Remember From Last Time:

The cumulative distribution function for any random variable X, denoted $F(x)$, is

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

The CDF has the following properties for all values x

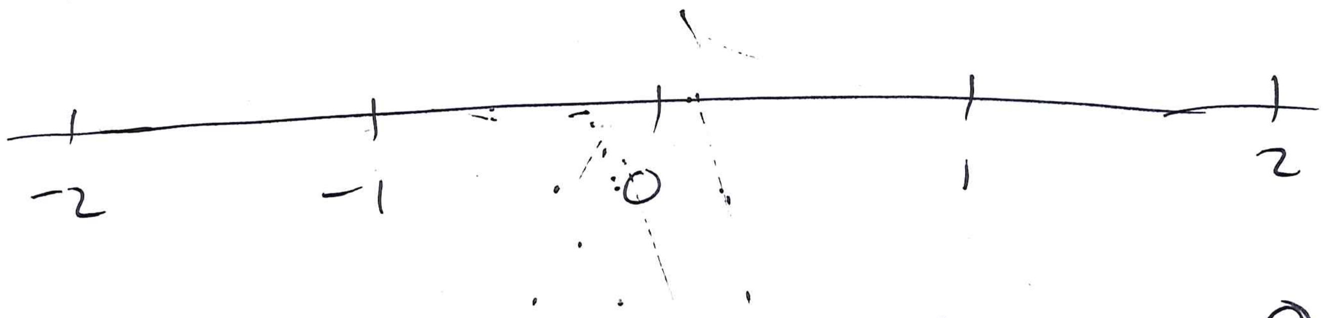
- $0 \leq F(x) \leq 1$
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $F(x)$ is a nondecreasing function of x

Also, if X is a continuous random variable then $Pr(X = x) = 0$ for any value x.

* Memorize properties + definitions *

Probability Density Function

ex: drop pen on a number line



but: cluster around 0

$Pr(X=x)=0$

$Pr(X=0)=0$

$Pr(X=2)=0$

How should we describe clustering?

$Pr(X \text{ "near" } 0) > Pr(X \text{ "near" } 2)$

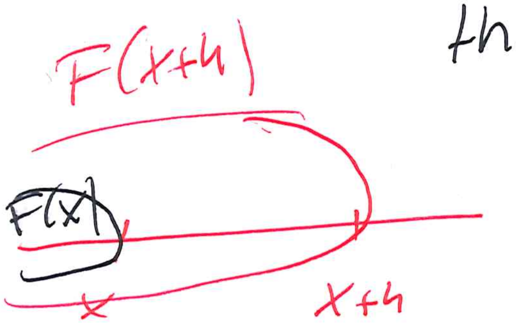
Density of points: $\frac{\# \text{pts}}{\text{length}}$

ex: 10 pts with [0, 1]:

density: $\frac{10}{1}$

$Pr\left(-\frac{1}{2} < X < \frac{1}{2}\right) > Pr(-2 < X < -1.5)$

On interval $[x, x+h]$, we can think of density as:



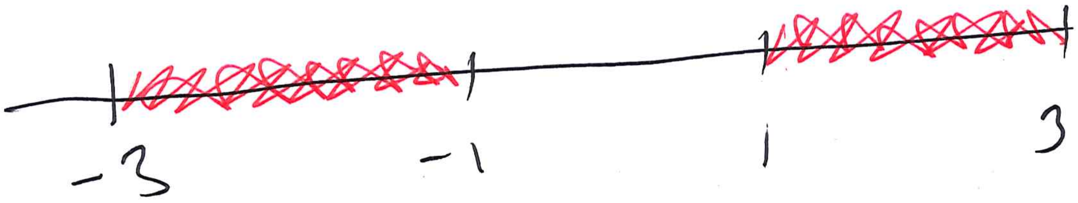
$$\frac{Pr(x \leq X \leq x+h)}{h} = \underbrace{\frac{F(x+h) - F(x)}{h}}_{\text{Looks like def of derivative!}}$$

