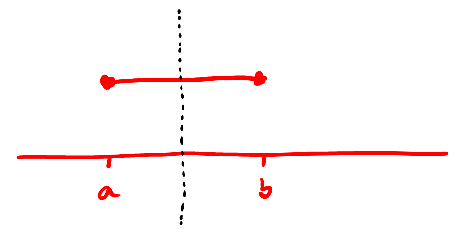
Math 3501 - Probability and Statistics I

3.1 - Random variables of the continuous type

Uniform distribution (from our last lesson)

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

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



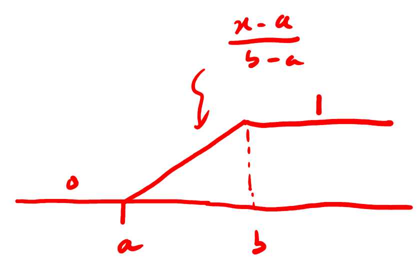
$\mu = \frac{b+a}{2}$

Its cdf

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

may be written as

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



Notation and terminology:

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

Compute the mean, variance, and mgf for a r.v. $X \sim \text{Uniform}(a,b)$.

Recall that the pdf of X is given by
$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

The mean of X is

$$\mu = E[X] = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx = \int_a^b x \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b x\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_{x=a}^{x=b}$$

because f is zero outside the interval $[a,b]$

$$= \frac{1}{b-a} \cdot \left(\frac{b^2}{2} - \frac{a^2}{2}\right)$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)}$$

$$= \frac{b+a}{2}$$

The variance of X is

$$\sigma^2 = \mathrm{Var}(X) \underset{\text{def}}{=} E\left[(X-\mu)^2\right] = \underbrace{E[X^2]}_{??} - \big(\underbrace{E[X]}_{\frac{b+a}{2}}\big)^2 = ??$$

$$E\left[X^2\right] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\, dx = \int_a^b x^2 \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b x^2\, dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_{x=a}^{x=b}$$

$$= \frac{1}{b-a}\left(\frac{b^3}{3} - \frac{a^3}{3}\right) = \frac{b^3 - a^3}{3(b-a)} = \frac{\cancel{(b-a)}(b^2+ab+a^2)}{3\cancel{(b-a)}} = \frac{b^2+ab+a^2}{3}$$

$$b^3 - a^3 = (b-a)(b^2+ab+a^2)$$

Review:

$$\underbrace{x^m - y^m} = (x-y) \cdot \underbrace{\text{polynomial on } x \text{ and } y}$$

quotient of $x^n - y^m$ divided by $x-y$

- long division
- synthetic division

Take $x^n - y^m$, if we set $x = y \Rightarrow y^n - y^n = 0 \Rightarrow y$ is a root of $x^n - y^n$

$\Rightarrow x-y$ is a factor of $x^n - y^n$

The variance is then

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} =$$

$$= \frac{4(b^2 + ab + a^2) - 3(b^2 + 2ab + a^2)}{12}$$

$$= \frac{4b^2 + 4ab + 4a^2 - 3b^2 - 6ab - 3a^2}{12} = \frac{b^2 - 2ab + a^2}{12}$$

$$= \frac{(b-a)^2}{12}$$

The mgf of X is

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\, dx = \int_a^b e^{tx} \cdot \frac{1}{b-a}\, dx = \frac{1}{b-a} \int_a^b e^{tx}\, dx =$$

For $t \neq 0$

$$\hookrightarrow = \frac{1}{b-a} \left[\frac{e^{tx}}{t} \right]_{x=a}^{x=b} = \frac{1}{b-a} \left(\frac{e^{bt}}{t} - \frac{e^{at}}{t} \right) = \frac{e^{bt} - e^{at}}{t(b-a)} \text{ for } t \neq 0$$

Recall that for $t=0$, we have

$$M(0) = E[e^{0 \cdot X}] = E[e^{0}] = E[1] = 1 \quad \Big\} \text{ true for any r.v.}$$

mean : $\mu = E[X] = M'(0)$

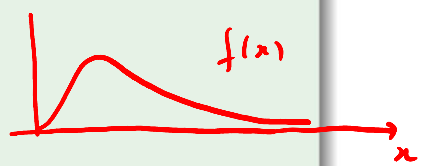
Conclusion :

$$M(t) = \begin{cases} \dfrac{e^{bt} - e^{at}}{t(b-a)}, & t \neq 0 \\ 1, & \text{if } t = 0 \end{cases}$$

Example

Let X have the pdf

$$f(x) = \begin{cases} xe^{-x} & 0 \leq x < \infty \\ 0 & \text{elsewhere} \end{cases}$$



Find the mgf of X, and use it to find the mean and variance of X.

