Probability 4: Continuous Random Variables

Module Outline

- Review of previous Prob. Module
- Covariance and Correlation
- Prob. distributions of continuous r.v.s: c.d.f. and density
- Expectation and Variance
- Common continuous distributions

Review: Prob. Module 3

- Random variables: Discrete

- Probability mass function (p.m.f.) $p(x) = Prob(\widetilde{X} = x)$
- Cumulative distribution fn. (c.d.f.):
  $Prob(\widetilde{X} \leq x) = \sum_{y:\ y \leq x} p(y)$

Review: Prob. Module 3

- Random variables: Discrete

- Probability mass function (p.m.f.) $p(x) = Prob(\widetilde{X} = x)$

  Cumulative distribution fn. (c.d.f.):
  $Prob(\widetilde{X} \leq x) = \sum_{y:\ y \leq x} p(y)$

- *Expectation*: $E\widetilde{X} = \mu_X = \sum_{\text{all } x} xp(x)$

  Variance: $\sigma_X^2 = E[(X - \mu_X)^2] = E(X^2) - (\overline{EX})^2$

  μ_X

Review: Prob. Module 3

- Common discrete distributions:

  Bernoulli (single trial): $\widetilde{Z} = 1$ or 0

  Binomial (number of successes in n trials):

  For $x \in \{0, 1, 2, \ldots, n\}$ $\quad Prob(\widetilde{X} = x) = \binom{n}{x} p^x (1-p)^{n-x}$

  n, p

Review: Prob. Module 3

- Common discrete distributions:

  Bernoulli (single trial): $\widetilde{Z} = 1$ or 0

  Binomial (number of successes in n trials):

  $$\text{For } x \in \{0, 1, 2, ..., n\} \quad Prob(\widetilde{X} = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- *Poisson* (number of occurrences in a time period/area)

  $$\text{For } x \in \{0, 1, 2, ..., \infty\} \quad Prob(\widetilde{X} = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$x! = x(x-1)(x-2)\cdot\cdots\cdot 1$

parameter $= \lambda =$ mean

Clicker question 1

- **Q**: Which one of the following would be appropriate to be modelled as a Poisson distribution?

  (a) Whether a roll of the die comes up even or not

  (b) Number of car-owners in a sample of 100 adults

  (c) Number of customers arriving at a bank during an hour

  (d) Number who get A in a class of 200 students

Solution to clicker question 1

- **Q**: Which one of the following would be appropriate to be modelled as a Poisson distribution?

  (a) Whether a roll of the die comes up even or not Bernoulli $\tilde{Z} = 1/0$

  (b) Number of car-owners in a sample of 100 adults Binomial $n = 100$

  (c) Number of customers arriving at a bank during an hour Poisson

  (d) Number who get A in a class of 200 students $n = 200$ Binomial

Covariance and Correlation

- **Expectation** $\mu_X = E\widetilde{X}$: mean/average value of $\widetilde{X}$

Variance $Var(\widetilde{X})$: dispersion of $\widetilde{X}$

σ_x^2

Covariance and Correlation

- **Expectation** $\mu_X = E\widetilde{X}$: mean/average value of $\widetilde{X}$

  Variance $Var(\widetilde{X})$: dispersion of $\widetilde{X}$

- **Covariance**$(\widetilde{X}, \widetilde{Y})$: relationship between $\widetilde{X}$ and $\widetilde{Y}$

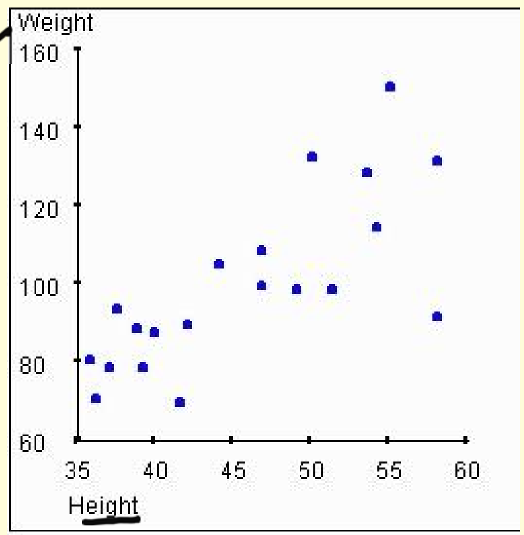
Covariance and Correlation

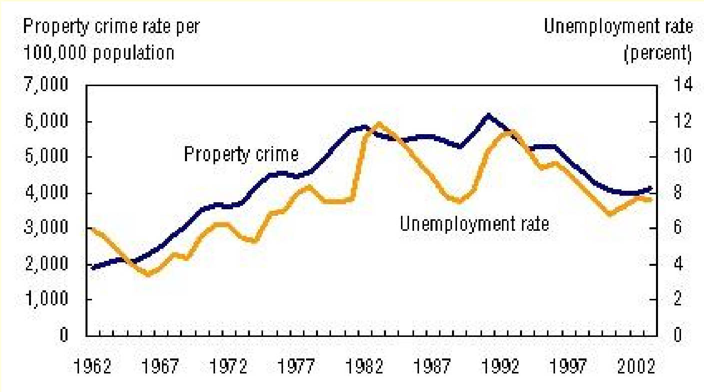
- **Expectation** $\mu_X = E\widetilde{X}$: mean/average value of $\widetilde{X}$

  Variance $Var(\widetilde{X})$: dispersion of $\widetilde{X}$

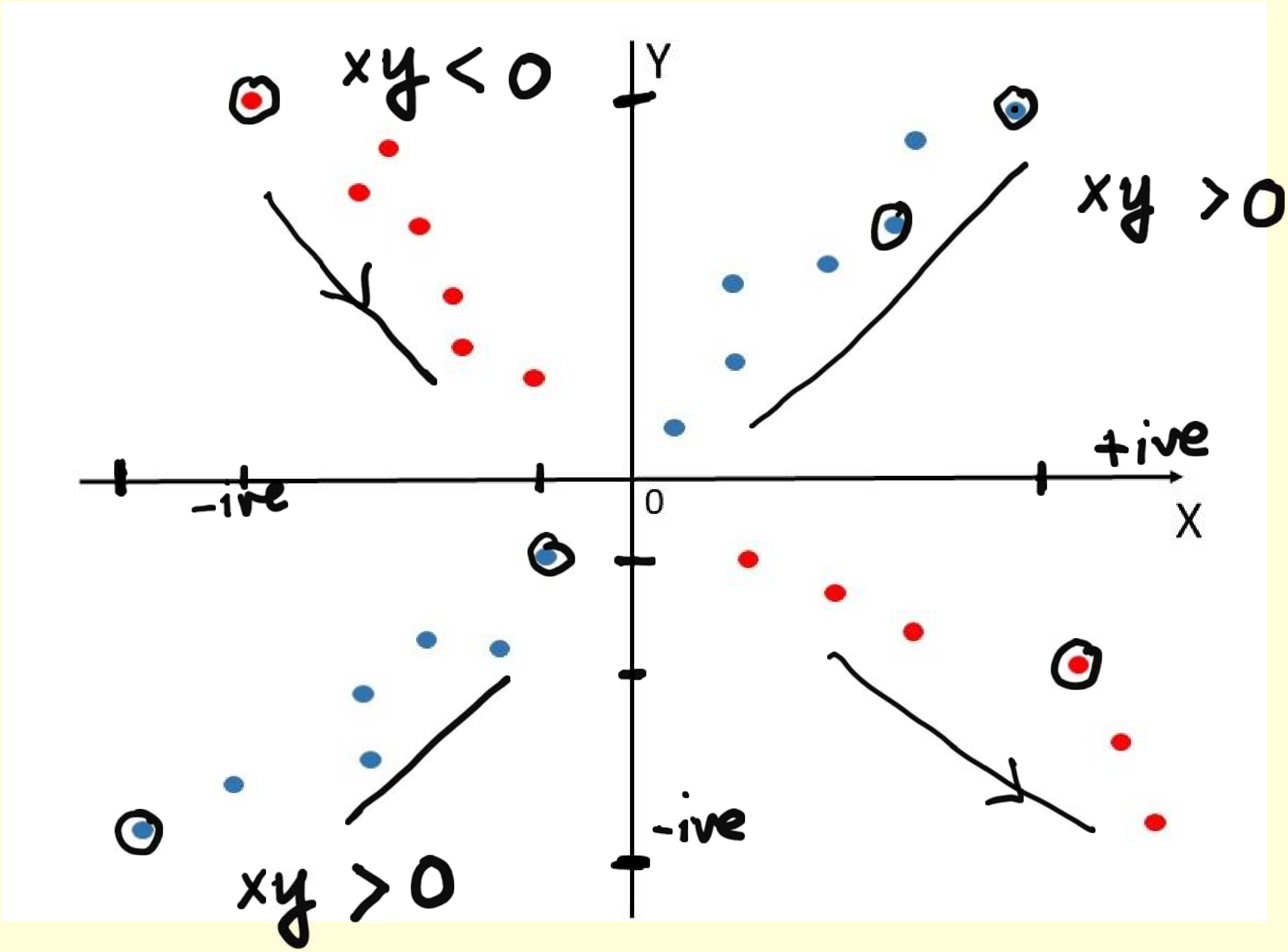
- **Covariance**$(\widetilde{X}, \widetilde{Y})$: relationship between $\widetilde{X}$ and $\widetilde{Y}$

- Examples



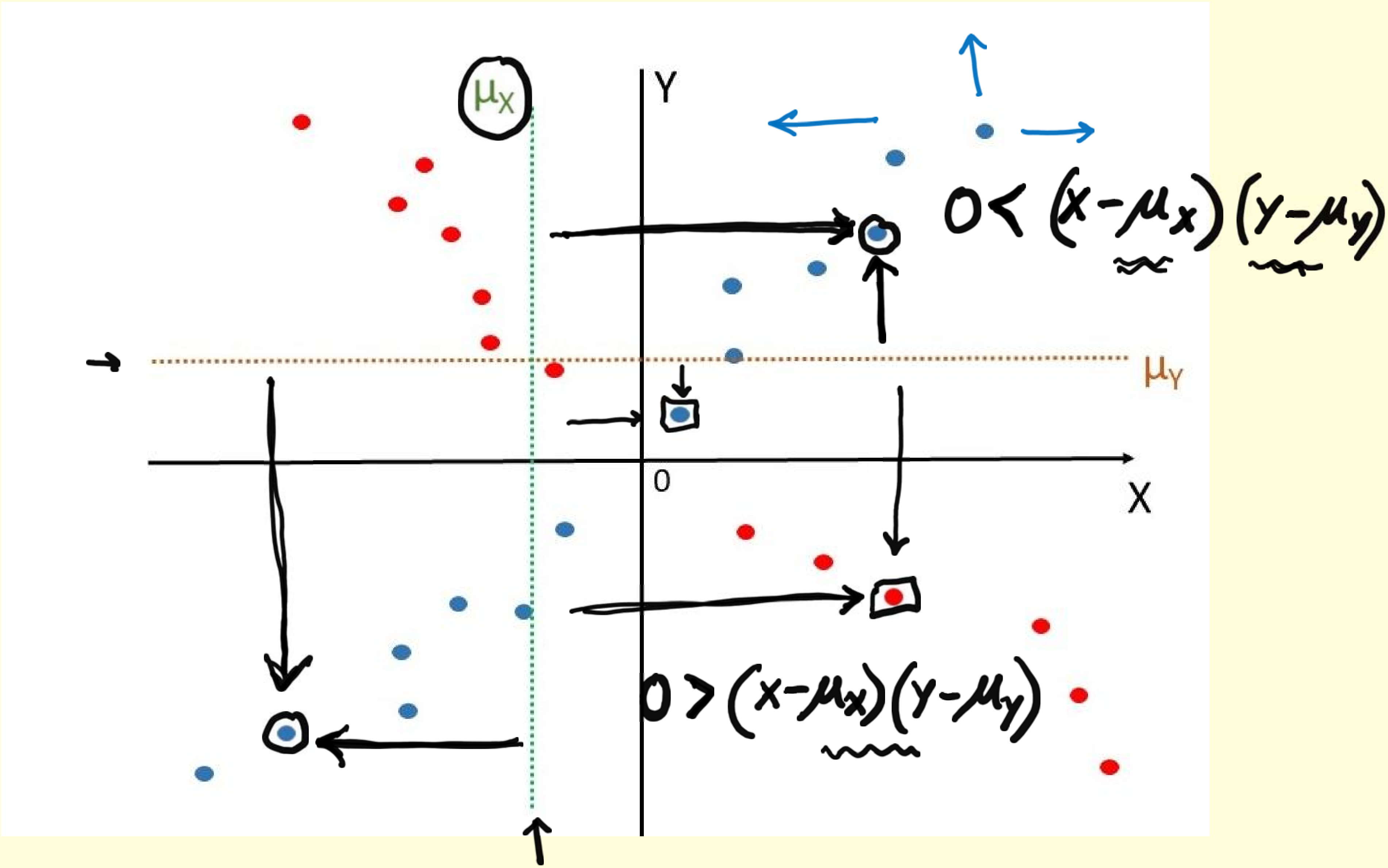


Covariance: Basic idea



•

Covariance: Basic idea



Covariance: definition

sigma

- $Cov(\widetilde{X}, \widetilde{Y}) = \sigma_{XY} = E(\widetilde{X} - \mu_X)(\widetilde{Y} - \mu_Y)$

$$= \sum_{\text{all } x,y} (x - \mu_X)(y - \mu_Y)\, p(x, y)$$

Covariance: an easier formula

$$Var(\tilde{X}) = E\tilde{X}^2 - (EX)^2$$

- $Cov(\widetilde{X}, \widetilde{Y}) = \sigma_{XY} = E\left[(\widetilde{X} - \mu_X)(\widetilde{Y} - \mu_Y)\right]$

$$= E\left[\tilde{X}\tilde{Y} - \tilde{X}\mu_y - \mu_x\tilde{Y} + \mu_x\mu_y\right]$$

$$= E[\tilde{X}\tilde{Y}] - E[\tilde{X}\mu_y] - E[\mu_x\tilde{Y}] + E[\mu_x\mu_y]$$

$$= E[\tilde{X}\tilde{Y}] - \mu_y E[\tilde{X}] - \mu_x E[\tilde{Y}] + \mu_x\mu_y$$

$$= E[\tilde{X}\tilde{Y}] - \mu_y\mu_x - \cancel{\mu_x\mu_y} + \cancel{\mu_x\mu_y}$$

$$= E[\tilde{X}\tilde{Y}] - E[Y]E[X]$$

Covariance: an easier formula

- $Cov(\widetilde{X}, \widetilde{Y}) = \sigma_{XY} = E(\widetilde{X} - \mu_X)(\widetilde{Y} - \mu_Y)$
- $Cov(\widetilde{X}, \widetilde{Y}) = E(\widetilde{X} - \mu_X)(\widetilde{Y} - \mu_Y) = \underline{E(\widetilde{X}\widetilde{Y}) - E(\widetilde{X})E(\widetilde{Y})}$

Computing covariance: an example

$$E X = 10 \times 0.6 + 20 \times 0.4 = 14$$

- Joint pmf of $\widetilde{X}$ and $\widetilde{Y}$:

			y		
	$f(x,y)$	**1**	**2**	**3**	$f_X(x)$
x	**10**	0	0.2	0.4	0.6
	20	0.2	0.2	0	0.4
	$f_Y(y)$	0.2	0.4	0.4	1

$$E Y = 1 \times 0.2 + 2 \times 0.4 + 3 \times 0.4 = 2.2$$

- $Cov(\widetilde{X}, \widetilde{Y}) = \sigma_{XY} = E(\widetilde{X}\widetilde{Y}) - E(\widetilde{X})E(\widetilde{Y}) =$ $28 - 14 \times 2.2 = -2.8$

$$E(XY) = 10 \times 0 + 20 \times 0.2 + 30 \times 0.4 + 20 \times 0.2 + 40 \times 0.2 + 60 \times 0 = 28$$

Clicker question 2

- **Q**: Which one of the following is measured by *covariance*?