Let $F(x)$ be the cumulative distribution function for a continuous random variable X. The probability density function (PDF) of X is:

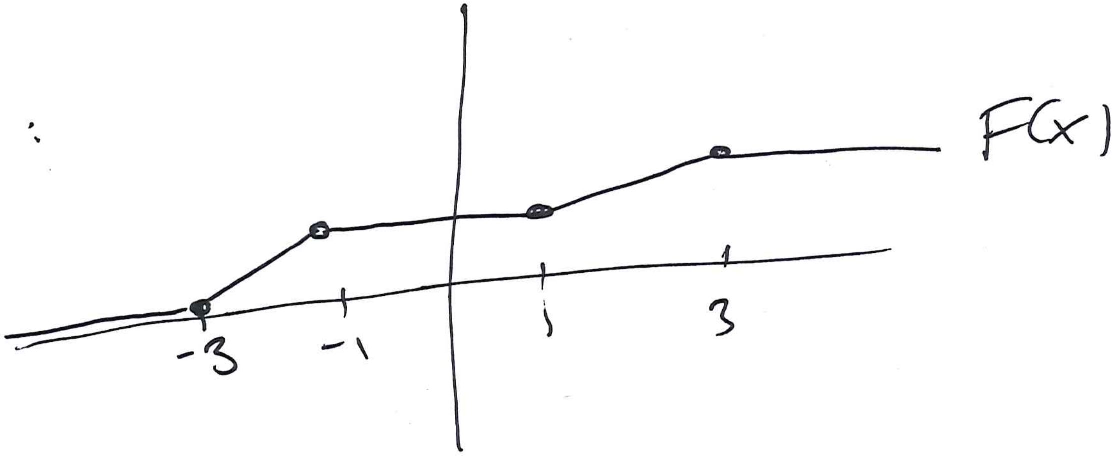
$$f(x) = \frac{d}{dx}\{F(x)\}$$

where the derivative exists.

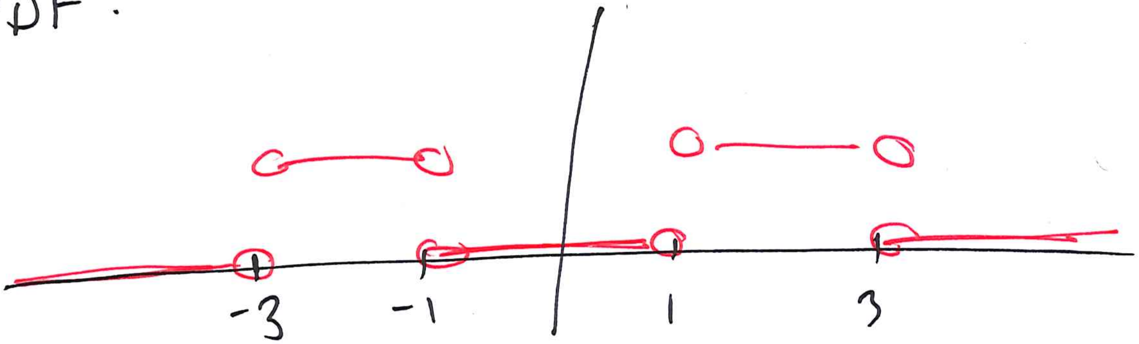
ex from last time:



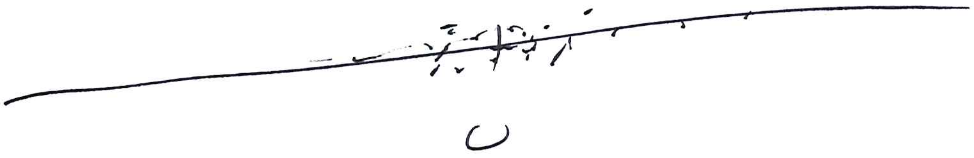
CDF:



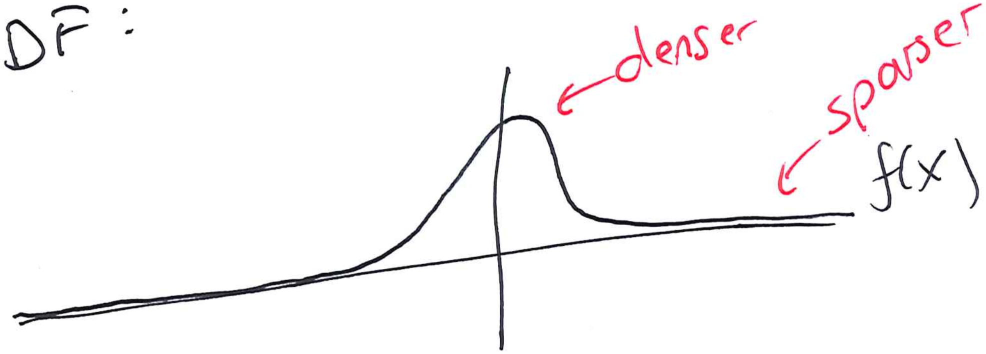
PDF:



Pen-dropping ex:



PDF:



Usally in practice we use

$$\underset{\text{CDF}}{F(x)} = \int_{-\infty}^{x} \underset{\text{PDF}}{f(t)}\,dt$$

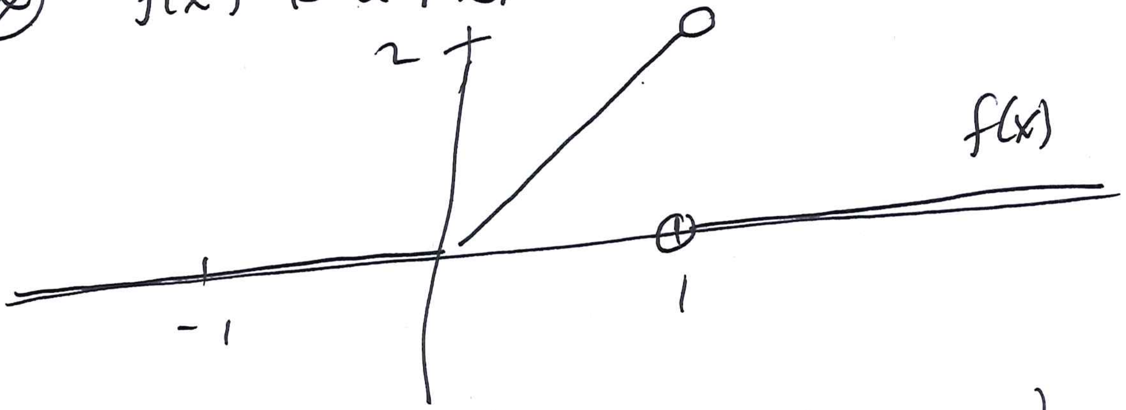
instead of

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

If $f(x)$ is probability density function (PDF) of a continuous random variable X:

$$\Pr(a \leq X \leq b) = \int_a^b f(x)\,dx \qquad \circledast$$

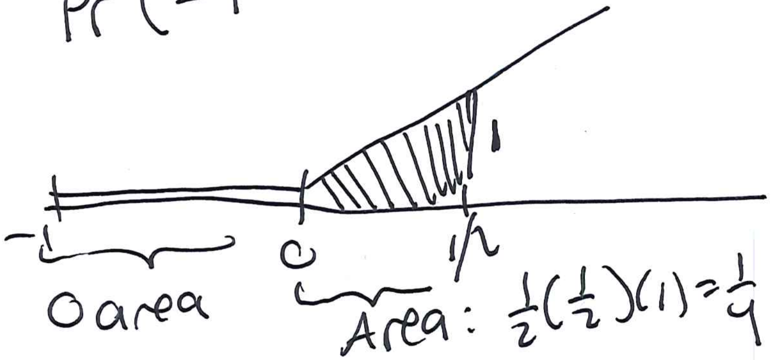
ⓔⓧ $f(x)$ is a PDF



Choices:

$\Pr(-1 < X < 0) = 0$ (area under $f(x)$)

$\Pr(-1 < X < 1/2) = 0 + \frac{1}{4} = \frac{1}{4}$



Note: If X is continuous

$Pr(X=x)=0$, so

$$\underline{\underline{Pr(X \leq x)}} = Pr(X<x) + \underbrace{P(X=x)}_{0}$$

$$= \underline{\underline{Pr(X<x)}}$$

Properties of a PDF:

If $f(x)$ is a PDF for continuous random variable X:

(a) $F(x) = Pr(X \leq x) = \int_{-\infty}^{x} f(t)dt$ ~~f(x)dx~~

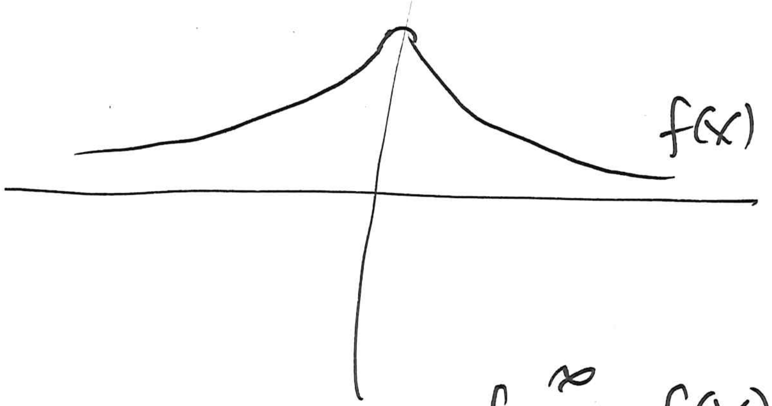
(b) $f(x) \geq 0$ for any x

(c) $\int_{-\infty}^{\infty} f(x)\,dx = Pr(-\infty \leq X \leq \infty) = 1$

understand => memorize

ⓔⓧ $f(x) = a e^{-|x|}$ for some constant a

If f(x) is a PDF, what is a?



Picture:

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

$$1 = \int_{-\infty}^{\infty} a e^{-|x|}\,dx = \underbrace{\int_{-\infty}^{0} a e^{-|x|}\,dx}_{\substack{x \leq 0, \text{ so} \\ |x| = -x \\ \text{then: } -|x| = \\ -(-x) = x}} + \overbrace{\int_{0}^{\infty} a e^{-|x|}\,dx}^{\substack{x \geq 0, \text{ so } |x| = x \\ -|x| = -x}}$$

$$= \int_{-\infty}^{0} a e^{x}\,dx + \int_{0}^{\infty} a e^{-x}\,dx$$

$$\lim_{r\to-\infty}\int_r^0 ae^{x}dx \quad + \quad \lim_{s\to\infty}\int_0^s ae^{-x}dx$$

$$\lim_{r\to-\infty}\left(ae^{x}\Big|_r^0 \right) + \lim_{s\to\infty}\left(-ae^{-x}\Big|_0^s \right)$$

$$\lim_{r\to-\infty}\left(\underbrace{ae^{0}}_{a} \underbrace{-ae^{r}}_{0} \right) + \lim_{s\to\infty}\left(\underbrace{-ae^{-s}}_{0} \underbrace{+ae^{0}}_{a} \right) \quad \Big\| \quad e^{-s} = \frac{1}{e^{s}} \to 0$$

$= 2a = 1$ (properties of a PDF)

