$$F(s) = P(X \leq s)$$

Use strategy 1: Break it up into easier-to-compute probabilities.

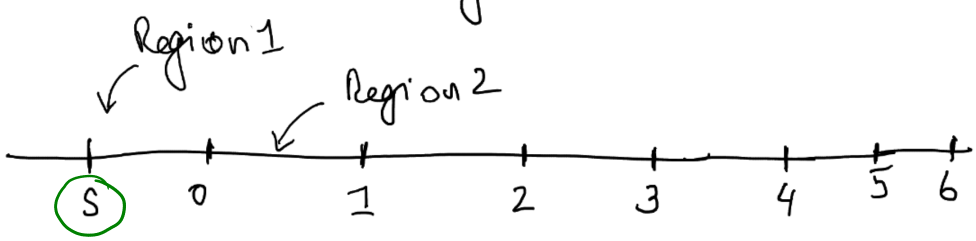
$$P(X \leq s) = \sum_{i=1}^{6} P(X \leq s, Y = i)$$

?

If roll 1, then Uniform $[0,1)$
roll 6 then Uniform $[0,6)$

Range of X : $[0,6)$.

To find $F(s)$: we will determine its value in various regions.



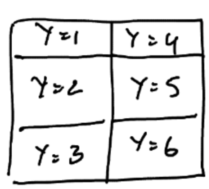
If $s \leq 0$ Then $F(s) = \mathbb{P}(X \leq s) = 0$

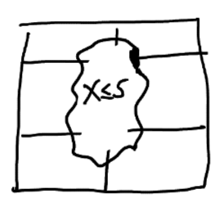
If $0 < s < 1$ Then what? Depends on die roll.

$$\mathbb{P}(X \leq s)$$
$$= \sum_{k=1}^{6} \mathbb{P}(X \leq s, Y = k) \quad (\star)$$

where Y = value of die roll.

Recall





$$\mathbb{P}(X \leq s \mid Y=1) = \mathbb{P}(\text{Uniform}[0,1] \leq s)$$

$$= \frac{s}{1}$$

$$\mathbb{P}(X \leq s \mid Y=2) = \frac{s}{2}$$

$\vdots$

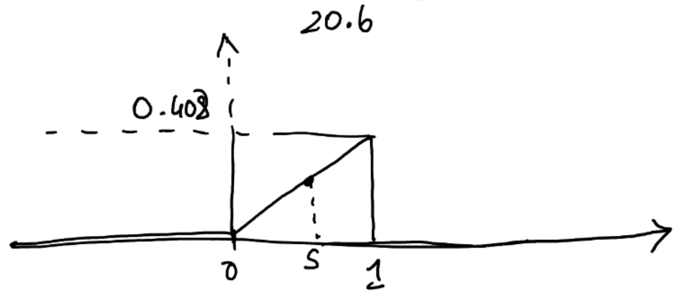
And so on. Therefore

$$\mathbb{P}(X \leq s, Y=i) = \mathbb{P}(X \leq s \mid Y=i)\mathbb{P}(Y=i)$$

$$= \frac{s}{i} \cdot \frac{1}{6}$$

$$\sum_{i=1}^{6} \mathbb{P}(X \leq s, Y=i) = \sum_{i=1}^{6} \frac{s}{i} \cdot \frac{1}{6} \qquad 0 < s < 1.$$

$$= \frac{49}{20.6} s = 0.408 s.$$



If $1 < s < 2$,

lets find

$$F(s) = \mathbb{P}(X \leq s)$$

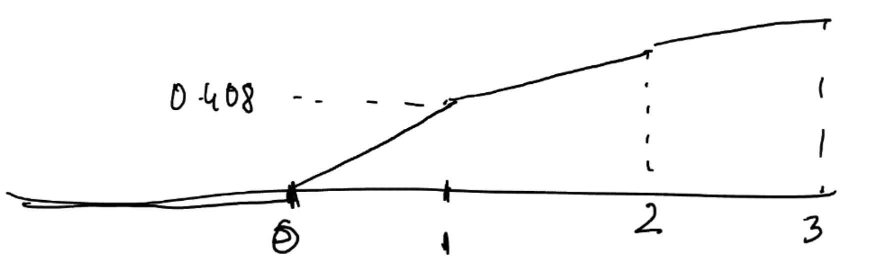
$$= \sum_{k=1}^{6} \mathbb{P}(X \leq s, Y = k)$$

$\mathbb{P}(X \leq s \mid Y = 1) = 1$ (since $1 < s < 2$)

$\mathbb{P}(X \leq s \mid Y = 2) = \frac{s}{2}$ and so on.

$$\begin{array}{r} 0.41 \\ -0.16 \\ \hline 0.25 \end{array}$$

$$\therefore \quad F(s) = \frac{1}{6} + \sum_{k=2}^{6} \frac{s}{k} \cdot \frac{1}{6} = \frac{1}{6} + 0.255s$$



and so on.