  (a) The average of a random variable

  (b) The dispersion of a random variable

  (c) The maximum value of a random variable

  (d) Relationship between two random variables

Solution to clicker question 2

- **Q**: Which one of the following is measured by *covariance*?

  (a) The average of a random variable $\rightarrow E\tilde{X}$

  (b) The dispersion of a random variable $\rightarrow$ Var

  (c) The maximum value of a random variable

  (d) Relationship between two random variables

Covariance: some properties

- $Cov(\widetilde{X}, \widetilde{Y}) = Cov(\widetilde{Y}, \widetilde{X})$

Covariance: some properties

- $Cov(\widetilde{X}, \widetilde{Y}) = Cov(\widetilde{Y}, \widetilde{X})$
- If $\widetilde{X}$ and $\widetilde{Y}$ are independent, $Cov(\widetilde{X}, \widetilde{Y}) = 0$.

Covariance: some properties

- $Cov(\widetilde{X}, \widetilde{Y}) = Cov(\widetilde{Y}, \widetilde{X})$
- If $\widetilde{X}$ and $\widetilde{Y}$ are independent, $Cov(\widetilde{X}, \widetilde{Y}) = 0$.
- $Cov(\widetilde{X} + c, \widetilde{Y} + d) = Cov(\widetilde{X}, \widetilde{Y})$

Covariance: some properties

- $Cov(\widetilde{X}, \widetilde{Y}) = Cov(\widetilde{Y}, \widetilde{X})$
- If $\widetilde{X}$ and $\widetilde{Y}$ are independent, $Cov(\widetilde{X}, \widetilde{Y}) = 0$.
- $Cov(\widetilde{X} + c, \widetilde{Y} + d) = Cov(\widetilde{X}, \widetilde{Y})$
- $Cov(c\widetilde{X}, d\widetilde{Y}) = cdCov(\widetilde{X}, \widetilde{Y})$

H W

m kg

yds lbs

$\frac{1}{0.9}$ $\times 2.2$

$\rightarrow = \frac{1}{0.9} \times 2.2 \; Cov(H_m, W_{kg})$

Covariance: some properties

- $Cov(\widetilde{X}, \widetilde{Y}) = Cov(\widetilde{Y}, \widetilde{X})$
- If $\widetilde{X}$ and $\widetilde{Y}$ are independent, $Cov(\widetilde{X}, \widetilde{Y}) = 0$.
- $Cov(\widetilde{X} + c, \widetilde{Y} + d) = Cov(\widetilde{X}, \widetilde{Y})$
- $Cov(c\widetilde{X}, d\widetilde{Y}) = cdCov(\widetilde{X}, \widetilde{Y})$
- $Cov(\widetilde{X}, \widetilde{X}) = Var(\widetilde{X})$

$$E(XY) - E(X)E(Y)$$
$$= E(X^2) - E(X)E(X) = Var(X)$$

Correlation coefficient

- $Corr(\widetilde{X}, \widetilde{Y}) = \rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y} = \dfrac{Cov(\widetilde{X}, \widetilde{Y})}{\sqrt{Var(\widetilde{X}) Var(\widetilde{Y})}}$ ✓ $= \dfrac{Cov}{s.d.(X)\, s.d.(Y)}$

ρ ↓ rho

Correlation coefficient

- $Corr(\widetilde{X}, \widetilde{Y}) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{Cov(\widetilde{X}, \widetilde{Y})}{\sqrt{Var(\widetilde{X}) Var(\widetilde{Y})}}$
- $-1 \leq \rho_{XY} \leq 1$

Correlation coefficient

- $Corr(\widetilde{X}, \widetilde{Y}) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{Cov(\widetilde{X}, \widetilde{Y})}{\sqrt{Var(\widetilde{X}) Var(\widetilde{Y})}}$
- $-1 \leq \rho_{XY} \leq 1$
- If $\rho_{XY} = 0$: $\widetilde{X}$ and $\widetilde{Y}$ are uncorrelated

e.g. price of houses $= \tilde{X}$
price of fish $= \tilde{Y}$

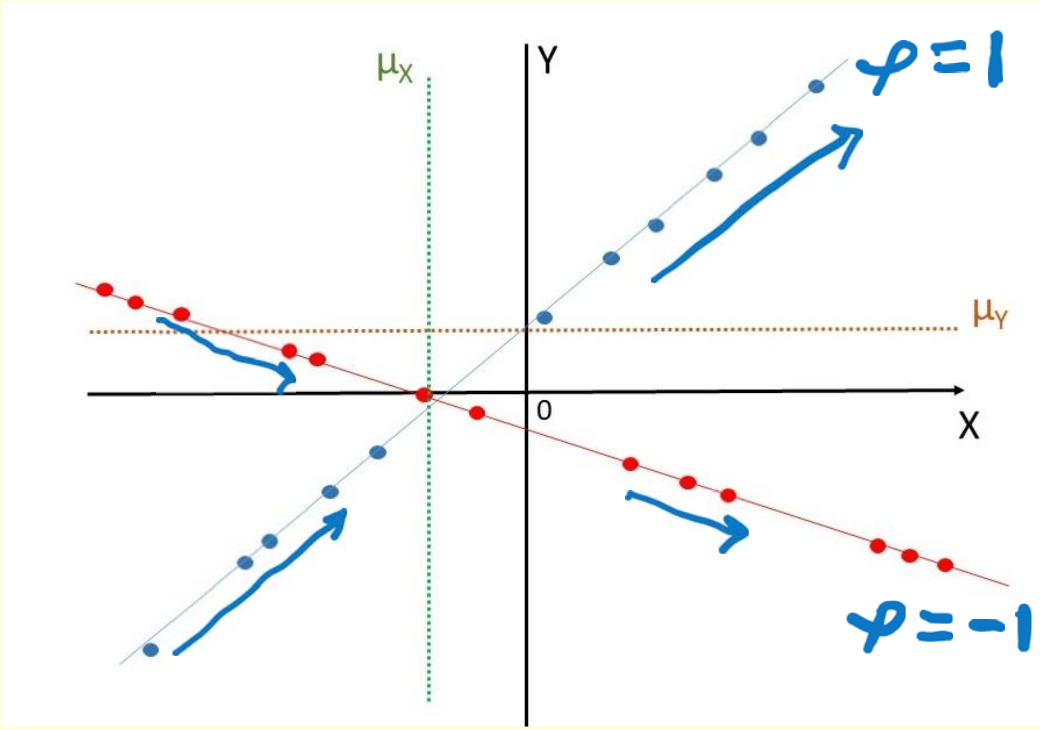
Correlation coefficient: Properties

- $-1 \leq \rho_{XY} \leq 1$

$\tilde{x}$ = price of houses
$\tilde{y}$ = rent

- If $\rho_{XY} > 0$: $\widetilde{X}$ and $\widetilde{Y}$ are positively correlated

  If $\rho_{XY} = 1$: $\widetilde{X}$ and $\widetilde{Y}$ are *perfectly, linearly, positively correlated*



$\tilde{x} = a + b\tilde{y}$
$b > 0$

$\tilde{x} = a - b\tilde{y}$

Correlation coefficient: Properties

- $-1 \leq \rho_{XY} \leq 1$
- If $\rho_{XY} > 0$: $\widetilde{X}$ and $\widetilde{Y}$ are positively correlated

  If $\rho_{XY} = 1$: $\widetilde{X}$ and $\widetilde{Y}$ are *perfectly, linearly, positively correlated*
- If $\rho_{XY} < 0$: $\widetilde{X}$ and $\widetilde{Y}$ are negatively correlated

  If $\rho_{XY} = -1$: $\widetilde{X}$ and $\widetilde{Y}$ are *perfectly, linearly, negatively correlated*

$\tilde{X}$ = growth-rate
$\tilde{Y}$ = birth-rate

Correlation coefficient: Properties

- $-1 \leq \rho_{XY} \leq 1$

- If $\rho_{XY} > 0$: $\widetilde{X}$ and $\widetilde{Y}$ are positively correlated

  If $\rho_{XY} = 1$: $\widetilde{X}$ and $\widetilde{Y}$ are *perfectly, linearly, positively correlated*

- If $\rho_{XY} < 0$: $\widetilde{X}$ and $\widetilde{Y}$ are negatively correlated

  If $\rho_{XY} = -1$: $\widetilde{X}$ and $\widetilde{Y}$ are *perfectly, linearly, negatively correlated*

- $Corr(\widetilde{X}, \widetilde{Y}) = Corr(a + c\widetilde{X}, b + d\widetilde{Y})$

Corr (H, W) → unit free

kg m

lb ft.

Correlation: an example

- Joint pmf of $\widetilde{X}$ and $\widetilde{Y}$:

			y		
	$f(x,y)$	**1**	**2**	**3**	$f_X(x)$
x	**10**	0	0.2	0.4	0.6
	20	0.2	0.2	0	0.4
	$f_Y(y)$	0.2	0.4	0.4	1

$$Var(\tilde{Y}) = E\tilde{Y}^2 - (E\tilde{Y})^2 = 5.4 - (2.2)^2 = 0.56$$

- $Corr(\widetilde{X},\widetilde{Y}) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{Cov(\widetilde{X},\widetilde{Y})}{\sqrt{Var(\widetilde{X})Var(\widetilde{Y})}} = \frac{-2.8}{\sqrt{24 \times 0.56}} = -0.76$

$Cov(\widetilde{X},\widetilde{Y}) = -2.8$, $Var(\widetilde{X}) =$, $Var(\widetilde{Y}) =$

$$Var(\tilde{X}) = E\tilde{X}^2 - (E\tilde{X})^2 = 220 - (14)^2 = 24$$

$$E\tilde{X}^2 = 10^2 \times 0.6 + 20^2 \times 0.4 = 220$$

Clicker question 3

- **Q**: Which one of the following correlation coefficients measures the strongest relationship between random variables $\widetilde{X}$ and $\widetilde{Y}$?

  (a) -0.8

  (b) -0.4

  (c) 0.3

  (d) 0.6

Solution to clicker question 3

- **Q**: Which one of the following correlation coefficients measures the strongest relationship between random variables $\widetilde{X}$ and $\widetilde{Y}$?

  (a) -0.8

  (b) -0.4

  (c) 0.3

  (d) 0.6

Random Variables: Continuous

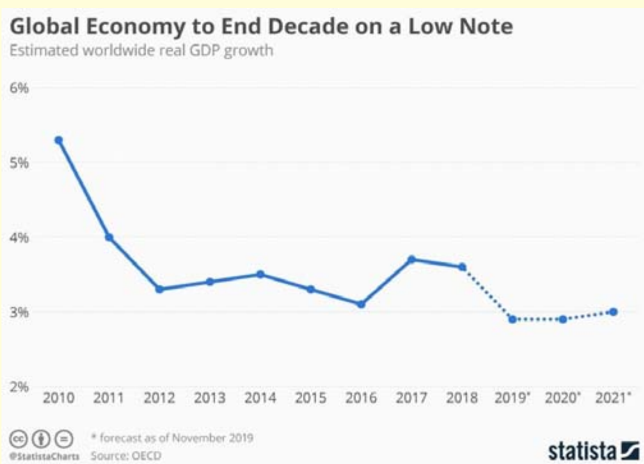
- *Random variable* $\widetilde{X}$: A variable taking on *numerical values* based on the outcome of some random event

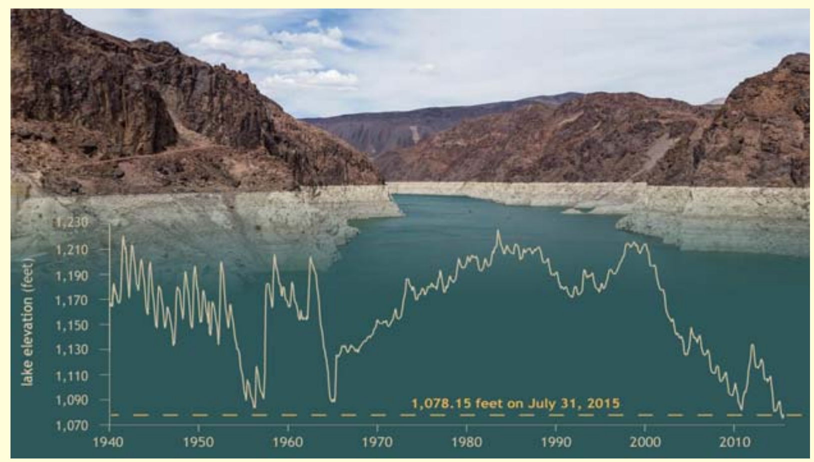
Random Variables: Continuous

- *Random variable* $\widetilde{X}$: A variable taking on *numerical values* based on the outcome of some random event
- It is *continuous* if $\widetilde{X}$ can take any value in some range.

Random Variables: Continuous

- *Random variable* $\widetilde{X}$: A variable taking on *numerical values* based on the outcome of some random event
- It is *continuous* if $\widetilde{X}$ can take any value in some range.
- Examples: growth rate, crime rate, profits/losses, water level in a lake, temperature,.....





Random Variables: Continuous

- *Random variable* $\widetilde{X}$: A variable taking on *numerical values* based on the outcome of some random event
- It is *continuous* if $\widetilde{X}$ can take any value in some range.
- Examples: growth rate, crime rate, profits/losses, water level in a lake, temperature,.....
- In practice, measurement limitations sometimes limit us to discrete values (e.g. temperature)

Continuous Random Variables: probabilities

- M = maximum possible value of $\widetilde{X}$

  m = minimum possible value of $\widetilde{X}$

  $Range = M - m$

  $70 - (-100) = 170$

Continuous Random Variables: probabilities

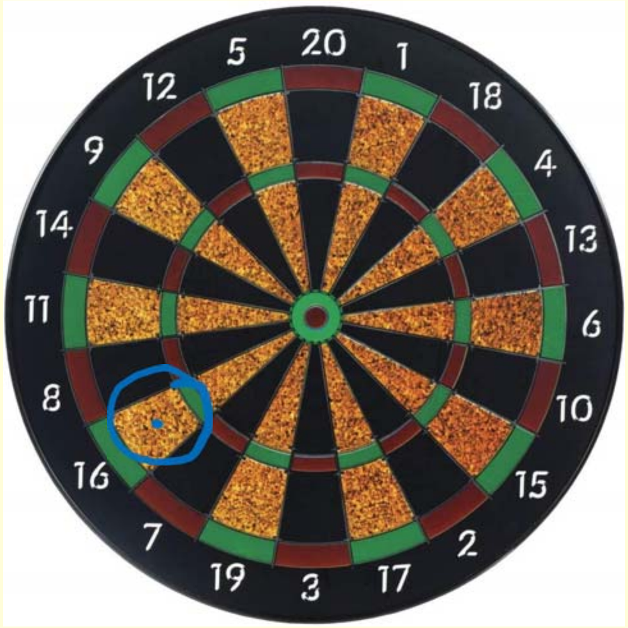
- $M =$ maximum possible value of $\widetilde{X}$

  $m =$ minimum possible value of $\widetilde{X}$

  $Range = M - m$

- Prob $[\widetilde{X} = x] =? 0$

$t \leq 3^\circ C$



Continuous Random Variables: probabilities

- $M =$ maximum possible value of $\widetilde{X}$

  $m =$ minimum possible value of $\widetilde{X}$

  $Range = M - m$

- Prob $[\widetilde{X} = x] =?$

- **Prob** $[\widetilde{X} = x] = 0.$

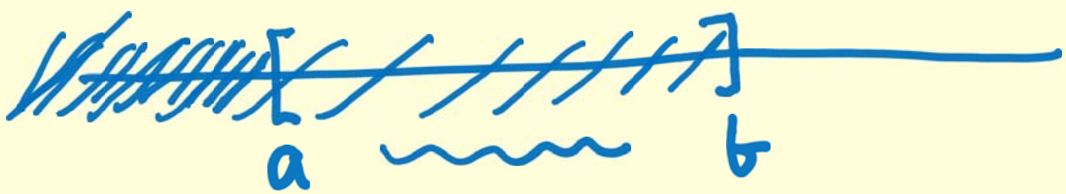
  Look at: **Prob** $[\widetilde{X} \leq x]$

Continuous Random Variables: c.d.f.