So: $a = 1/2$

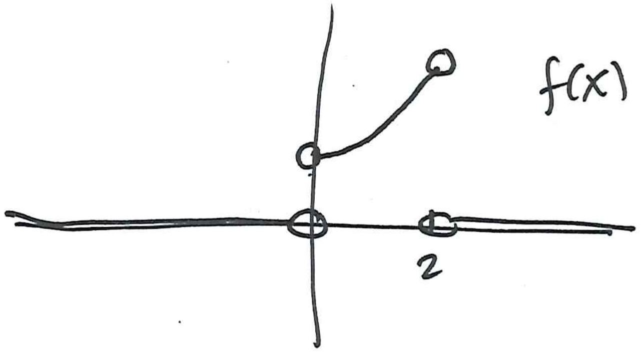
PDF: $f(x) = \frac{1}{2}e^{-|x|}$

ex

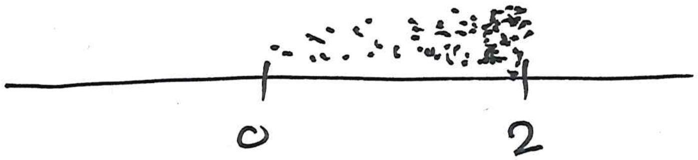
$$f(x) = \begin{cases} k(3x^2+1) & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

is a PDF

density



Values:



What is the CDF of X ?

Properties:

$$F(x) = Pr(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

eg $x = 5$

$F(5) = Pr(X \le 5) = 1$

$F(6) = Pr(X \le 6) = 1$

$F(1.9) = Pr(X \le 1.9) <$

If $x < 0$: $F(x) = 0$

If $x > 2$: $F(x) = Pr(X \le x) = 1$

If x in $[0, 2]$:

$$F(x) = Pr(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

$$= \underbrace{\int_{-\infty}^{0} f(t)\,dt}_{0} + \int_{0}^{x} f(t)\,dt$$

$$= \int_{0}^{x} \underbrace{k(3t^2+1)}_{f(t)}\,dt$$

$$= k\left(t^3+t\Big|_0^x\right) = k(x^3+x)$$

$$F(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ k(x^3+x) & \text{if} \quad 0 \le x \le 2 \\ 1 & \text{if} \quad x > 2 \end{cases}$$

Expected Value

Motivating Example

3 studenst 70%
2 students 50%

Compute average:
"weight" each value
by its likelihood

Average:

$$\frac{70+70+70+50+50}{5}$$

$$= \frac{3\cdot 70 + 2\cdot 50}{5}$$

$$= \frac{3}{5}(70) + \frac{2}{5}(50)$$

$\uparrow$ $\frac{3}{5}$ of students scored 70

$\uparrow$ $\frac{2}{5}$ of students scored 50

The expected value (also called "expectation" or "mean") of a continuous random variable X with probability density function f(x) is

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

(memorize)

$$\int_{-\infty}^{\infty} x \cdot f(x) \, dx$$

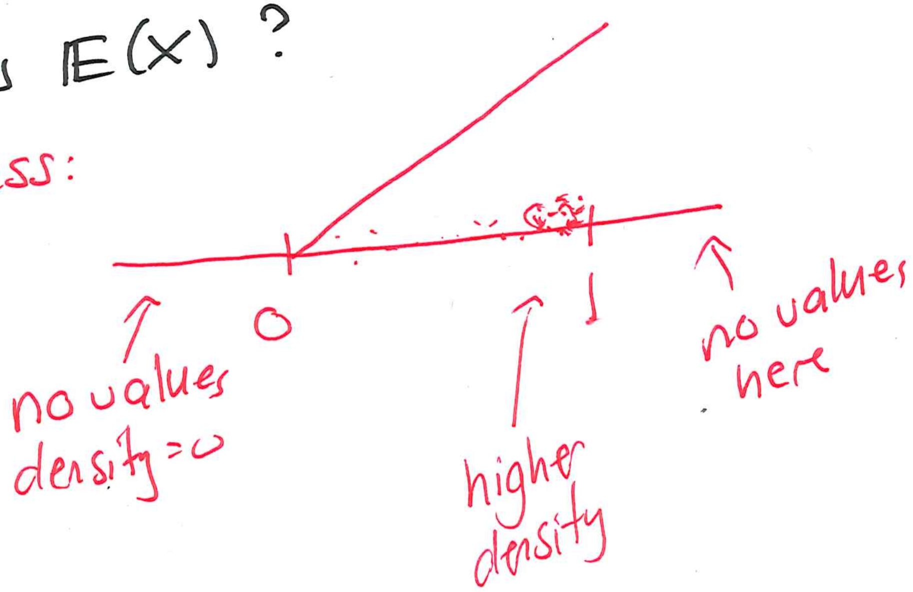
likelihood

all possible values

ⓔⓧ Suppose $f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2x & \text{if } 0 \le x \le 1 \\ 0 & x > 1 \end{cases}$ is PDF.