The m.g.f of X is given by

$$e^{tx} \cdot e^{-x} = e^{tx-x} = e^{(t-1)x}$$

$$M(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\, dx = \int_0^{\infty} e^{tx} \cdot x e^{-x}\, dx = \underbrace{\int_0^{\infty} x \cdot e^{(t-1)x}\, dx}_{\text{IMPROPER INTEGRAL}}$$

QUESTION: For which values of t does the integral converge? We need to have $t < 1$ so that the coefficient of x on the exponential function is negative

if $t \geq 1$, then $\lim_{x \to \infty} x e^{(t-1)x} \neq 0$ and so the integral must diverge.

We will now see that the integral converges whenever $t < 1$ and obtain the limit in the process

Let us evaluate the following integral (with $t < 1$):

$$\int_0^{\infty} x\, e^{(t-1)x}\, dx = \lim_{b \to \infty} \int_0^{b} x\, e^{(t-1)x}\, dx =$$

$$= \lim_{b \to \infty} \left\{ \left[\frac{x\, e^{(t-1)x}}{t-1} \right]_{x=0}^{x=b} - \int_0^{b} \frac{e^{(t-1)x}}{t-1}\, dx \right\}$$

$$= \lim_{b \to \infty} \left\{ \frac{b\, e^{(t-1)b}}{t-1} - 0 - \left[\frac{e^{(t-1)x}}{(t-1)^2} \right]_{x=0}^{x=b} \right\}$$

$$= \lim_{b \to \infty} \left\{ \underbrace{\frac{b\, e^{(t-1)b}}{t-1}} - \underbrace{\frac{e^{(t-1)b}}{(t-1)^2}} + \underbrace{\frac{1}{(t-1)^2}} \right\}$$

Remember $t < 1$ and so $t - 1 < 0$

$$= 0 - 0 + \frac{1}{(t-1)^2} = \frac{1}{(t-1)^2}$$

CONCLUSION: The mgf of X is $\boxed{M(t) = \frac{1}{(t-1)^2}, \qquad t < 1}$

We now find the mean and variance of X by computing derivatives of $M(t)$:

$$M(t) = (t-1)^{-2} \implies M'(t) = -2(t-1)^{-3} \implies M''(t) = 6(t-1)^{-4}$$

The mean is then

$$\mu = E[x] = M'(0) = -2(0-1)^{-3} = 2$$

To compute the variance, we will use $Var(x) = E[x^2] - (E[x])^2$

Recall that $E[x^2] = M''(0) = 6(0-1)^{-4} = 6$ and so

$$Var(x) = E[x^2] - (E[x])^2 = 6 - 2^2 = 6 - 4 = 2$$

Why is the m.g.f helpful in this example?

If we try to compute $E[X]$ and $E[X^2]$ from the definition, we would have to evaluate

$$E[X] = \int_0^\infty x \cdot x e^{-x} \, dx = \int_0^\infty x^2 e^{-x} \leftarrow$$

improper integral **AND** integrate by parts twice

$$E[X^2] = \int_0^\infty x^2 \cdot x e^{-x} \, dx = \int_0^\infty x^3 e^{-x} \leftarrow$$