What is $F(6)$?

POLL

Can you find $F(s)$ for $2 < s < 3$?

Recall that $Y \in \{1, 2, \ldots, 6\}$. $X \sim \text{Uniform}[0, Y]$

$$P(X \leq s) = \sum_{i=1}^{6} P(X \leq s \mid Y = i)\, P(Y = i)$$

Relationship between pdf and cdf

FUNDAMENTAL THM OF CALC. RELATES PDF and CDF if X continuous

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

$$\frac{d}{dx} F(x) = f(x)$$ if X has a density.

So lets find the density of above.

$$f(s) = \begin{cases} 0.408 & 0 < s < 1 \\ \frac{d}{ds}\left(\frac{1}{6} + 0.25s\right) & 1 < s < 2 \\ & 2 < s < 3 \end{cases}$$

WHAT HAPPENS AT THE POINT $S=1$?

$F(s)$ is not differentiable, and so $f(s)$ ______

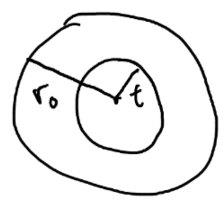
It doesn't matter too much if $f(s)$ ______

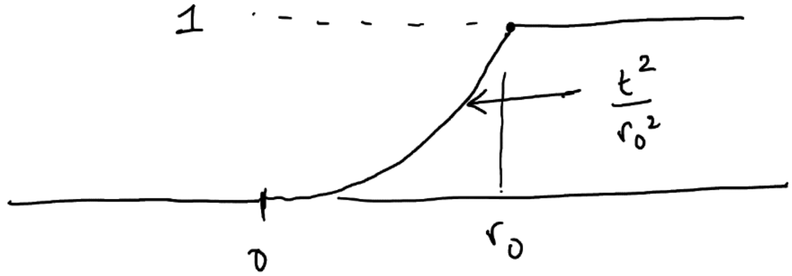
does not exist at a single point.

<u>3.14</u> Dart board radius r_0. Find cdf of

$R = \sqrt{x^2+y^2}$ (distance from origin).

$$\mathbb{P}(R \leq t) = \mathbb{P}\left(\sqrt{x^2+y^2} \leq t\right) = \frac{\pi t^2}{\pi r_0^2} \quad 0<t<r_0$$





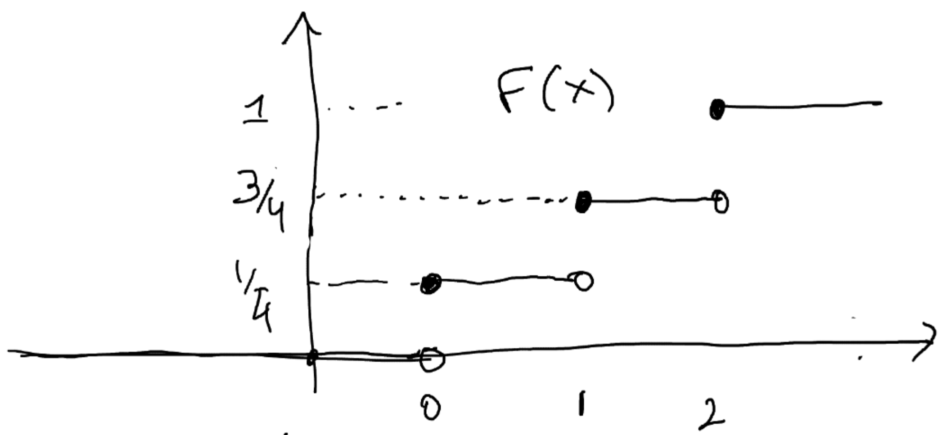
What kind of curve is this for $0<t<r_0$

