Math 3501 - Probability and Statistics I

3.1 - Random variables of the continuous type

Random variable of continuous type

set of values
that X may take

Definition

A random variable X is said to be of *continuous type* if its space S is either an interval or an union of intervals and, moreover, it holds that $P(X = x) = 0$ for each $x \in S$.

The distribution of probability of X is also said to be of the continuous type.

unlike the case of discrete r.v.s for which we would have $P(X = x_i) > 0$ for $x_i \in S$

Examples:

– time between two consecutive airplane arrivals at JFK

– Height of a BC student selected at random

– lifetime of a light bulb ← quality control

– time until death of an insured life ← actuarial math

– time before a stock price exceeds a given threshold ← financial math

– weight of contents of a box of cereal ← quality control

Definition (Probability density function)

S interval or union of intervals

Let X be a random variable of the *continuous type* with space S.

The *probability density function* (abbreviated pdf) of X is a real-valued integrable function f with support S for which the following properties hold:

$\left\{\begin{array}{l} f \text{ has support } S \text{ if } f(x) = 0 \text{ for all } x \notin S \\ \text{and } f(x) \neq 0 \text{ for all } x \in S \end{array}\right.$

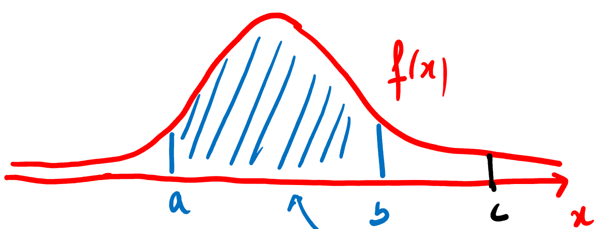
$f(x) > 0$ for all $x \in S$ ⇝ (a) $f(x) \geq 0$ for all $x \in S$

(b) $\displaystyle\int_S f(x)\,\mathrm{d}x = 1$

(c) If $(a, b) \subseteq S$, then the probability of the event $\{a < X < b\}$ is given by

$$\boxed{P(a < X < b) = \int_a^b f(x)\,\mathrm{d}x\ .}$$

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



area shaded in blue $= \int_a^b f(x)\,dx = P(a < X < b)$

Also $P(X = c) = \int_c^c f(x)\,dx = 0$

Cumulative distribution function

The cumulative distribution function (cdf) or distribution function of a random variable X of the continuous type, is related with the pdf of X via

by def of cdf $\longrightarrow$ $$\boxed{F(x) = P(X \le x)} = \int_{-\infty}^{x} \underbrace{f(t)}_{\text{pdf}} \, \mathrm{d}t \,, \quad x \in \mathbb{R}$$

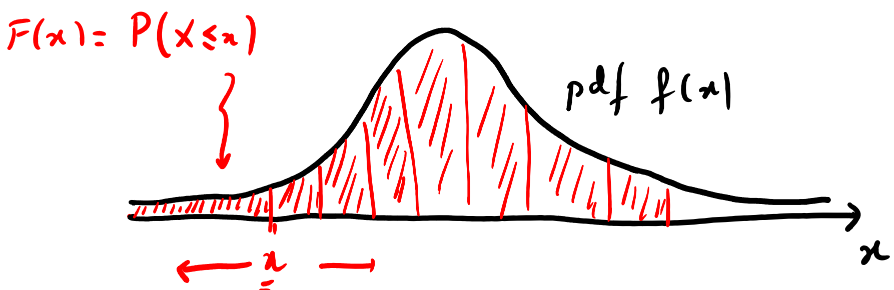
indefinite integral due to the variable x as the upper limit of integration

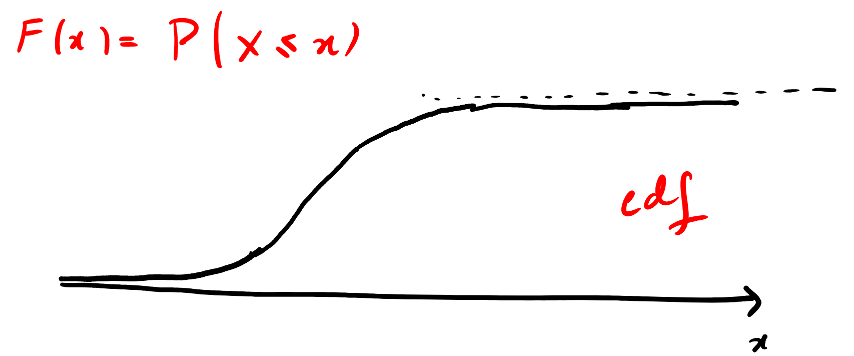
Note: From the fundamental theorem of calculus, we have that

$$\boxed{F'(x) = f(x)}$$

at points of continuity pdf of X is the derivative of the cdf of X

for all values of x at which f is continuous.