What is $\mathbb{E}(X)$?

Guess:



all values in [0,1]
so expect
mean is [0,1]

More big #s than small:
avg should be closer to 1 than to 0
(> 1/2)

Using definition:

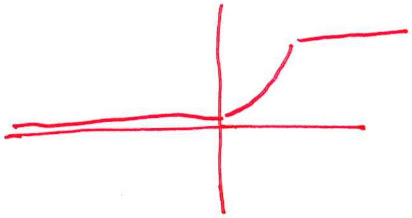
$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$ ⊛⊛ ⊛

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

$$\int_0^1 x(2x)\, dx = \frac{2}{3}x^3\Big|_0^1 = \boxed{\frac{2}{3}}$$

If I do X many, many times, average results, should be 2/3

ⓔⓧ $F(x) = \begin{cases} 0 & x < 0 \\ x^2 + \frac{3}{2}x & 0 \le x \le \frac{1}{2} \\ 1 & x > \frac{1}{2} \end{cases}$ is a CDF for X



What is $\mathbb{E}(X)$?

Def of $\mathbb{E}(X)$ uses $f(x)$ (density)

not $F(x)$ (cumulative)

Recall: $F'(x) = f(x)$

① Find $f(x)$ by differentiating $F(x)$

② Find $\mathbb{E}(X)$ def: $\int_{-\infty}^{\infty} x f(x)\, dx$

① $F'(x) = f(x) = \begin{cases} 0 & x < 0 \\ 2x + \frac{3}{2} & 0 < x < \frac{1}{2} \\ 0 & x > 1/2 \end{cases}$ (deriv of constant)

② $$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \underbrace{\int_{-\infty}^{0} x \underbrace{f(x)}_{0} dx}_{0} + \int_{0}^{1/2} \underline{\underline{\left(2x+\tfrac{3}{2}\right)}}\, x\, dx + \underbrace{\int_{1/2}^{\infty} x \underbrace{f(x)}_{0} dx}_{0}$$

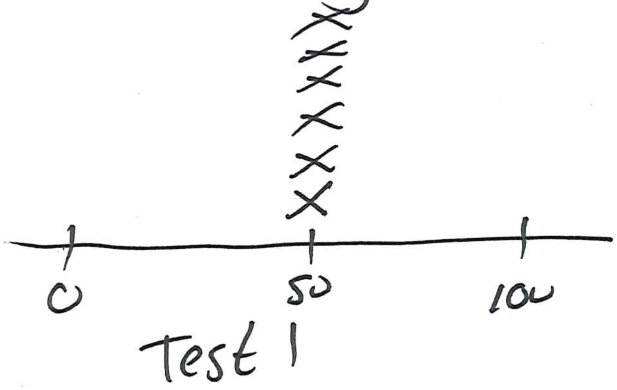
$$= \int_{0}^{1/2} \left(2x^2 + \tfrac{3}{2}x\right) dx = \tfrac{2}{3}x^3 + \tfrac{3}{4}x^2 \Big|_{0}^{1/2}$$

$$= \tfrac{2}{3}\left(\tfrac{1}{8}\right) + \tfrac{3}{4}\left(\tfrac{1}{4}\right) = \tfrac{1}{12} + \tfrac{3}{16} = \tfrac{4}{48} + \tfrac{9}{48} = \boxed{\tfrac{13}{48}}$$