Density

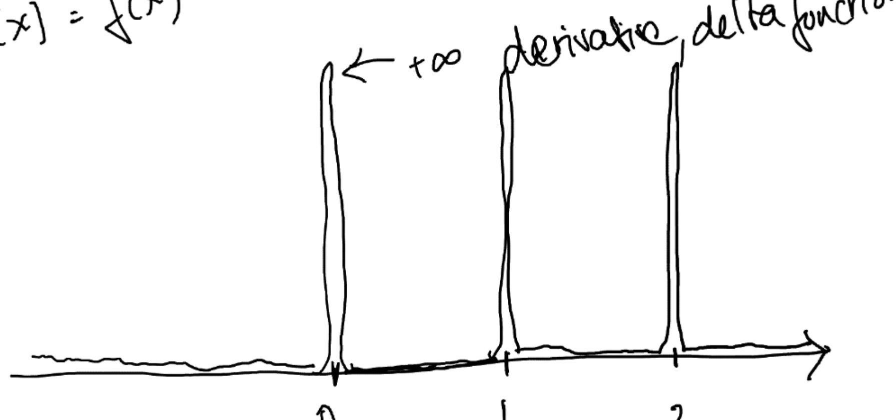
$$f(s) = \begin{cases} 0 & t<0 \\ 2t/r_0 & 0<t<r_0 \\ 0 & t>r_0 \end{cases}$$

CDF and PDF for discrete rvs

Recall, for $X \sim \text{Bin}(2, 1/2)$



"$\frac{d}{dx} F(x) = f(x)$"



Not a (normal) function. So, we say, the density does not exist.

Notice that you can FIND the pmf by looking at JUMPS:

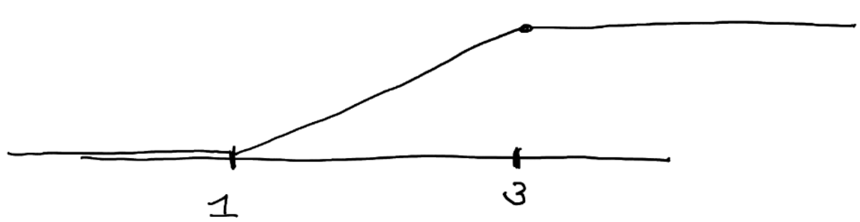
$$F(1) - F(1-\epsilon) = P(X = 1)$$

$$= \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$$

"Difference of cdf" gives pmf of discrete rv.

Derivative of pdf gives pdf of continuous rv.

Ex 3.15 Let $F(s) = \begin{cases} 0 & s < 1 \\ \frac{s-1}{2} & 1 \leq s \leq 3 \\ 1 & s > 3 \end{cases}$

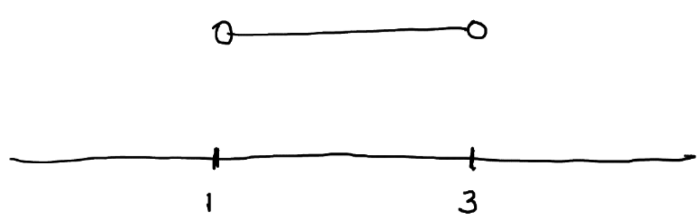


What kind of rv has this cdf? What is its pmf or pdf?

$$F'(s) = \begin{cases} 0 & s < 1 \\ \frac{1}{2} & 0 < s < 3 \\ 0 & s > 3 \end{cases}$$

We can then simply let $f(s) = F'(s)$ be the pdf:

$f(s)$



Notice that $F'(s)$ DID NOT EXIST at $s = 1$ and $s = 3$. But this did not matter for f. You can assign ANY value to f there and it will determine the same probabilities for X.

Here $X \sim$ Uniform $[1,3]$

3.20 You can have cdfs that have a "continuous part" and a discrete part, and so represent a random variable that is neither continuous nor discrete.

An insurance policy has a 500 deductible. An accident costs Y a Uniform $[100, 1500]$.

Let X be the amount that you have to pay. Find the cdf of X:

If $Y \leq 500$, then $X = Y$. Otherwise if $Y > 500$, then you would pay your deductible and the insurance pays the rest.

$$P(X \leq t) = 0 \qquad 0 \leq t \leq 100.$$
$$P(X \leq t) = P(100 < Y \leq t) \qquad 100 \leq t < 500$$

$$P(X \leq 500) = 1$$
$$P(X \leq t) = 1 \qquad 500 \leq t < \infty$$

So all we have to do is to find

$$P(100 < Y \leq t) = \frac{t-100}{1400}$$