Properties: For any random variable X, its cdf satisfies:

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

hold in general!!!

1) $F(x)$ is non-decreasing

2) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$

Moreover, for a continuous random variable X, we have:

$$P(a < X \leq b) = \int_a^b \overbrace{f(x)}^{\text{pdf}}\,\mathrm{d}x = \underbrace{F(b) - F(a)}_{\left[F(x)\right]_{x=a}^{x=b}}$$

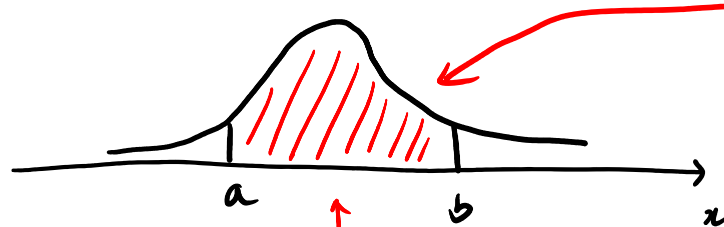
cdf

FTC

$<$ $<$

$\leq$ $\leq$

$\leq$ $<$



$P(a < X \leq b)$

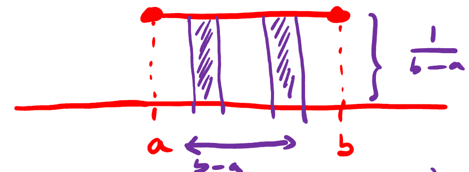
Uniform distribution

(a, b)

The random variable X is said to have a *uniform distribution* on the interval $[a, b]$ if its pdf is constant on $[a, b]$, that is

$f(x) = 0$ for all values of x not in $[a, b]$

$$f(x) = \frac{1}{b-a}\,, \quad a \le x \le b\,.$$



We call this distribution UNIFORM because f assigns equal probability to subintervals of $[a, b]$ of equal width!

Its cdf

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$$

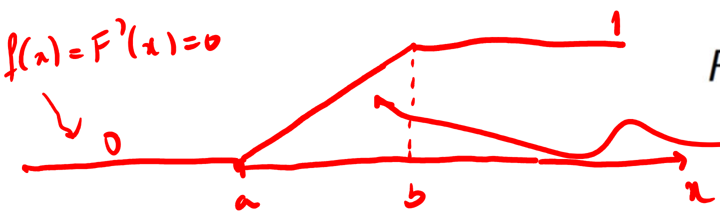
For $x \in [a, b]$

$$F(x) = \int_a^x f(y)\,dy$$

because $f(x) = 0$ for $x < a$

may be written as

$f(x) = F'(x) = 0$



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

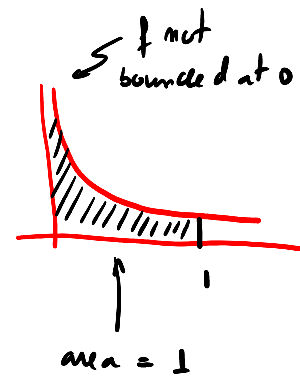
$y = \frac{x-b}{b-a}$ is the equation of the line through $(a,0)$ and $(b,1)$

Notation and terminology:

- we may also say that X is $U(a, b)$ or write $X \sim U(a, b)$.

Remarks:

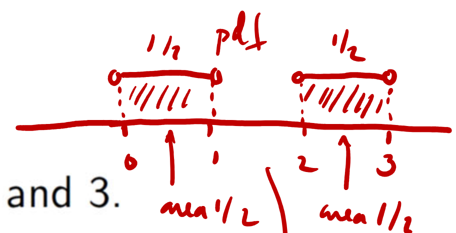
1) Unlike a pmf of a random variable of the discrete type (which is always bounded above by 1), the pdf of a continuous random variable does not have to be bounded.



2) The area between the graph of a pdf and the horizontal axis equals 1.

3) The pdf of a continuous random variable does not need to be a continuous function. For instance, the function

$$f(x) = \begin{cases} 1/2 & \text{if} \quad 0 < x < 1 \quad \text{or} \quad 2 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$



enjoys the properties of a pdf and yet has discontinuities at $x = 0, 1, 2$ and 3.