- *Recall*: Prob $[\widetilde{X} \leq x] = F(x)$: c.d.f. of $\widetilde{X}$

Continuous Random Variables: c.d.f.

- *Recall*: Prob $[\widetilde{X} \leq x] = F(x)$: c.d.f. of $\widetilde{X}$
- **Prob** $\{\widetilde{X} \in [a, b]\} = ?$



Continuous Random Variables: c.d.f.

- *Recall*: Prob $[\widetilde{X} \leq x] = F(x)$: c.d.f. of $\widetilde{X}$
- **Prob** $\{\widetilde{X} \in [a, b]\} =?$
- **Prob** $\{\widetilde{X} \in [a, b]\} = P[\widetilde{X} \leq b] - P[\widetilde{X} \leq a] = F(b) - F(a)$

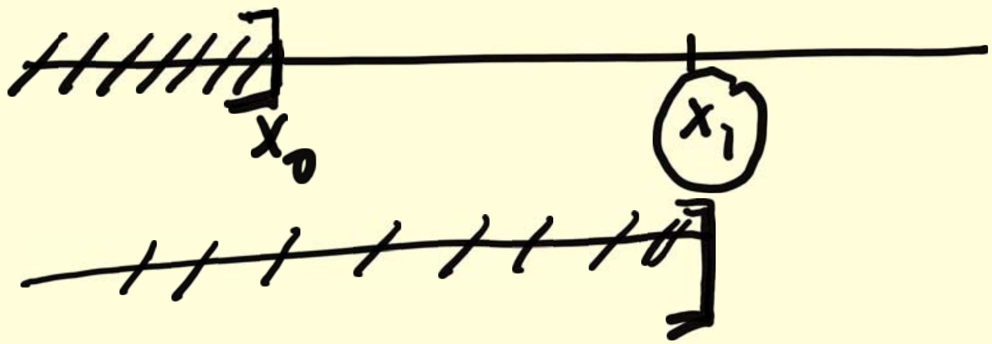
[ ]

Continuous Random Variables: c.d.f.

- *Recall*: Prob $[\widetilde{X} \leq x] = F(x)$: c.d.f. of $\widetilde{X}$
- **Prob** $\{\widetilde{X} \in [a, b]\} = ?$
- **Prob** $\{\widetilde{X} \in [a, b]\} = P[\widetilde{X} \leq b] - P[\widetilde{X} \leq a] = F(b) - F(a)$
- How to get an idea of which points are more/less likely?

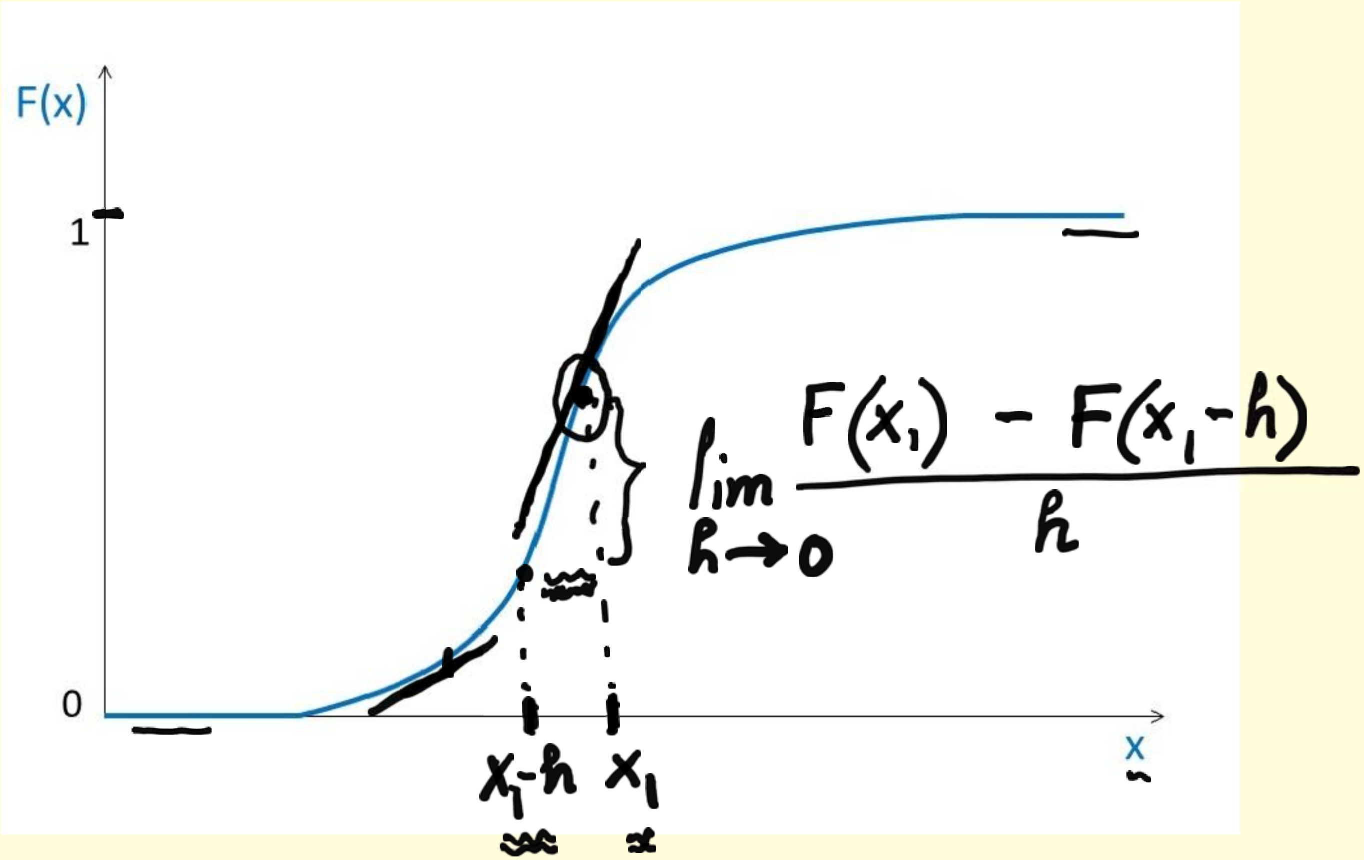
Continuous r.v.s: prob. density function (pdf)

- *Recall in discrete case*: $F(x_1) - F(x_0) = p(x_1)$



Continuous r.v.s: prob. density function

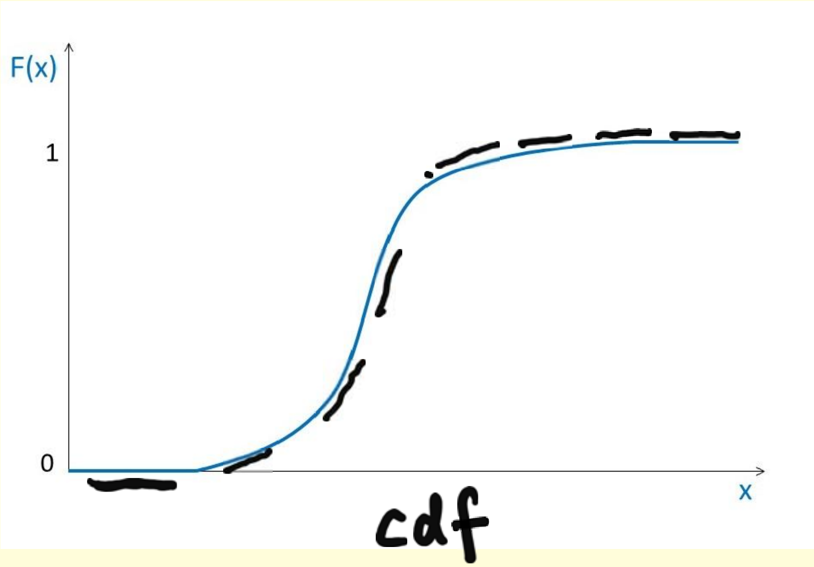
- *Recall in discrete case*: $F(x_1) - F(x_0) = p(x_1)$
- *Continuous r.v.*: Look at rate of change in cdf

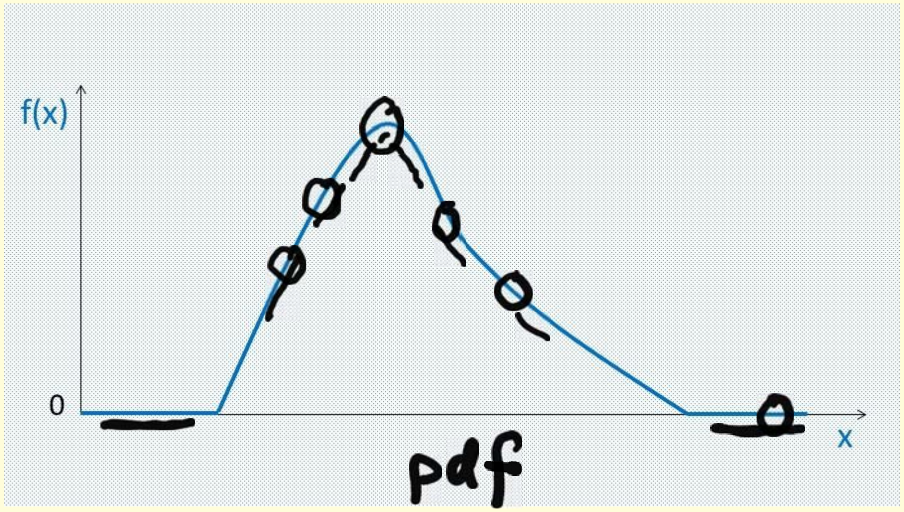


Continuous r.v.s: prob. density function

- *Recall in discrete case*: $F(x_1) - F(x_0) = p(x_1)$
- *Continuous r.v.*: Look at rate of change in cdf
- **Prob. density function (p.d.f.)** $f(x)$**:**

$$f(x) = \lim_{h \to 0} \frac{F(x) - F(x-h)}{h} = \frac{d}{dx} F(x) = F'(x) = f(x)$$





Continuous r.v.s: prob. density function

- *Recall in discrete case*: $F(x_1) - F(x_0) = p(x_1)$
- *Continuous r.v.*: Look at rate of change in cdf
- **Prob. density function (p.d.f.)** $f(x)$**:**

$$f(x) = \lim_{h \to 0} \frac{F(x) - F(x-h)}{h} = \frac{d}{dx} F(x) = F'(x)$$

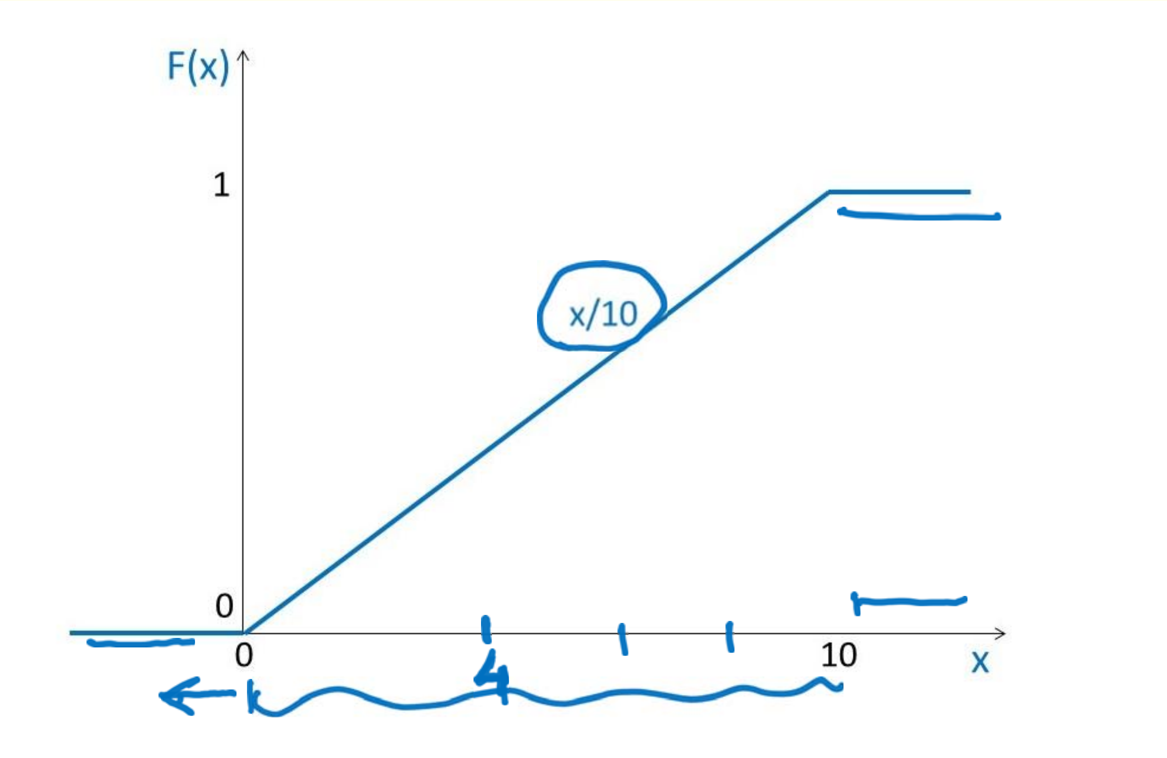
- Interpretation of pdf $f(x)$:

  $f(x)$ *high*: relative likelihood of occurrence of x is high

  $f(x)$ *low*: relative likelihood of occurrence is low

Cdf and pdf: an example

- *Uniform distribution over* $[0, 10]$: $F(x) = 0$ **for** $x < 0$, $F(x) = \frac{x}{10}$ **for** $x \in [0, 10]$, $F(x) = 1$ **for** $x > 10$



$F(4) = \frac{4}{10}$
$= 0.4$

Cdf and pdf: an example

- *Uniform distribution over* $[0, 10]$: $F(x) = 0$ **for** $x < 0$, $F(x) = \frac{x}{10}$ **for** $x \in [0, 10]$, $F(x) = 1$ **for** $x > 10$
- Relative likelihoods of different values? pdf

Cdf and pdf: an example

- *Uniform distribution over* $[0, 10]$: $F(x) = 0$ **for** $x < 0$,

  $F(x) = \frac{x}{10}$ **for** $x \in [0, 10]$, $F(x) = 1$ **for** $x > 10$

- Relative likelihoods of different values?