improper integral AND integrate by parts three times

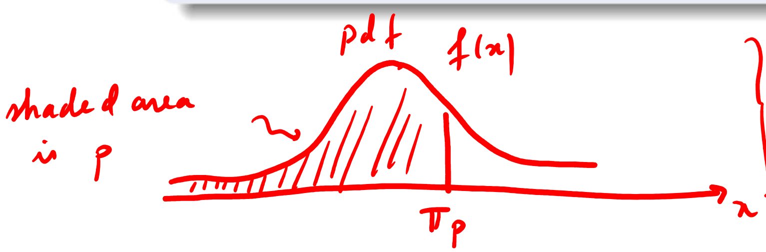
$(100p)$th percentile

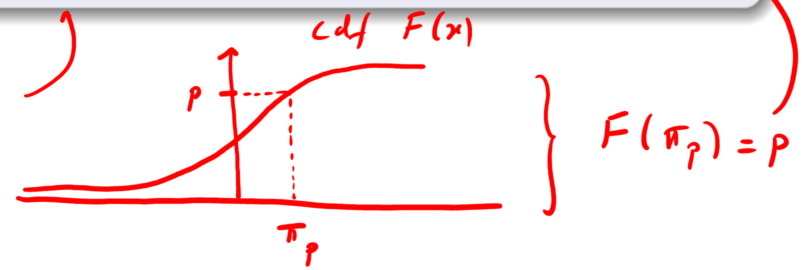
Definition

Let X be a continuous random variable with pdf $f(x)$ and cdf $F(x)$.

The $(100p)$*th percentile* is a number π_p such that the area under $f(x)$ to the left of π_p is p, that is

$$F\left(\pi_p\right) = \int_{-\infty}^{\pi_p} f(x)\,\mathrm{d}x = p\ .$$





Notation and terminology:

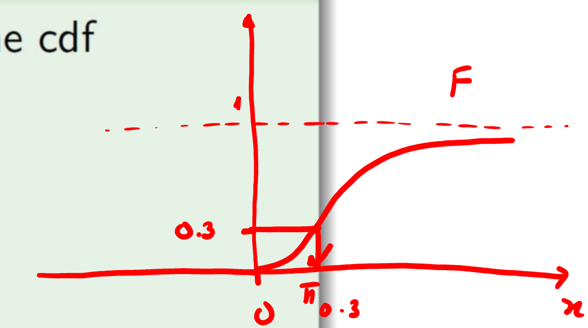
- The 50th percentile is called the median: $m = \pi_{0.50}$.
- The 25th and 75th percentiles are called the first and third quartiles, respectively, and are denoted by $q_1 = \pi_{0.25}$ and $q_3 = \pi_{0.75}$.
- The median $m = \pi_{0.50} = q_2$ is also called the second quartile.

Example

The time X in months until the failure of a certain product has the cdf

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ 1 - e^{-(x/4)^3} & 0 \le x < \infty \end{cases}$$

Find its 30th percentile $\pi_{0.3}$.



We want to find the value $\pi_{0.3}$ such that $\boxed{F(\pi_{0.3}) = 0.3}$ from the definition!

All we have to do is to solve the equation $F(x) = 0.3$ for x

We know from the definition of F that $x > 0$:

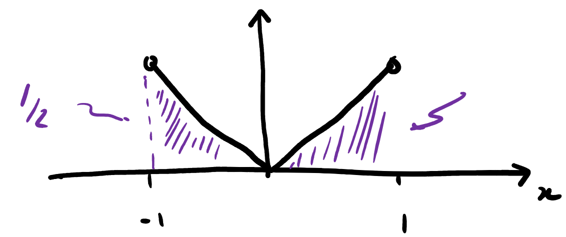
solve $1 - e^{-(x/4)^3} = 0.3$ for x to get

$$e^{-(x/4)^3} = 0.7 \iff -\left(\frac{x}{4}\right)^3 = \ln 0.7 \iff \left(\frac{x}{4}\right)^3 = -\ln 0.7 \iff x = 4(-\ln 0.7)^{1/3}$$

Conclusion $\pi_{0.3} = 4(-\ln 0.7)^{1/3} = 4\left(\ln \frac{10}{7}\right)^{1/3}$

EXAMPLE Let X be a continuous r.v. with pdf given by

$$f(x) = \begin{cases} |x| \,, & \text{if } -1 < x < 1 \\ 0 \,, & \text{otherwise} \end{cases}$$



Note that

① $f(x) \geq 0$ for all $x \in (-1,1)$

② $\int_{-1}^{1} f(x)\, dx = 1$

f is indeed a pdf!

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

Compute:

① $P(-1/2 < X < 3/4)$

② cdf of F

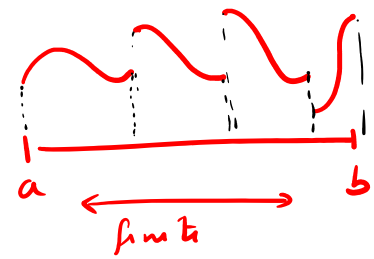
③ mean of X

④ variance of X

⑤ first quantile and 90th percentile.