4) The cdf associated with a distribution of the continuous type is always a continuous function.

not true for discrete r.v.s

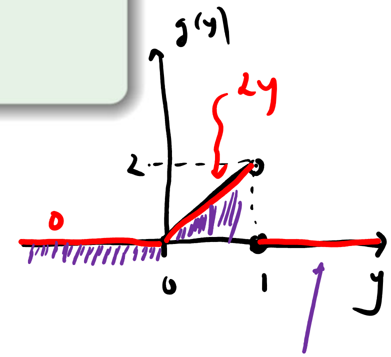
Example

Let Y be a continuous random variable with pdf $g(y) = 2y$ for $0 < y < 1$.

Determine the cdf of Y.

Find $P\left(\frac{1}{2} < Y \leq \frac{3}{4}\right)$. $= ?$

Let Y be the r.v. with pdf $g(y) = \begin{cases} 2y & \text{if } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$



The cdf of Y is

$$G(y) \underset{\text{def}}{=} P(Y \leq y) = \int_{-\infty}^{y} g(t)\,dt = ??$$

Three cases to consider:

i) if $y \leq 0$, then we know that $g(y) = 0$ for all $y < 0$ and so $G(y) = 0$

ii) if $0 < y < 1$, then $G(y) = \int_{-\infty}^{y} g(t)\,dt = \int_{0}^{y} g(t)\,dt = \int_0^y 2t\,dt = y^2$

because $g(t) = 0$ for $t \leq 0$

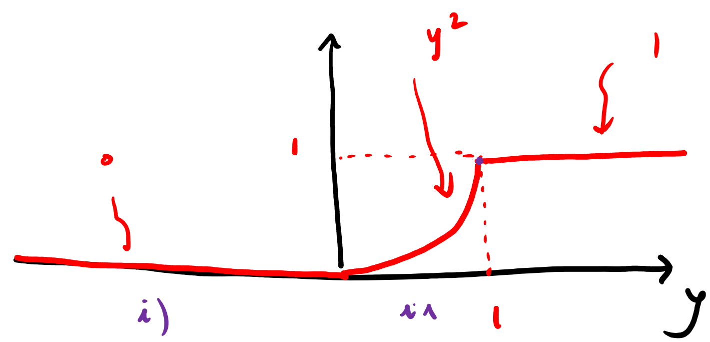
$G'(y) = 0$

iii) if $\overbrace{y \geq 1}$, then $g(y) = 0$ for all $y \geq 1$ and so

$$G(y) = \int_{-\infty}^{y} g(t)\,dt = \int_{0}^{1} g(t)\,dt = 1$$

CONCLUSION: the cdf is

$$\rightarrow \quad G(y) = \begin{cases} 0 & \text{if } y < 0 \\ y^2 & \text{if } 0 < y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

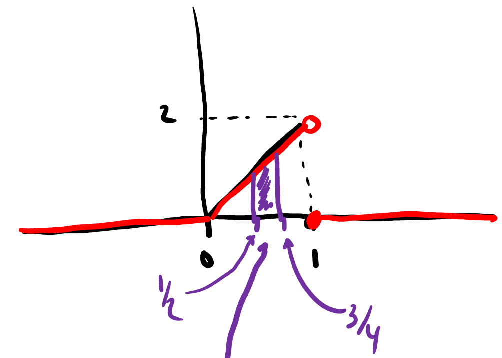


To compute $P\left(\frac{1}{2} < Y \le \frac{3}{4}\right) = ??$

best strategy whenever we're only given the pdf

Option 1 : use the pdf of Y

$g(y) = 2y$ for $0 < y < 1$



$$P\left(\frac{1}{2} < Y \le \frac{3}{4}\right) = \int_{1/2}^{3/4} g(y)\,dy = \int_{1/2}^{3/4} 2y \, dy$$

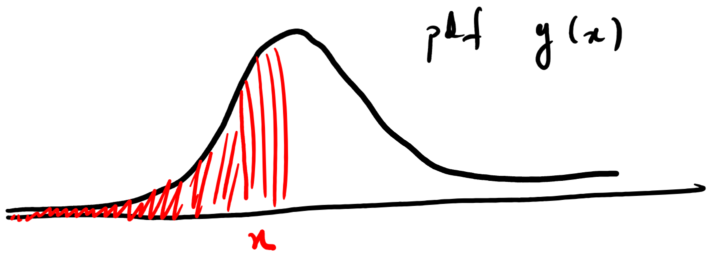
$$= \left[\, y^2 \,\right]_{y=1/2}^{y=3/4} = \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{5}{16}$$