- *Prob. density function* $f(x) = \frac{d}{dx} F(x) =$ $f(x) = \begin{cases} 0 \text{ for } x < 0 \\ \frac{1}{10} \text{ for } 0 \leq x \leq 10 \\ 0 \text{ for } x > 10 \end{cases}$

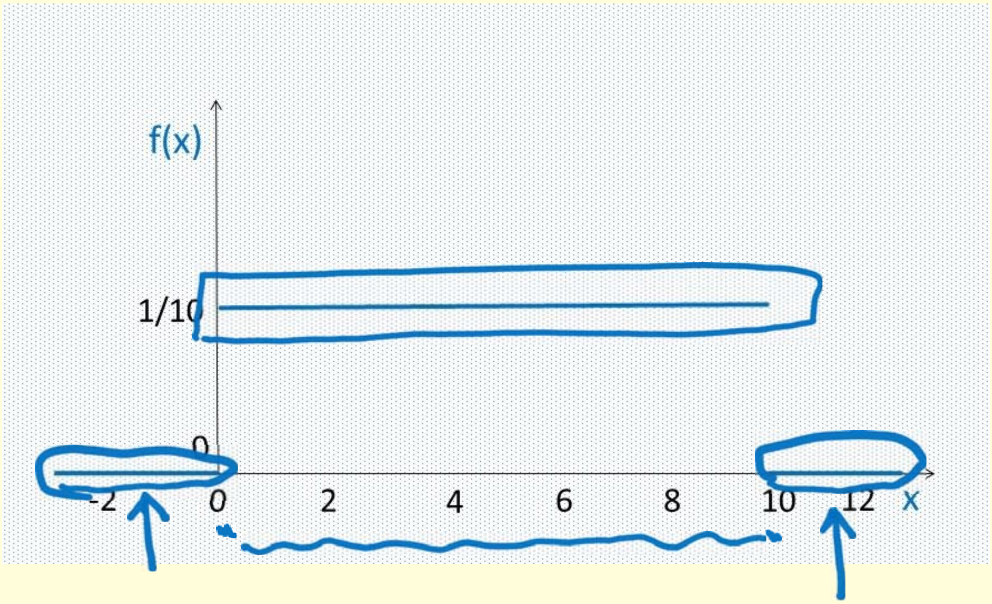
$\frac{d}{dx} 0 = 0$

$\frac{d}{dx} 1 = 0$

$$f(x) = \frac{d}{dx}\left(\frac{x}{10}\right) = \frac{1}{10} \underbrace{\frac{d}{dx}(x)}_{=1} = \frac{1}{10}$$

Cdf and pdf: an example

- *Uniform distribution over* $[0, 10]$: $F(x) = 0$ **for** $x < 0$, $F(x) = \frac{x}{10}$ **for** $x \in [0, 10]$, $F(x) = 1$ **for** $x > 10$
- Relative likelihoods of different values?
- *Prob. density function* $f(x) = \frac{d}{dx} F(x) =$



Clicker question 4

- **Q**: What does c.d.f. and p.d.f. stand for? Pick the option that has both terms correct.

  (a) concentrated density function, probability density function

  (b) cumulative density function, probability density function

  (c) cumulative density function, probability distribution function

  (d) cumulative distribution function, probability density function

Solution to clicker question 4

- **Q**: What does c.d.f. and p.d.f. stand for? Pick the option that has both terms correct.

  (a) concentrated density function, probability density function

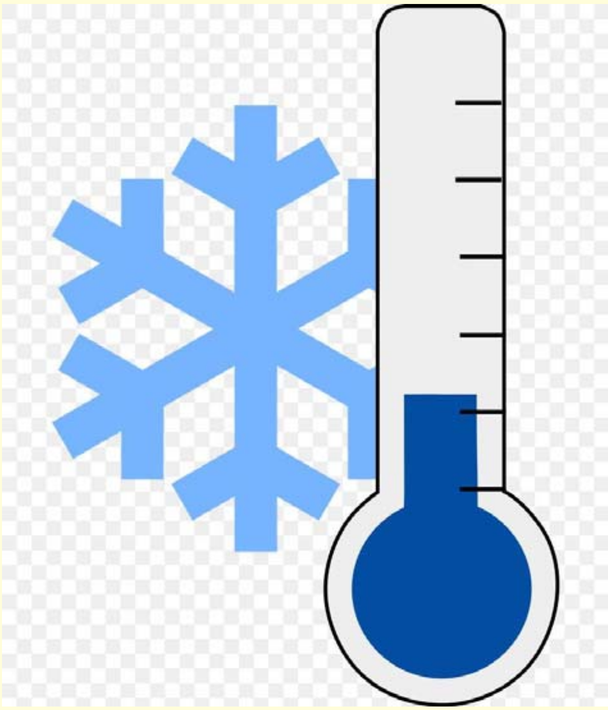
  (b) cumulative density function, probability density function

  (c) cumulative density function, probability distribution function

  (d) cumulative distribution function, probability density function

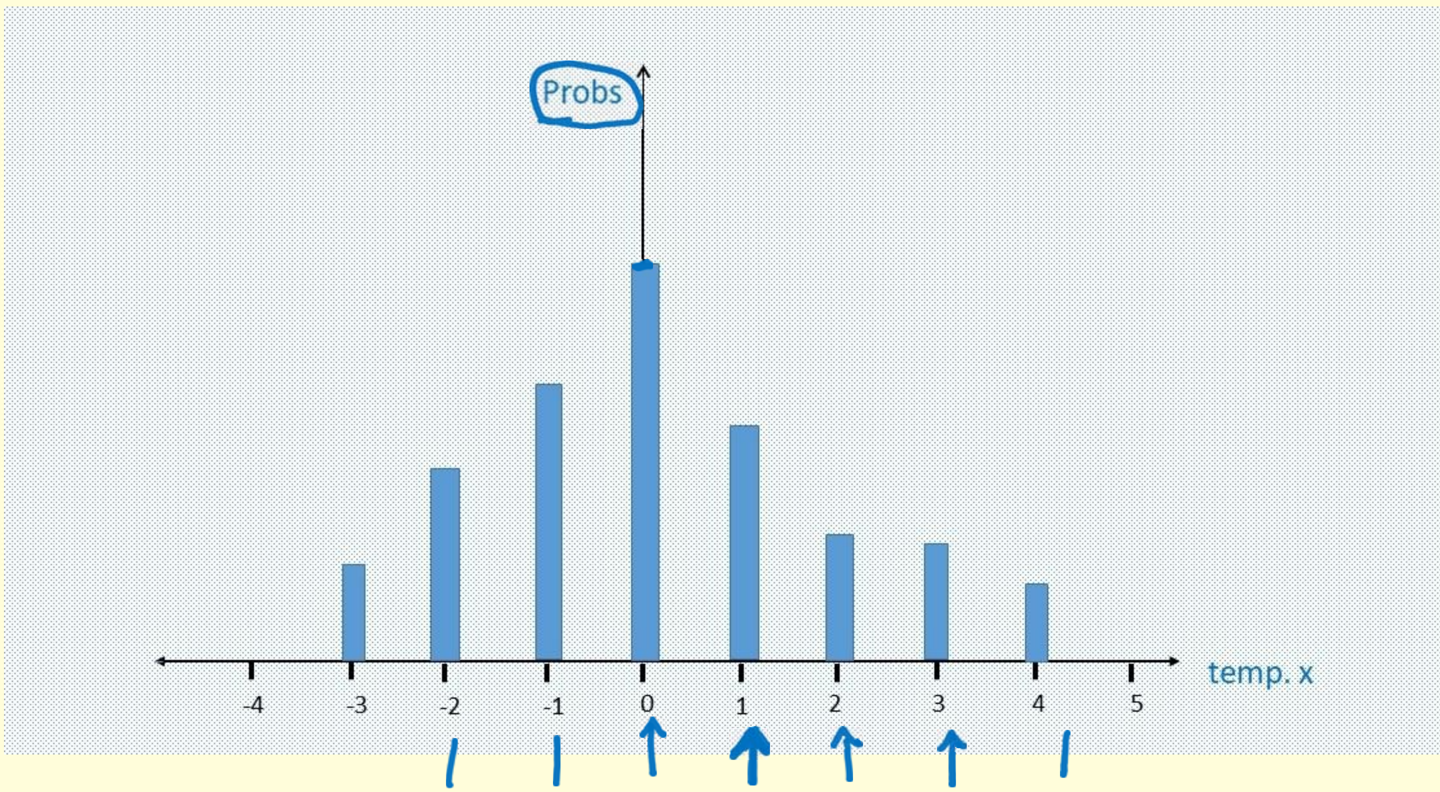
Prob. density fn.: An alternative view

- $\widetilde{X}$ = temp. in a city



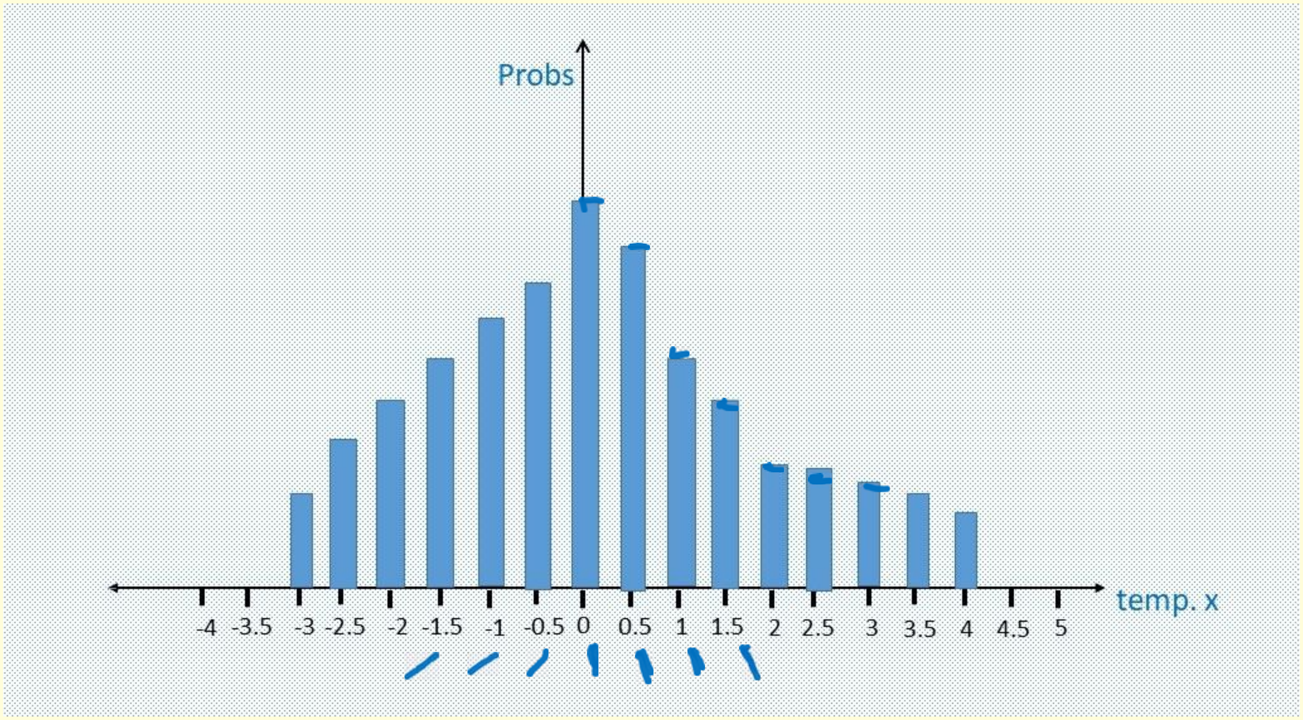
Prob. density fn.: An alternative view

- $\widetilde{X}$ = temp. in a city
- Measure $\widetilde{X}$ in whole degrees: probs. $= p(x)$



Prob. density fn.: An alternative view

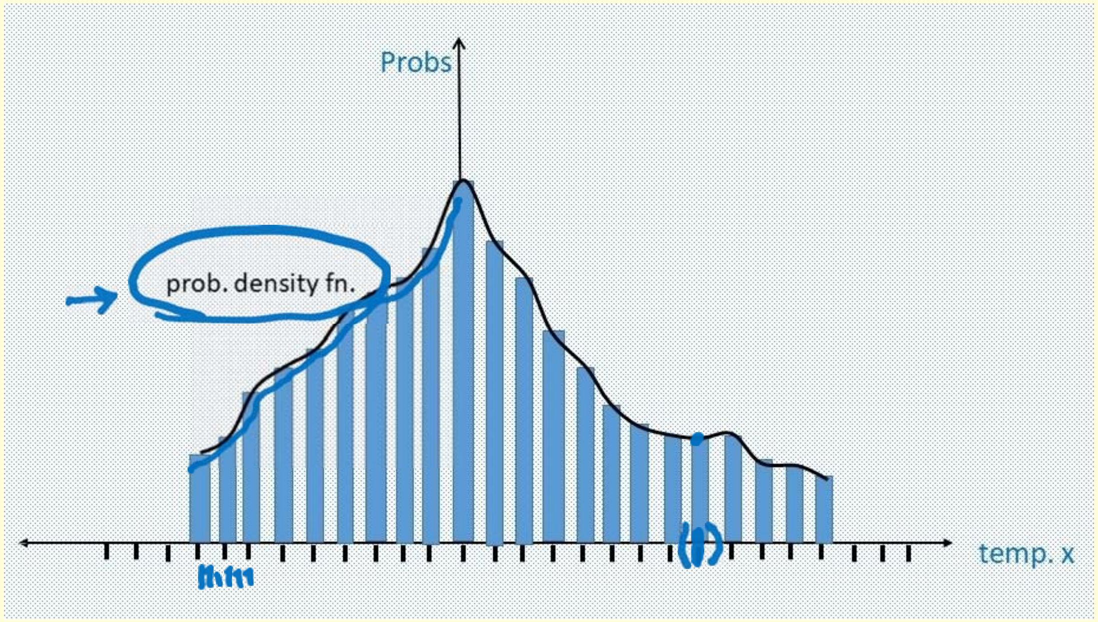
- $\widetilde{X}$ = temp. in a city
- Measure $\widetilde{X}$ in whole degrees: probs. $= p(x)$
- Measure $\widetilde{X}$ in 0.5 degrees: probs. $= p(x)$



Prob. density fn.: An alternative view

- $\widetilde{X}$ = temp. in a city
- Measure $\widetilde{X}$ in whole degrees: probs. $= p(x)$
- Measure $\widetilde{X}$ in 0.5 degrees: probs. $= p(x)$
- Measure in finer and finer increments:

  arrive at p.d.f. $f(x)$



C.d.f. and p.d.f.: Connections

- *c.d.f.*: Prob $[\widetilde{X} \leq x] = F(x)$

  p.d.f.: $f(x) = \frac{d}{dx}F(x) = F'(x)$

Clicker question 5

- For a random variable $\widetilde{X}$, the cdf is given by $F(x) = x^2$ for $x \in [0, 1]$, $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$.

  Which of the following is its pdf for values of $x \in [0, 1]$?

  (a) 0

  (b) $2x$

  (c) x^2

  (d) $2x^2$

Solution to clicker question 5

- For a random variable $\widetilde{X}$, the cdf is given by $F(x) = x^2$ for $x \in [0, 1]$, $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$.

  Which of the following is its pdf for values of $x \in [0, 1]$?