(1)

$$P\left(-\tfrac{1}{2} < X < \tfrac{3}{4}\right) = \int_{-1/2}^{3/4} f(x)\,dx = \int_{-1/2}^{3/4} |x|\,dx$$



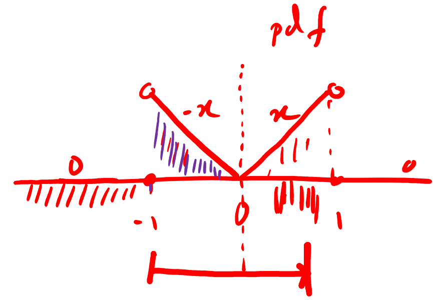
$$= \int_{-1/2}^{0} -x\,dx + \int_{0}^{3/4} x\,dx = \left[-\frac{x^2}{2}\right]_{x=-1/2}^{x=0} + \left[\frac{x^2}{2}\right]_{x=0}^{x=3/4}$$

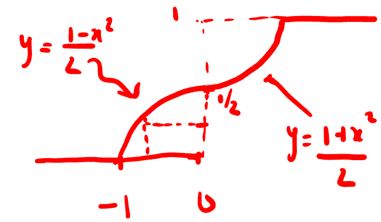
$$= -0 + \frac{(-1/2)^2}{2} + \frac{(3/4)^2}{2} - 0 = \frac{9}{32} + \frac{1}{8} = \cdots$$

(2) cdf of X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = ???$$

$$F(x) = \begin{cases} 0 & , \quad x < -1 \\ \dfrac{1-x^2}{2} & , \quad -1 \leq x \leq 0 \\ \dfrac{1+x^2}{2} & , \quad 0 < x < 1 \\ 1 & , \quad x \geq 1 \end{cases}$$





For $x \in [-1,0]$, we have:

$f(t) = -t$ for all $t \in [-1,0]$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-1}^{x} -t\,dt = -\frac{t^2}{2}\Big]_{t=-1}^{t=x} = -\frac{x^2}{2} + \frac{(-1)^2}{2}$$

$$= \frac{1-x^2}{2}$$ ⇝ note that when $x = -1$, $\frac{1-(-1)^2}{2} = 0$

and when $x = 0$, $\frac{1-0^2}{2} = \frac{1}{2}$

For $x \in (0,1)$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \underbrace{\int_{-1}^{0} -t\,dt}_{1/2} + \int_{0}^{x} t\,dt = \frac{1}{2} + \left[\frac{t^2}{2}\right]_{t=0}^{t=x} = \frac{1}{2} + \frac{x^2}{2} = \frac{1+x^2}{2}$$

(3) mean of X is

$$\mu = E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_{-1}^{1} x \cdot |x|\, dx = \int_{-1}^{0} x(-x)\, dx + \int_{0}^{1} x \cdot x\, dx$$

because $|x| = -x$ when $x < 0$

because $|x| = x$ when $x > 0$

$$= \int_{-1}^{0} -x^2\, dx + \int_{0}^{1} x^2\, dx$$

$$= \left[-\frac{x^3}{3}\right]_{x=-1}^{x=0} + \left[\frac{x^3}{3}\right]_{x=0}^{x=1} = -0 + \frac{(-1)^3}{3} + \frac{1^3}{3} - 0$$

$$= 0$$

(4) similar to (3), we $Var(X) = E[X^2] - (E[X])^2$

(5) First quantile:

$\pi_{0.25}$ satisfies $F(\pi_{0.25}) = 0.25$ ← $\pi_{0.25}$ must be between -1 and 0 because the cdf grows from 0 at $x=-1$ to $1/2$ at $x=0$

Need to use the branch of the cdf with $\underline{\underline{-1 < x < 0}}$ ←

Solve $\dfrac{1-x^2}{2} = \dfrac{1}{4}$ $\iff$ $1 - x^2 = \dfrac{1}{2}$ $\iff$ $x^2 = \dfrac{1}{2}$

$\iff$ $x = -\sqrt{\dfrac{1}{2}} = -\dfrac{\sqrt{2}}{2}$

For $\pi_{0.9}$, we would use the branch for $0 < x < 1$ ...