area shaded purple is $\frac{5}{16}$

Option 2: use the cdf of Y

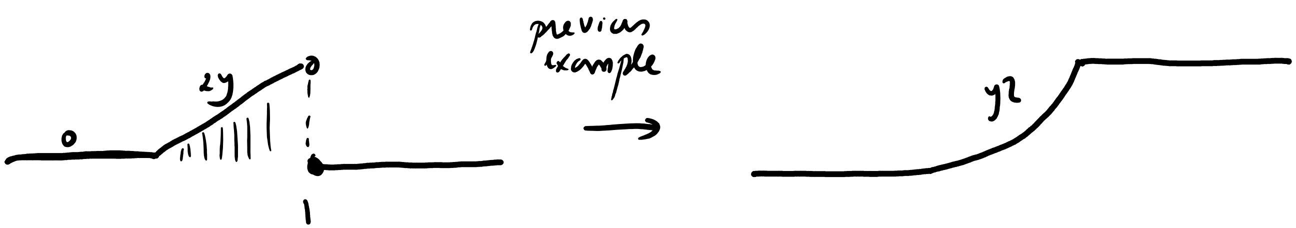
$$P\left(\frac{1}{2} < Y \le \frac{3}{4}\right) = \underbrace{G\left(\frac{3}{4}\right)}_{\text{area under graph of } g \text{ to the left of } 3/4} - \underbrace{G\left(\frac{1}{2}\right)}_{\text{area under graph of } g \text{ to the left of } 1/2} = \left(\frac{3}{4}\right)^2 - \left(\frac{1}{2}\right)^2 = \frac{5}{16}$$

area under graph of g between $1/2$ and $3/4$



cdf: $G(x) = P(X \le x) = \int_{-\infty}^{x} g(t)\,dt$ ← area under the graph of g to the left of x

FTC $\Rightarrow$ $G'(x) = g(x)$ whenever g is continuous: ← rate of change of the accumulated area under the graph of g is given by g



Mathematical Expectation

Let X be a continuous random variable with a pdf $f(x)$. Then:

In general, for a real-valued function u, we have:

$$\underbrace{E[u(X)]} = \int_{-\infty}^{\infty} \underbrace{u(x)} \cdot \underbrace{f(x)} \, dx$$

- the expected value of X, or the mean of X, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x$$

- the variance of X is

$$\sigma^2 = \mathrm{Var}(X) = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \, \mathrm{d}x$$

it still holds that

$$\underbrace{\mathrm{Var}(X) = E[X^2] - (E[X])^2}_{\text{true in general!}}$$

- the standard deviation of X is

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

- the moment-generating function (mgf), if it exists, is

$$M(t) = E\left(\mathrm{e}^{tX}\right) = \int_{-\infty}^{\infty} \mathrm{e}^{tx} f(x) \, \mathrm{d}x \,, \quad -h < t < h$$

Remarks

1) In the continuous case it is also valid that:

- $\sigma^2 = E\left(X^2\right) - \mu^2$ $\leftarrow$ $E[X^2] - (E[X])^2$

can still use mgf to compute moments!!!

- $\mu = M'(0)$
- $\sigma^2 = M''(0) - [M'(0)]^2$

2) The mgf, if it is finite for $-h < t < h$ for some $h > 0$, completely determines the distribution.

3) In both the discrete and continuous cases, if the rth moment, $E\left(X^r\right)$, exists and is finite, then the same is true of all lower-order moments

- the converse is not true: the first moment may exist and be finite, but the second moment is not necessarily finite.

4) If $M(t) = E\left(e^{tX}\right)$ exists and is finite for $-h < t < h$, then all moments exist and are finite, but the converse is not necessarily true.

$\rightarrow E[X^n] = M^{(n)}(0)$

Example (previous example continued)

Let Y be a continuous random variable with pdf $g(y) = 2y$ for $0 < y < 1$.

Find the mean and variance of Y.

Let Y be the r.v with pdf $g(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$

The mean of Y is

$$\mu = E[Y] = \int_{-\infty}^{\infty} y \cdot g(y)\, dy = \int_0^1 y \cdot 2y\, dy = \int_0^1 2y^2\, dy = \left[\frac{2}{3} y^3\right]_{y=0}^{y=1} = \frac{2}{3}$$

because $g(y) = 2y$ when $y \in (0,1)$ and is zero elsewhere

To determine the variance, we use the property:

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2$$

Since $E[Y^2] = \int_{-\infty}^{\infty} y^2 \cdot g(y)\, dy = \int_0^1 y^2 \cdot 2y\, dy = \int_0^1 2y^3\, dy$

$$= \left[\frac{2y^4}{4}\right]_{y=0}^{y=1} = \frac{1}{2}$$

and so

$$\mathrm{Var}(Y) = E[Y^2] - (E[Y])^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$