  (a) 0

  (b) $2x$

  (c) x^2

  (d) $2x^2$

$F(x) = x^2$

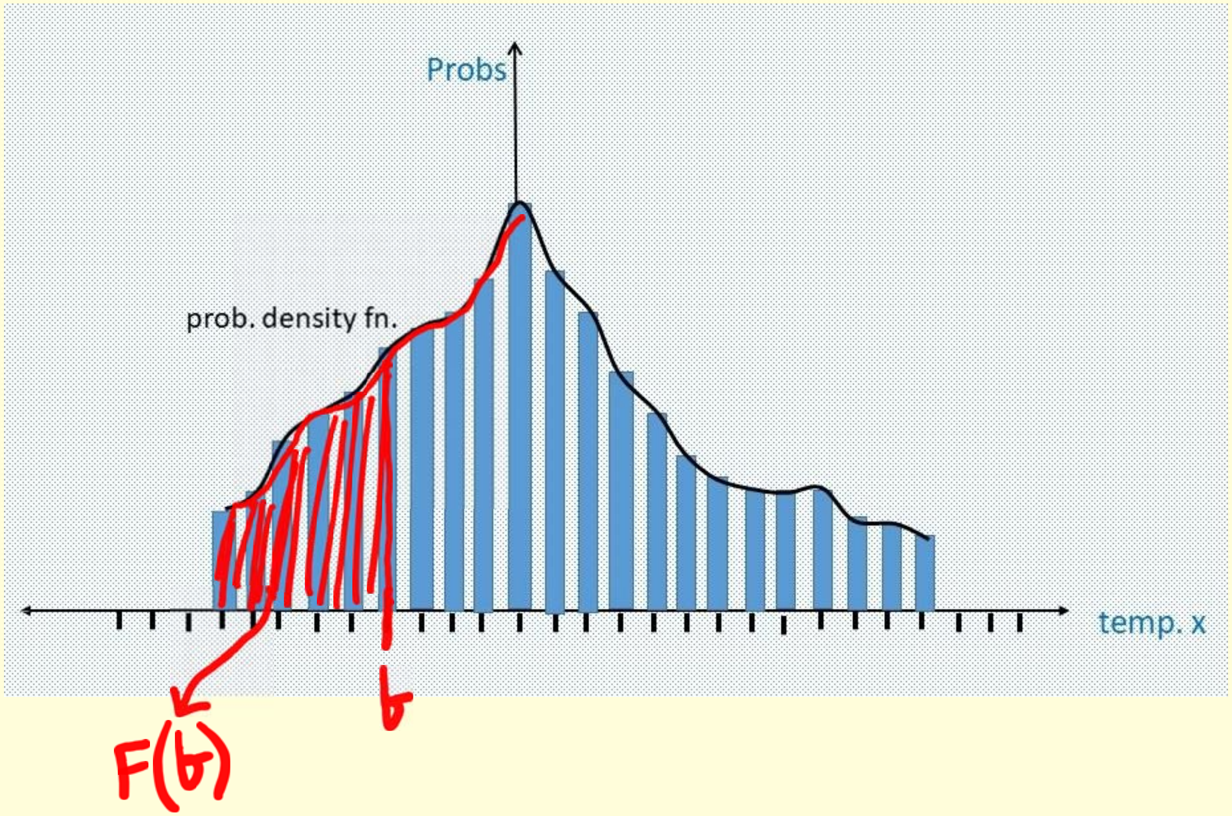
$f(x) = F'(x) = 2x$

$\frac{d}{dx} x^n = n x^{n-1}$

C.d.f. and p.d.f.: Connections

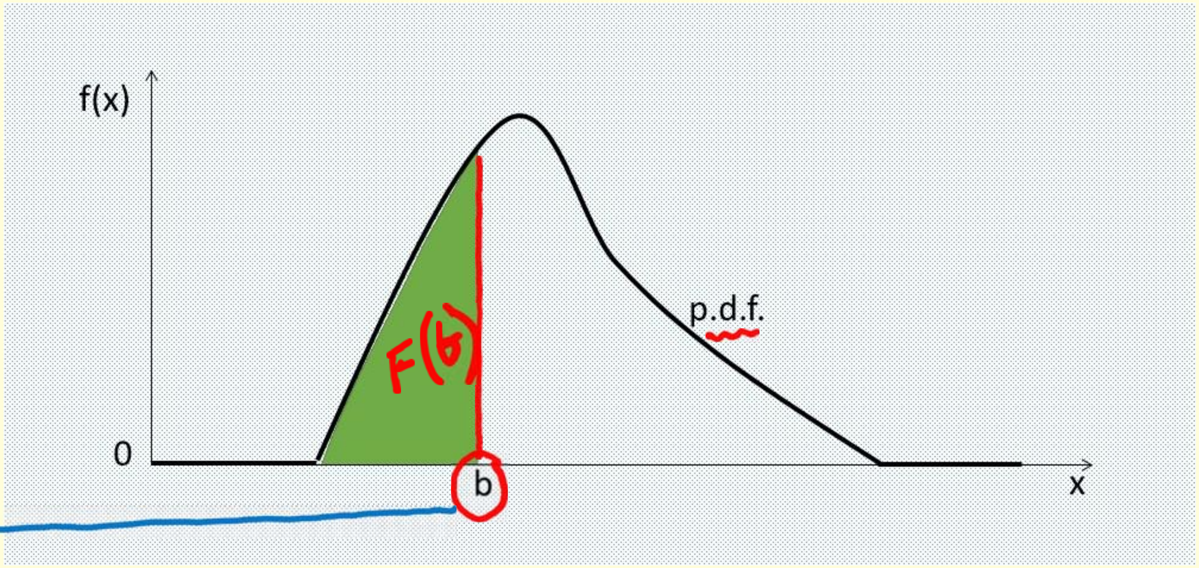
- *c.d.f.*: Prob $[\widetilde{X} \leq x] = F(x)$ *p.d.f.*: $f(x) = \frac{d}{dx}F(x) = F'(x)$

- *Opposite*:



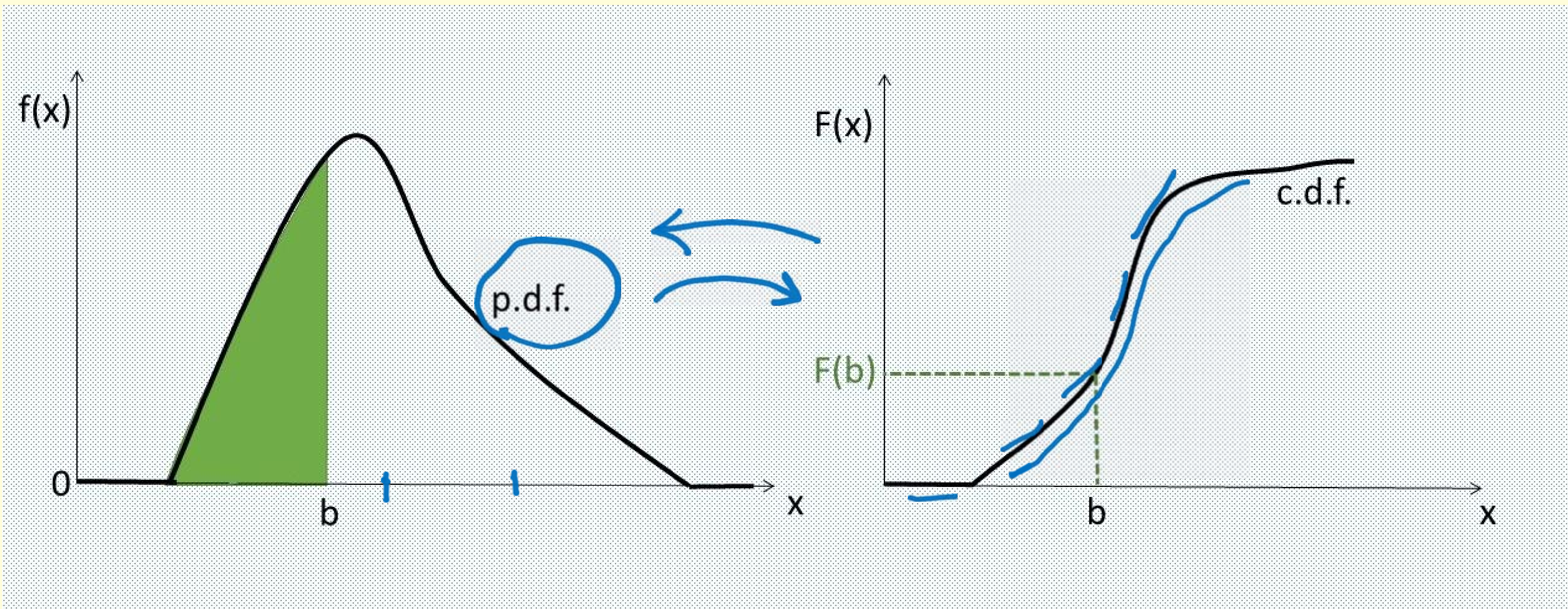
C.d.f. and p.d.f.: Connections

- *c.d.f.*: Prob $[\widetilde{X} \leq x] = F(x)$ *p.d.f.*: $f(x) = \frac{d}{dx}F(x) = F'(x)$
- *Opposite*:
- $F(b)$ = Area under p.d.f. upto $b = \int_{-\infty}^{b} f(x)dx$



C.d.f. and p.d.f.: Connections

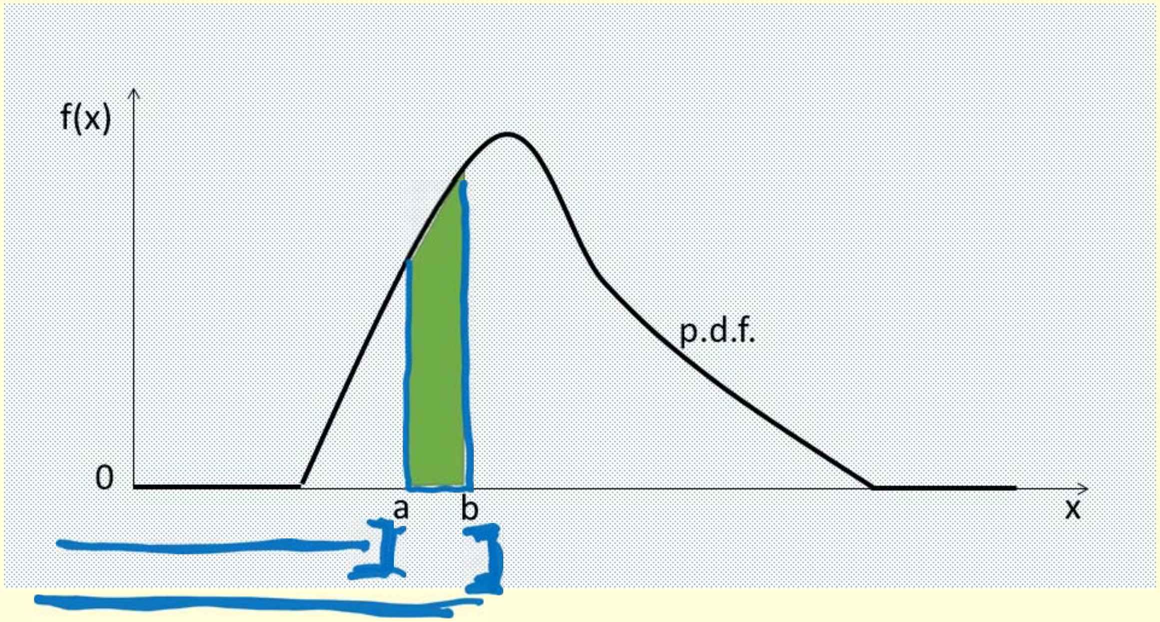
- $F(b) = \int_{-\infty}^{b} f(x)dx$



C.d.f. and p.d.f.: Connections

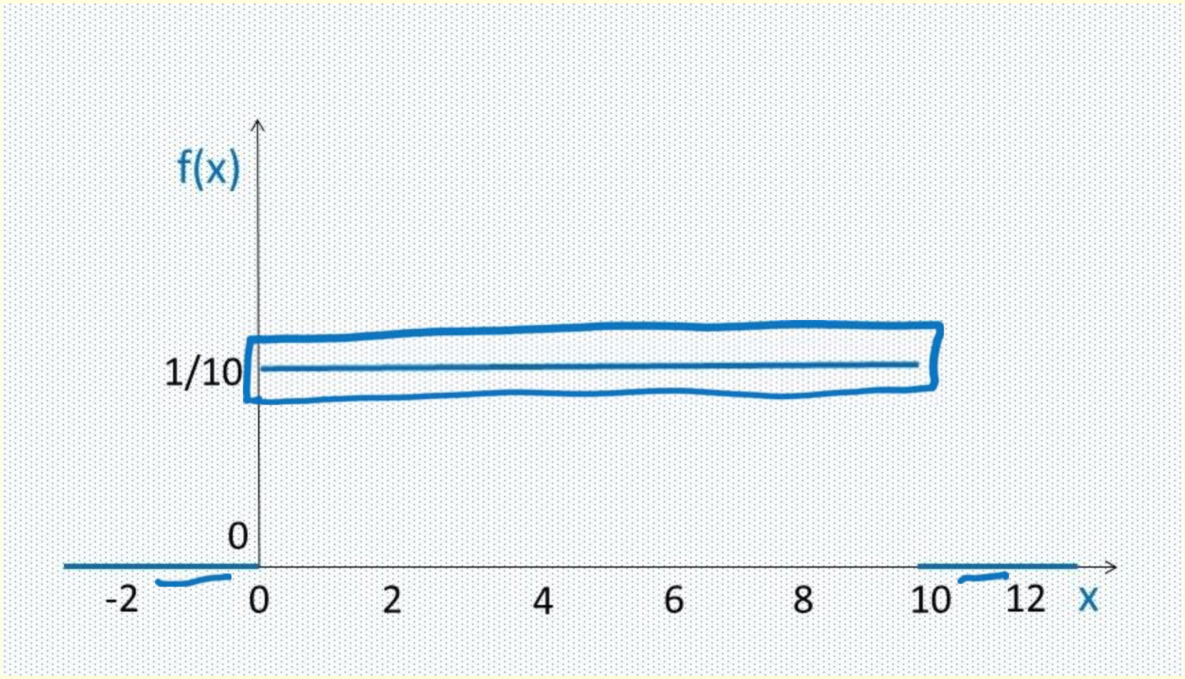
- $F(b) = \int_{-\infty}^{b} f(x)dx$
- **Prob** $\{\widetilde{X} \in [a,b]\} = Prob[\widetilde{X} \leq b] - Prob[\widetilde{X} \leq a] =$

$$F(b) - F(a) = \int_a^b f(x)dx$$



Cdf and pdf: an example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$, $f(x) = 0$ elsewhere



Cdf and pdf: an example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,

  $f(x) = 0$ elsewhere

- cdf $F(x)$?

Cdf and pdf: an example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$, $f(x) = 0$ elsewhere
- cdf $F(x)$?
- $F(b) =$ Area under p.d.f. upto $b = \int_{-\infty}^{b} f(x)dx$

Cdf and pdf: an example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,

  $f(x) = 0$ elsewhere

- cdf $F(x)$?