If we average ∞ many samples from this random variable, we get 13/48

Variance & Standard Deviation

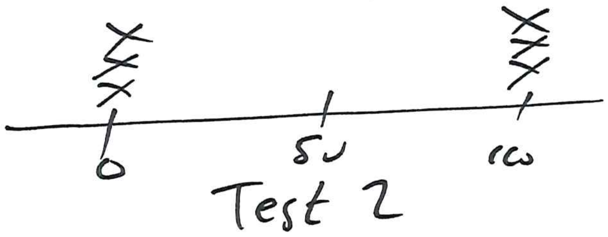
Motivating Example



Avg: $\frac{6 \cdot 50}{6} = \boxed{50}$

Everyone gets exactly the average

Low variance



Avg: $\frac{1}{2}(0) + \frac{1}{2}(100)$

$= \boxed{50}$

Nobody gets the average
Everybody is 50 pts away from average

High Variance

Idea : $x - \mathbb{E}(X)$: diff b/w value x and the average (could be + or −)

$(x - \mathbb{E}(X))^2$: destroying information about sign

Average this: $$\int_{-\infty}^{\infty} \underbrace{(x - \mathbb{E}(x))^2}_{\text{value to avg}} \cdot \underset{\text{likelihood}}{f(x)} \, dx$$

The variance of a continuous random variable X with PDF $f(x)$ is

$$Var(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x) \, dx$$

Fact: also $$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

(alternate way to calculate – give same answer)

The <u>standard deviation</u> of X is

$$\underset{\substack{\nearrow \\ \text{sigma}}}{\sigma}(X) = \sqrt{Var(X)}$$

---

ex PDF from before: $f(x) = \begin{cases} 0 & x < 0 \\ 2x & 0 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$, $\mathbb{E}(X) = 2/3$

Find $Var(X)$, $\sigma(X)$

$$Var(X) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x)\,dx$$

$$= 0 + \int_0^1 \left(x - \frac{2}{3}\right)^2 \cdot 2x\,dx$$

$$= \int_0^1 2x\left(x^2 - \frac{4}{3}x + \frac{4}{9}\right)dx$$

$$= \int_0^1 \left(2x^3 - \frac{8}{3}x^2 + \frac{8}{9}x\right)dx$$

$$= \frac{2}{4}x^4 - \frac{8}{9}x^3 + \frac{4}{9}x^2 \Big|_0^1$$

$$= \frac{1}{2} - \frac{8}{9} + \frac{4}{9} = \frac{1}{2} - \frac{4}{9} = \frac{9-8}{18} = \boxed{\frac{1}{18}} = \mathrm{Var}(X)$$

(Method 1)

Method 2: $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$

$$= \mathbb{E}(X^2) - \left(\frac{2}{3}\right)^2$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} \underset{\substack{\uparrow \\ \text{values} \\ \text{to average}}}{x^2} \; \underset{\substack{\uparrow \\ \text{likelihood}}}{f(x)} \, dx = 0 + \int_0^1 x^2(2x)\,dx$$

$$= \int_0^1 2x^3\,dx = \frac{2}{4}x^4\Big|_0^1 = \frac{1}{2}$$

$$\mathrm{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9-8}{18} = \boxed{\frac{1}{18}} \quad \text{(same)}$$

Standard dev:

$$\sigma(X) = \sqrt{Var(X)} = \sqrt{\frac{1}{18}} = \boxed{\frac{1}{3\sqrt{2}}}$$

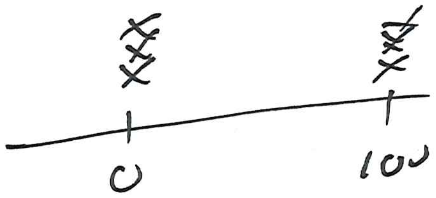
Idea: "average" outcome of experiment

$$\mathbb{E}(X) = 2/3$$

"average" difference between "reality + expectation"

$$\sigma(X) = \frac{1}{3\sqrt{2}}$$

High Var: avg not good descriptor for individuals



Low var: avg is good descriptor for individuals