- $F(b) =$ Area under p.d.f. upto $b = \int_{-\infty}^{b} f(x)dx$

- Here:

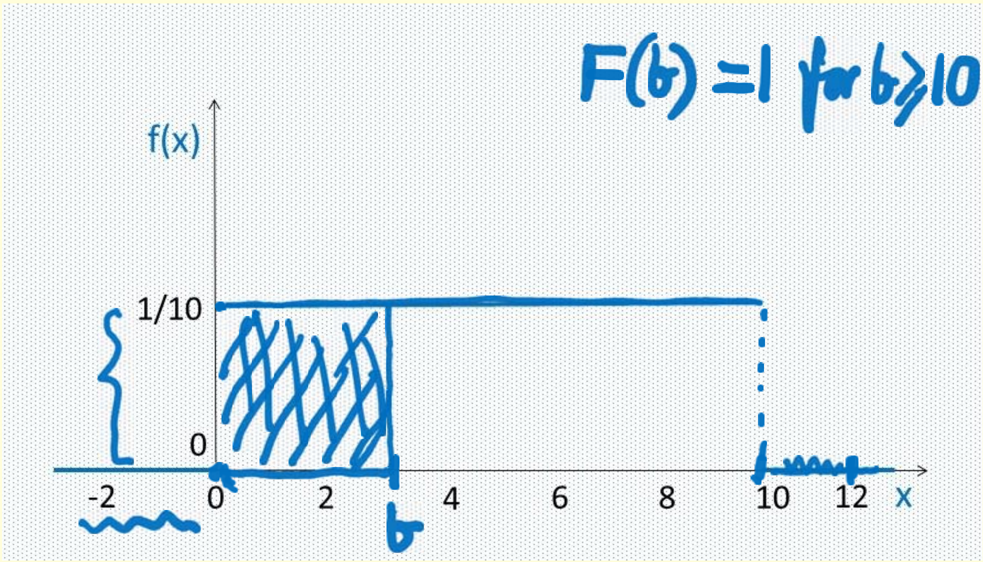
$F(12) = \frac{1}{10}\,10$

$= 1$

$F(b) = b \times \frac{1}{10} = \frac{b}{10}$

for $b \in [0, 10]$

$F(b) = 0$ for $b < 0$



Cdf and pdf: an example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$, $f(x) = 0$ elsewhere

- cdf $F(b) = \int_{-\infty}^{b} f(x)dx =$

$$\int_0^b \frac{1}{10}\,dx = \frac{1}{10}\int_0^b 1\,dx$$

$$= \frac{1}{10}\, x \Big|_0^b$$

$$= \frac{1}{10}\,[b - 0] = \frac{b}{10}$$

Cdf and pdf: an example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,

  $f(x) = 0$ elsewhere

  $\tilde{X}$ = returns on a stock $\sim U[0,10]$

- cdf $F(b) = \int_{-\infty}^{b} f(x)dx = \frac{b}{10}$

- Prob $[\tilde{X} \leq 5] = F(5) = \frac{5}{10} = \frac{1}{2}$

Cdf and pdf: an example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$, $f(x) = 0$ elsewhere

- cdf $F(b) = \int_{-\infty}^{b} f(x) dx =$

- Prob $[\widetilde{X} \leq 5] = F(5) =$

- Prob $\{\widetilde{X} \in [3, 5]\} = F(5) - F(3) = \frac{5}{10} - \frac{3}{10} = \frac{2}{10} = 0.2$

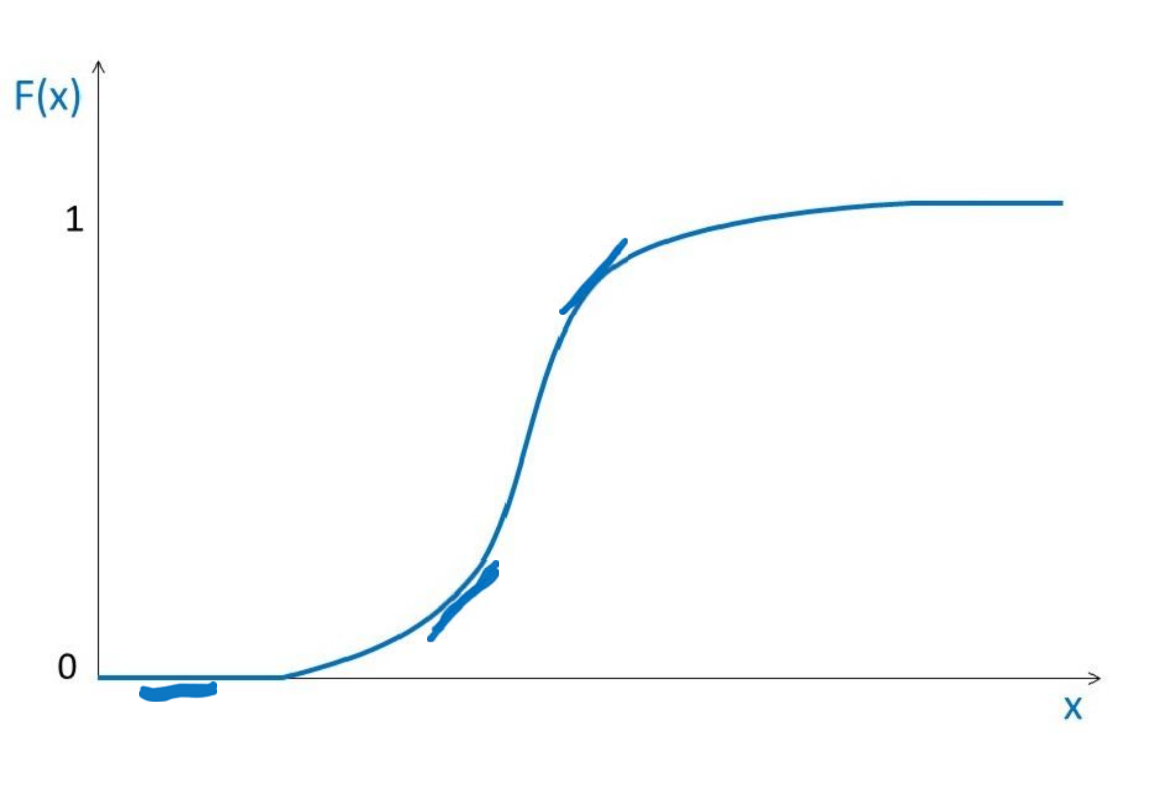
$F(b) = \frac{b}{10}$

P.d.f.: properties

- *p.d.f.*: $f(x) = F'(x)$ *c.d.f.*: $F(b) = \int_{-\infty}^{b} f(x)dx$

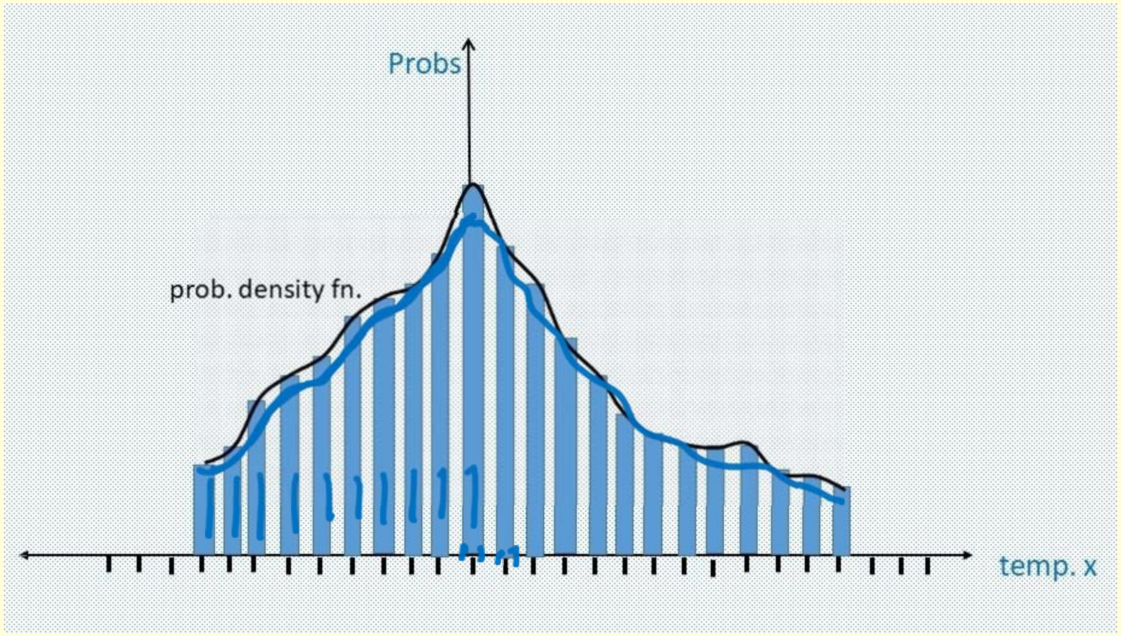
P.d.f.: properties

- p.d.f.: $f(x) = F'(x)$ c.d.f.: $F(b) = \int_{-\infty}^{b} f(x)dx$
- $f(x) \geq 0$



P.d.f.: properties

- p.d.f.: $f(x) = F'(x)$ c.d.f.: $F(b) = \int_{-\infty}^{b} f(x)dx$
- $f(x) \geq 0$
- Total area under pdf $= \int_{-\infty}^{\infty} f(x)dx = 1$



P.d.f.: properties

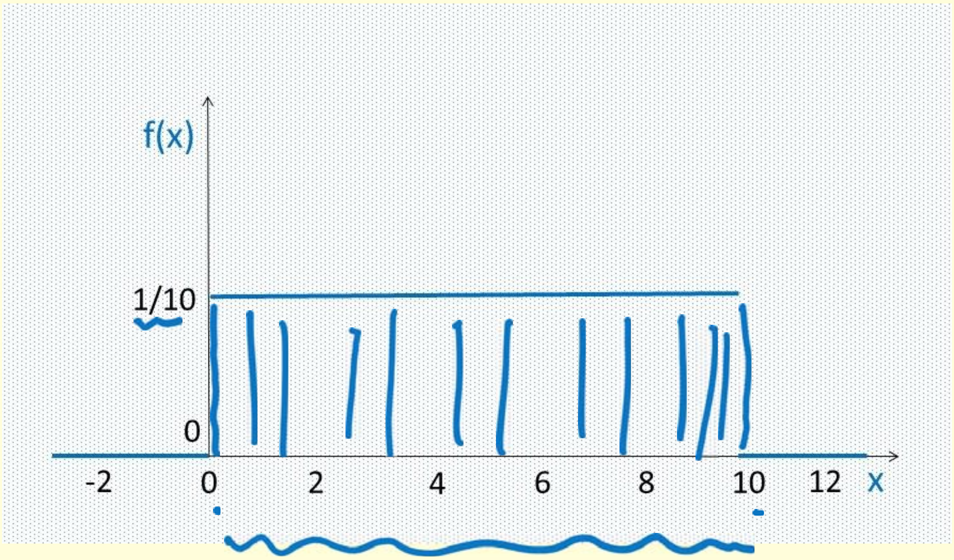
- <u>*p.d.f.*</u>: $f(x) = F'(x)$ <u>*c.d.f.*</u>: $F(b) = \int_{-\infty}^{b} f(x)dx$
- $f(x) \geq 0$
- Total area under pdf $= \int_{-\infty}^{\infty} f(x)dx = 1$
- <u>Example</u>: *Uniform distn pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,

  $f(x) = 0$ elsewhere

P.d.f.: properties

- <u>*p.d.f.*</u>: $f(x) = F'(x)$ <u>*c.d.f.*</u>: $F(b) = \int_{-\infty}^{b} f(x)dx$
- $f(x) \geq 0$
- Total area under pdf $= \int_{-\infty}^{\infty} f(x)dx = 1$
- <u>Example</u>: *Uniform distn pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$, $f(x) = 0$ elsewhere
- Total area $=$

$\frac{1}{10} \times 10 = 1$



Expectation and Variance

- **Expected value** (*mean*) of $\widetilde{X}$:

$\int \sum_x x p(x)$ $f(x)$

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

Expectation and Variance

- **Expected value** (*mean*) of $\widetilde{X}$:

$$E\widetilde{X} = \mu_X = \int_{-\infty}^{\infty} xf(x)dx$$

- $Eh(\widetilde{X}) = \int_{-\infty}^{\infty} h(x)f(x)dx$

EX^2

$E\sqrt{X}$

Expectation and Variance

- **Expected value** (*mean*) of $\widetilde{X}$:

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

- $Eh(\widetilde{X}) = \int_{-\infty}^{\infty} h(x) f(x)\,dx$

- **Variance** $\sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\,dx$

Expectation and Variance

- **Expected value** (*mean*) of $\widetilde{X}$:

$$E\widetilde{X} = \mu_X = \int_{-\infty}^{\infty} xf(x)dx$$

- $Eh(\widetilde{X}) = \int_{-\infty}^{\infty} h(x)f(x)dx$

- **Variance** $\sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx$

- <u>Easier formula</u>: *Variance* $\sigma_X^2 = E(X^2) - (EX)^2$

sigma^2

Expectation and Variance: An example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$, $f(x) = 0$ elsewhere
- $E\widetilde{X} = \mu_X = \int_{-\infty}^{\infty} xf(x)dx$

$$= \int_0^{10} x \cdot \frac{1}{10} dx = \frac{1}{10} \int_0^{10} x\, dx$$

$$\frac{x^2}{2}$$

$$= \frac{1}{10} \left(\frac{x^2}{2} \right) \Bigg|_0^{10} = \frac{1}{10} \left(\frac{100}{2} - 0 \right)$$

$$= \frac{1}{10} \times 50 = 5$$

Expectation and Variance: An example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,

  $f(x) = 0$ elsewhere

- $E\widetilde{X} = \mu_X = \int_{-\infty}^{\infty} xf(x)dx$

- $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$ $= \int_0^{10} x^2 \cdot \frac{1}{10} dx = \frac{1}{10} \int_0^{10} x^2 dx$

$= \frac{1}{10} \cdot \left. \frac{x^3}{3} \right|_0^{10} = \frac{1}{10} \left[\frac{1000}{3} - 0 \right] = \frac{100}{3}$

$\boxed{\int x^n dx = \frac{x^{n+1}}{n+1}}$

Expectation and Variance: An example

- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$, $f(x) = 0$ elsewhere
- $E\widetilde{X} = \mu_X = \int_{-\infty}^{\infty} xf(x)dx$ $= 5$
- $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$ $= \frac{100}{3}$
- Variance of $\widetilde{X} = \sigma_X^2 = E(X^2) - (EX)^2$ $= \frac{100}{3} - 25$
  $= 8.33$

Common Continuous Distributions 1: Uniform Distn.

- $\widetilde{X}$ distributed over an interval $[A, B]$

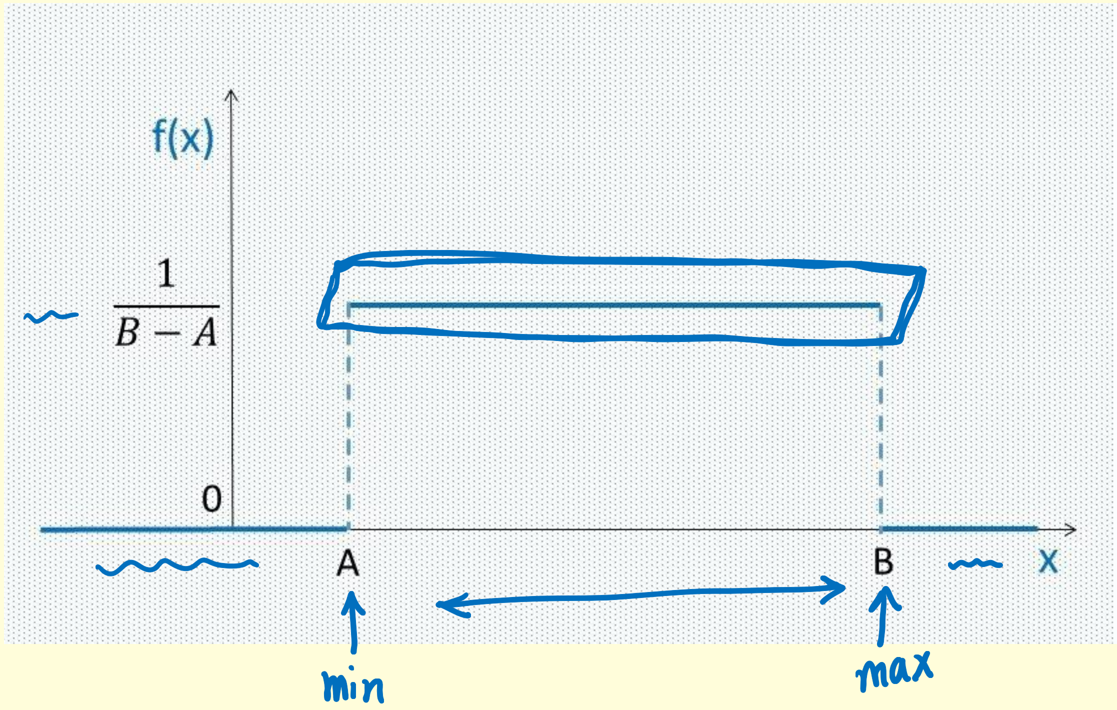
Range = B − A

Common Continuous Distributions 1: Uniform Distn.

- $\widetilde{X}$ distributed over an interval $[A, B]$
- Range $= B - A$

Common Continuous Distributions 1: Uniform Distn.

- $\widetilde{X}$ distributed over an interval $[A, B]$
- Range $= B - A$
- *p.d.f*: $f(x) = \frac{1}{B-A}$ **for** $x \in [A, B]$,
  $f(x) = 0$ elsewhere



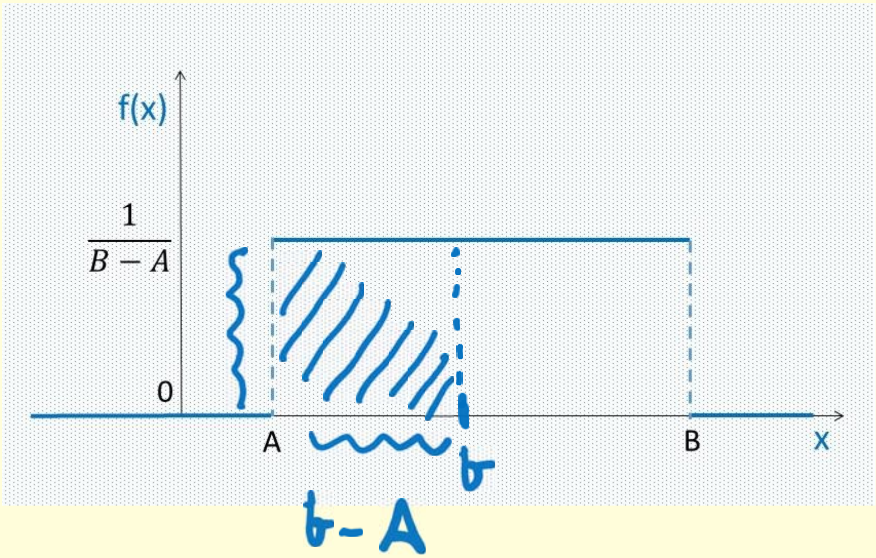
Common Continuous Distributions 1: Uniform Distn.

- $\widetilde{X}$ distributed over an interval $[A, B]$
- Range $= B - A$
- *p.d.f*: $f(x) = \frac{1}{B-A}$ **for** $x \in [A, B]$,

  $f(x) = 0$ elsewhere

- cdf $F(b) = \int_{-\infty}^{b} f(x)dx = \frac{b-A}{B-A}$

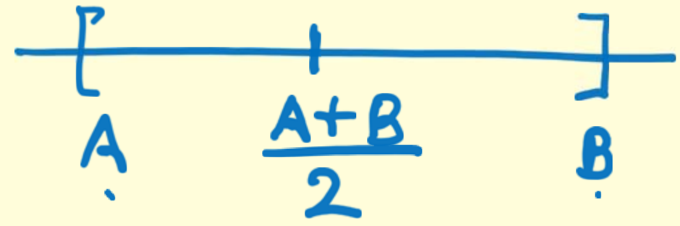
$$= (b-A) \cdot \frac{1}{(B-A)}$$



Common Continuous Distributions 1: Uniform Distn.

- $\widetilde{X}$ distributed over an interval $[A, B]$
- Range $= B - A$
- $p.d.f$: $f(x) = \frac{1}{B-A}$ **for** $x \in [A, B]$,

  $f(x) = 0$ elsewhere
- cdf $F(b) = \int_{-\infty}^{b} f(x)dx = \frac{b-A}{B-A}$
- Mean $= \frac{A+B}{2}$, Variance $= \frac{(B-A)^2}{12}$



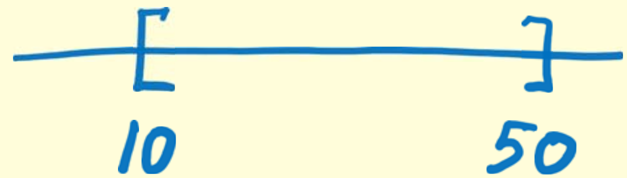
Uniform Distribution: Example

- Examples

Uniform Distribution: Example

- Examples
- A new stock is launched. $\widetilde{X}$ = its price at the end of the day.

  Uniformly distributed over $[10, 50]$



Uniform Distribution: Example

- Examples
- A new stock is launched. $\widetilde{X}$ = its price at the end of the day. Uniformly distributed over $[10, 50]$

- Mean price = $\frac{50+10}{2} = 30$, Variance $= \frac{(50-10)^2}{12}$

$$= \frac{40^2}{12}$$

$$= \frac{1600}{12} = \cdots$$

Uniform Distribution: Example

- Examples
- A new stock is launched. $\widetilde{X}$ = its price at the end of the day.
  Uniformly distributed over $[10, 50]$
- Mean price =
- $Prob[\widetilde{X} \leq 20] = F(20) = \frac{20-10}{50-10} = \frac{1\cancel{0}}{4\cancel{0}} = 0.25$

$$F(b) = \frac{b-A}{B-A}$$

Uniform Distribution: Example

- Examples
- A new stock is launched. $\widetilde{X}$ = its price at the end of the day. Uniformly distributed over $[10, 50]$
- Mean price =
- $Prob[\widetilde{X} \leq 20] = F(20) =$
- $Prob[\widetilde{X} > 40] =$ $1 - F(40) = 1 - \frac{40-10}{50-10}$

$$= 1 - \frac{3\cancel{0}}{4\cancel{0}} = \frac{1}{4} = 0.25$$

Clicker question 6

- The water level in a lake, $\widetilde{X}$, is uniformly distributed over $[500, 600]$. What is the mean water level?

(a) 100

(b) 500

(c) 550

(d) 600

Solution to clicker question 6

- The water level in a lake, $\widetilde{X}$, is uniformly distributed over $[500, 600]$. What is the mean water level?

550

(a) 100

(b) 500

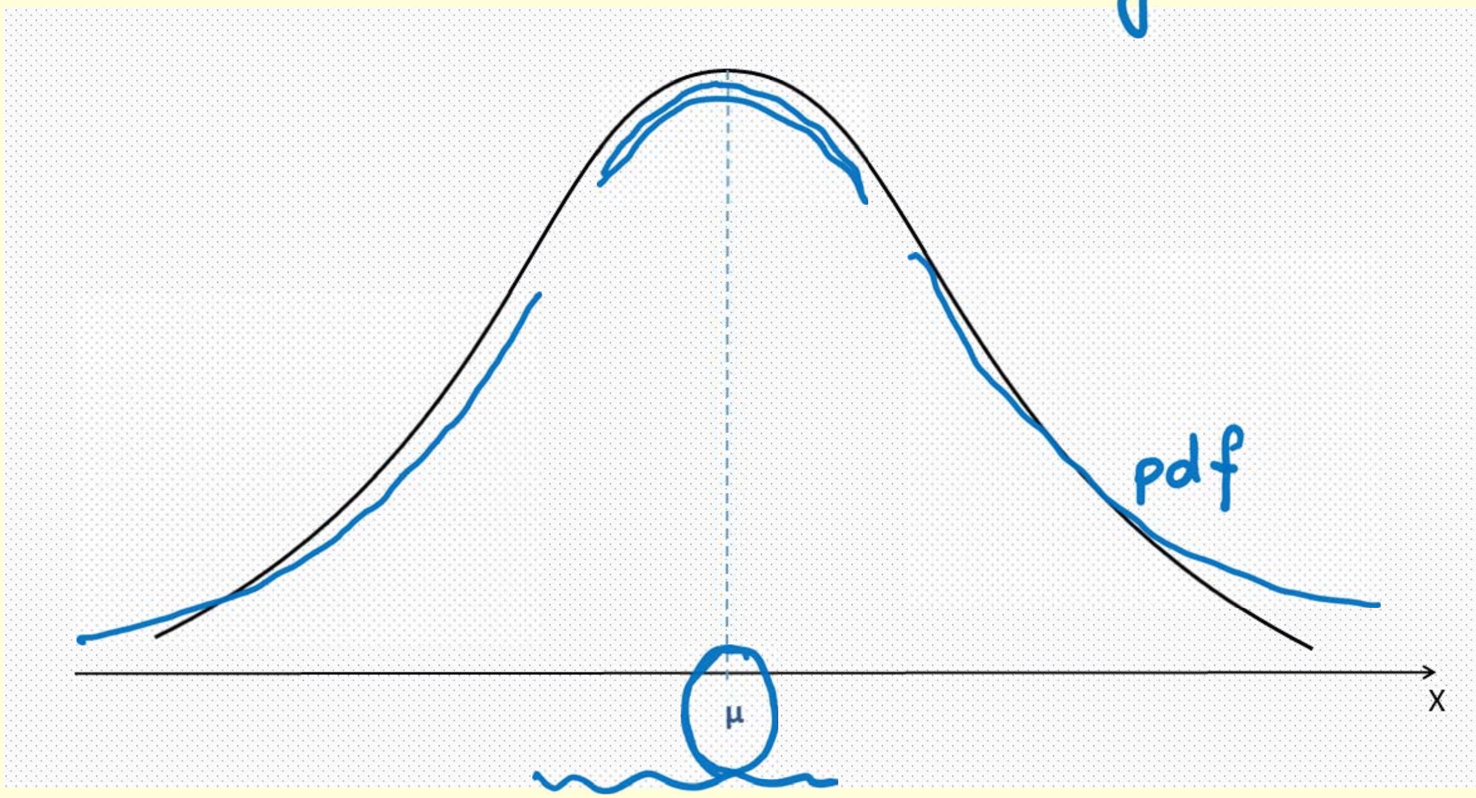
(c) 550

(d) 600

Common Continuous Distributions 2: Normal Distn.

- *Most commonly used distribution:* Test scores, weight, height, IQ, stock returns.......

Bell curve
Gaussian distn



Common Continuous Distributions 2: Normal Distn.

- *Most commonly used distribution:* Test scores, weight, height, IQ, stock returns.......
- Has 2 parameters: *mean* μ, *variance* σ^2

Common Continuous Distributions 2: Normal Distn.

- *Most commonly used distribution:* Test scores, weight, height, IQ, stock returns.......
- Has 2 parameters: *mean* μ, *variance* σ^2
- $\widetilde{X} \sim N(\mu, \sigma^2)$: *p.d.f*:

s. devn = σ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty \leq x \leq \infty$$

pdf

Common Continuous Distributions 2: Normal Distn.

- *Most commonly used distribution:* Test scores, weight, height, IQ, stock returns.......
- Has 2 parameters: *mean* μ, *variance* σ^2
- $\widetilde{X} \sim N(\mu, \sigma^2)$: *p.d.f*:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty \leq x \leq \infty$$

- cdf $F(b) = \int_{-\infty}^{b} f(x)dx$: Use software or prob. tables

  Excel $:= NORM.DIST(\cancel{x}\, b, mean, std.dev., TRUE)$

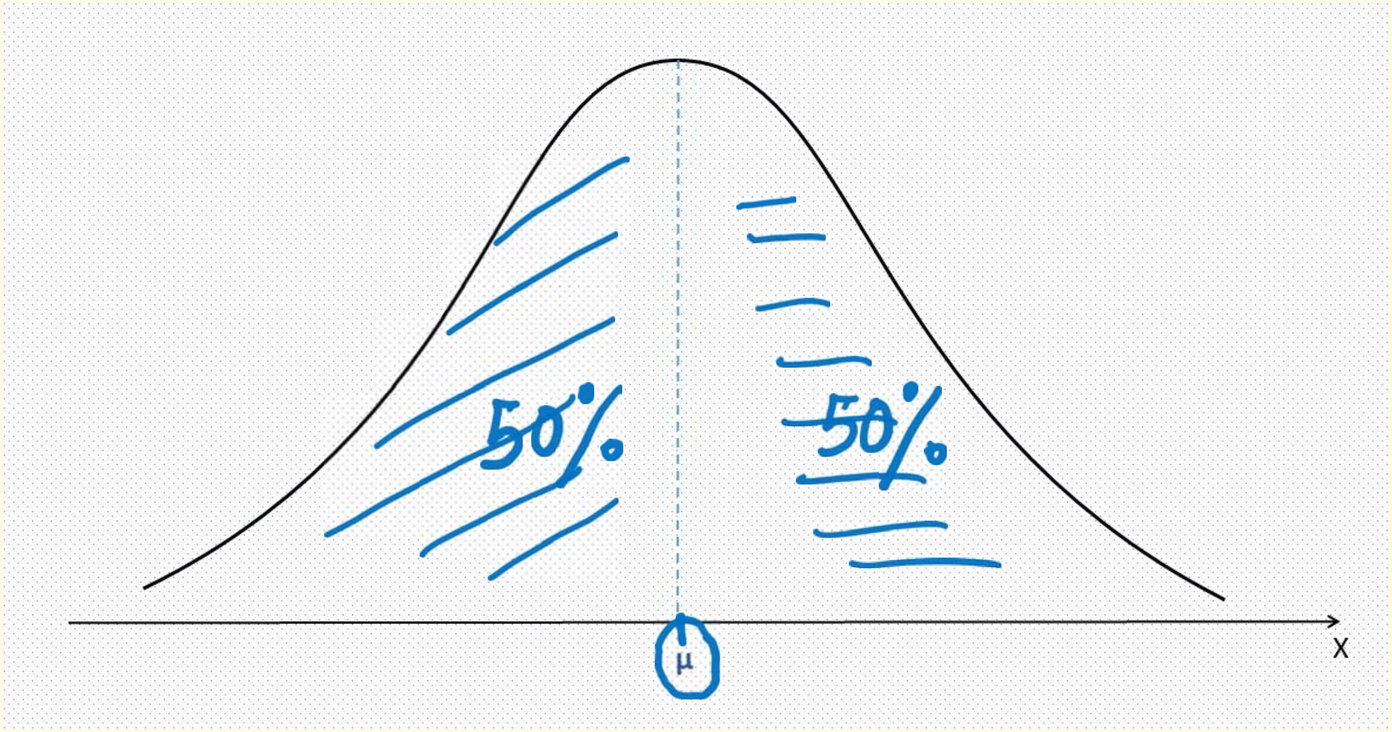
  GoogleSheets : *NORMDIST*

cdf

FALSE : pdf

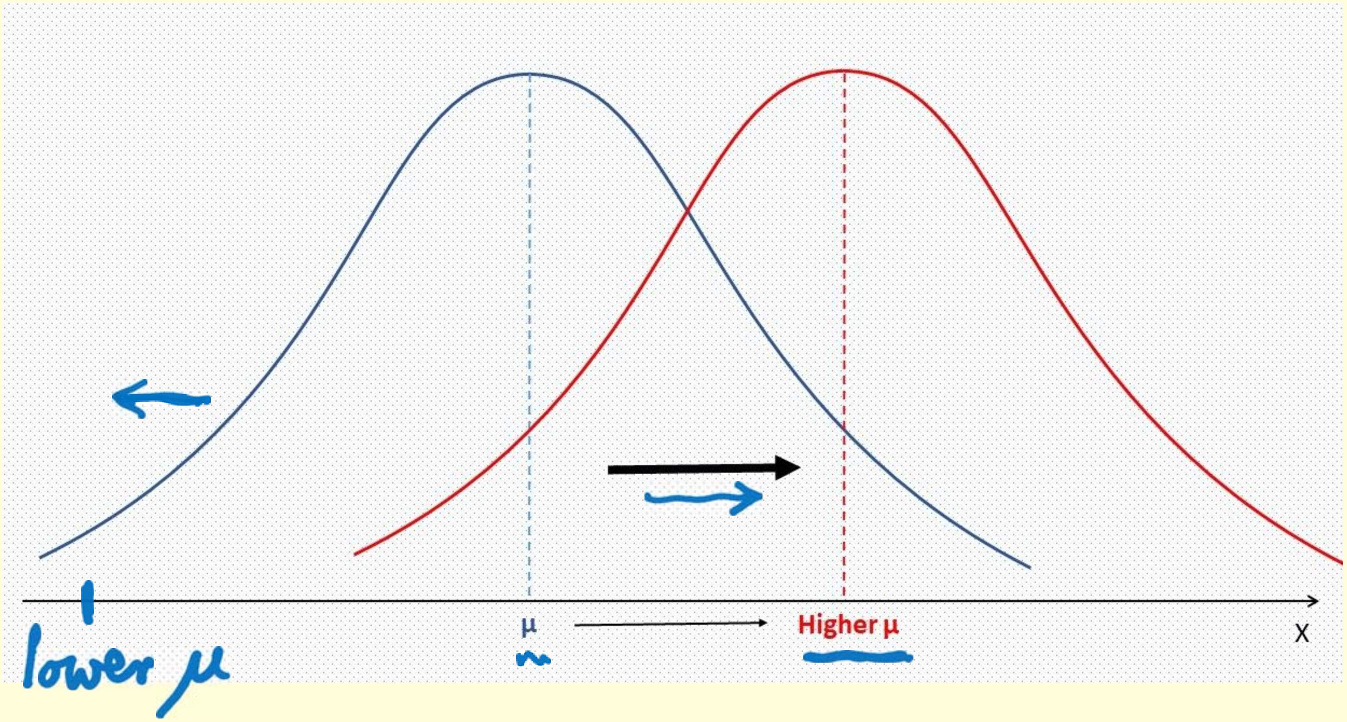
Normal Distn.: Shape

- Symmetric around mean μ :



Normal Distn.: Shape

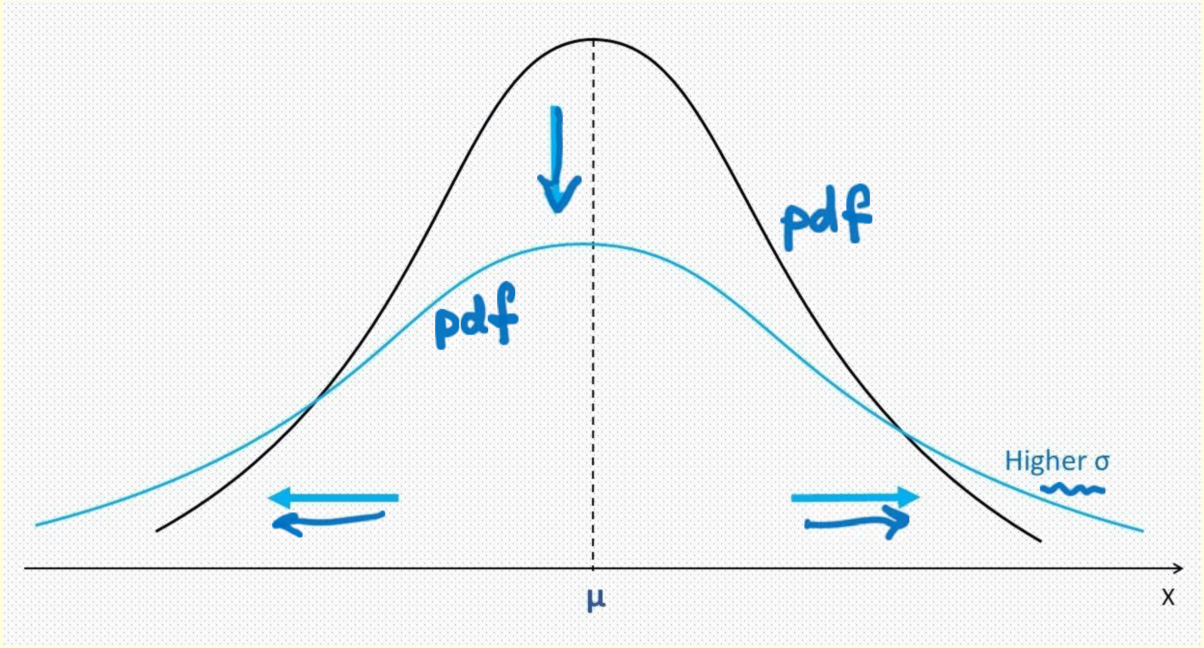
- Symmetric around mean μ :
- As μ changes: move left or right



Normal Distn.: Shape

- Symmetric around mean μ :
- As μ changes: move left or right
- As σ changes: flatten or become steeper

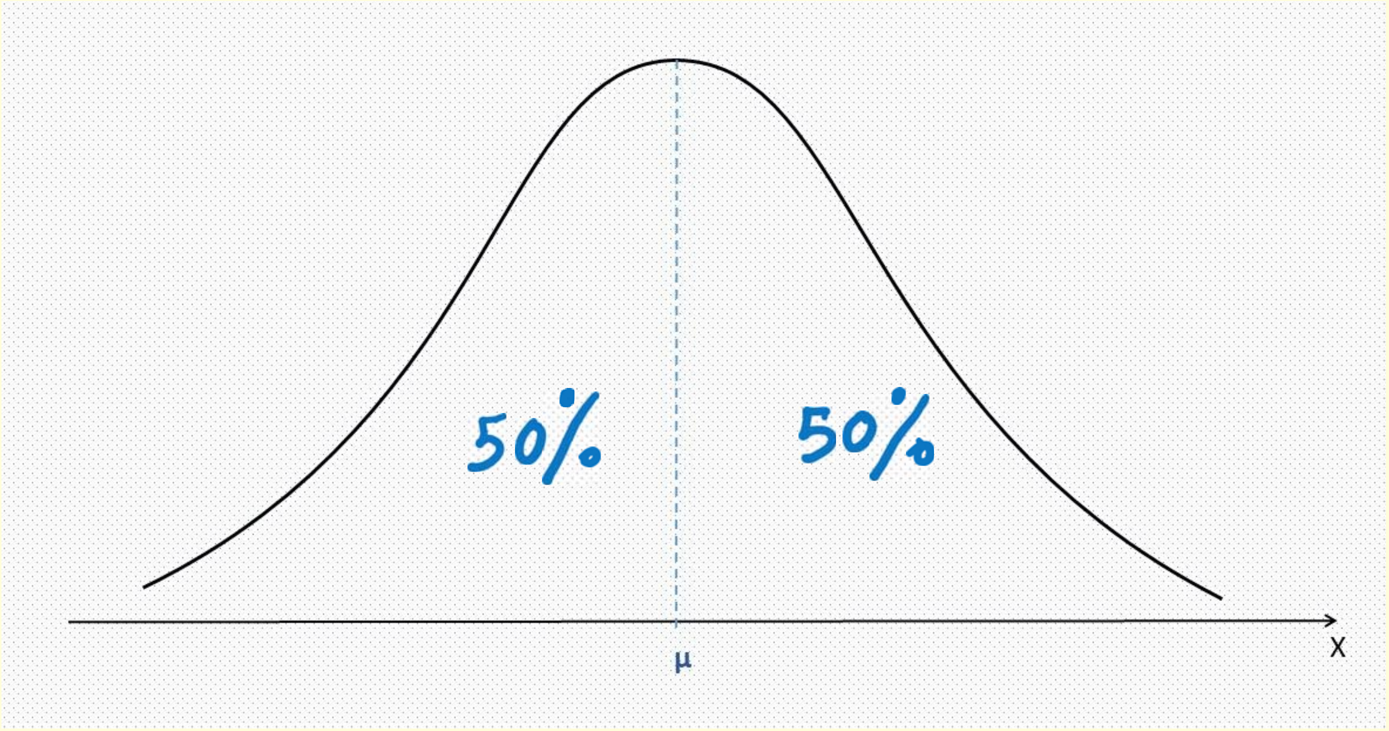
s.d.



Normal Distn. Properties: Properties

- *Symmetric around mean μ :*

  $Prob[\widetilde{X} \leq \mu] = Prob[\mu \leq \widetilde{X}] = 0.5$



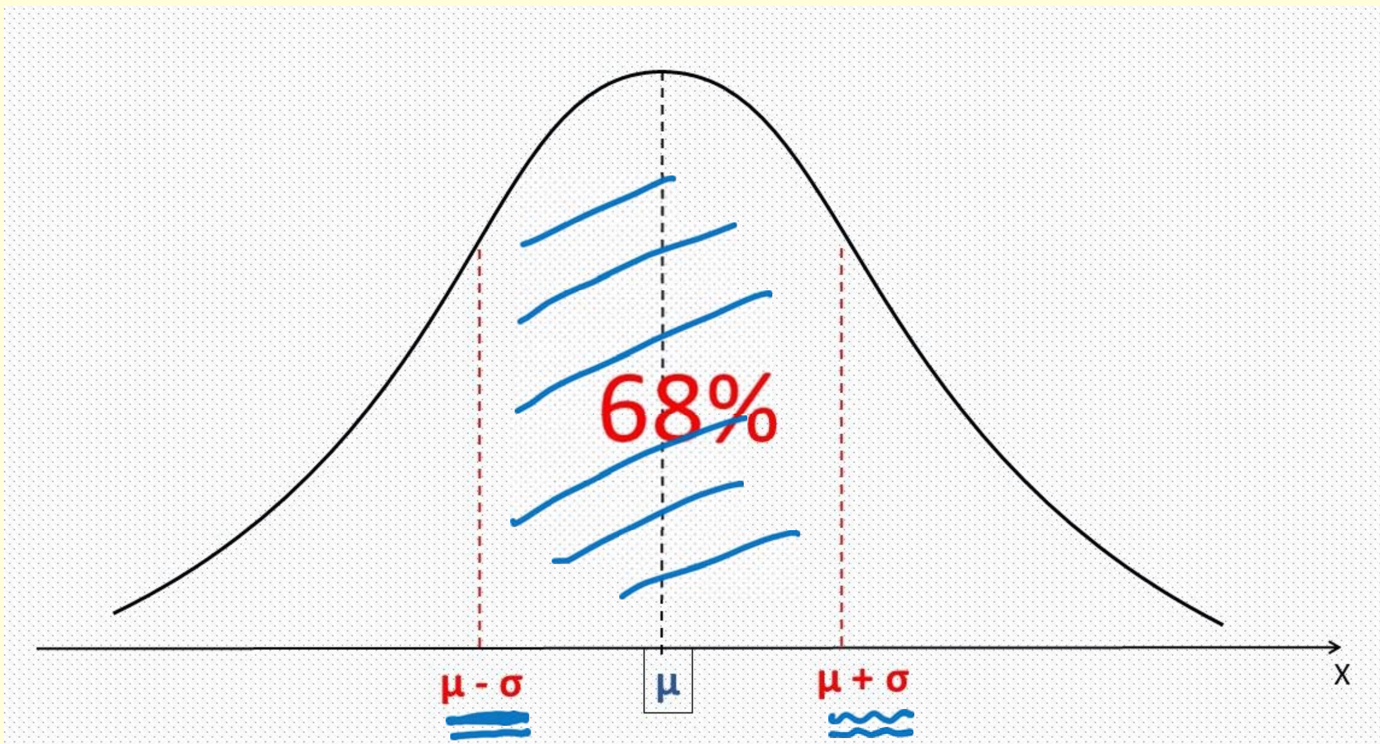
Normal Distn. Properties: ~~Properties~~

68 – 95 – 99.7 rule

- *Symmetric around mean* μ :

  $Prob[\widetilde{X} \le \mu] = Prob[\mu \le \widetilde{X}] = 0.5$

- $Prob[\mu - \sigma \le \widetilde{X} \le \mu + \sigma] = 0.68$

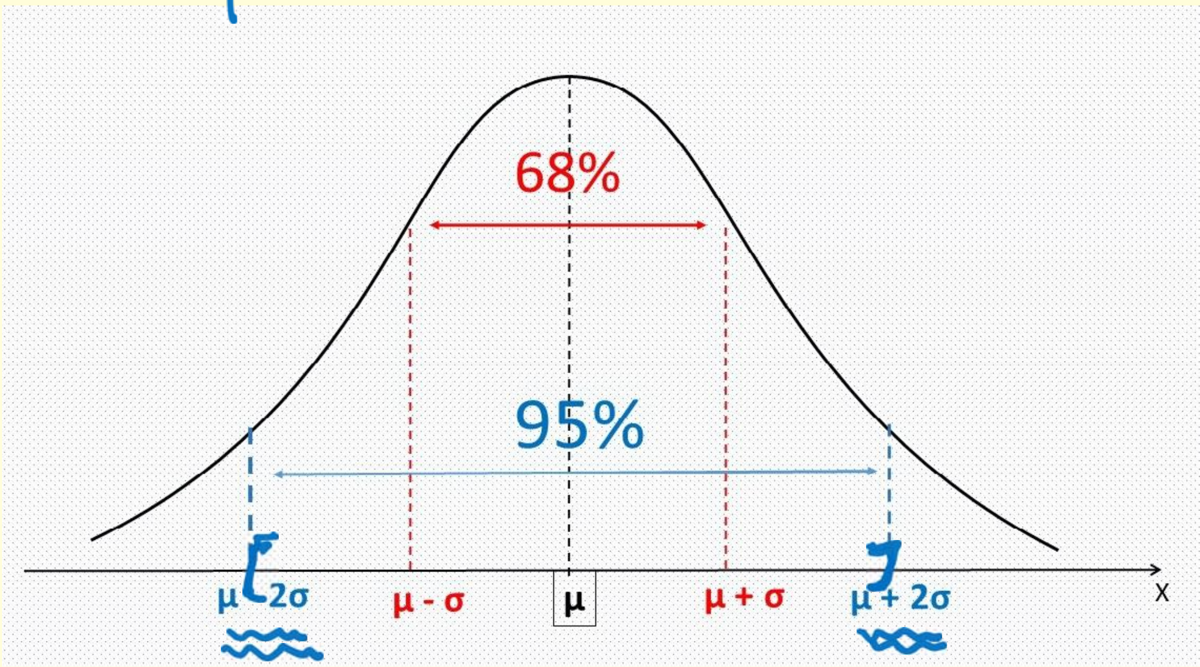


Normal Distn. Properties: Properties

68 - 95 - 99.7

- *Symmetric around mean* μ :

  $Prob[\widetilde{X} \leq \mu] = Prob[\mu \leq \widetilde{X}] = 0.5$
- $Prob[\mu - \sigma \leq \widetilde{X} \leq \mu + \sigma] = 0.68$
- $Prob[\mu - 2\sigma \leq \widetilde{X} \leq \mu + 2\sigma] = 0.95$

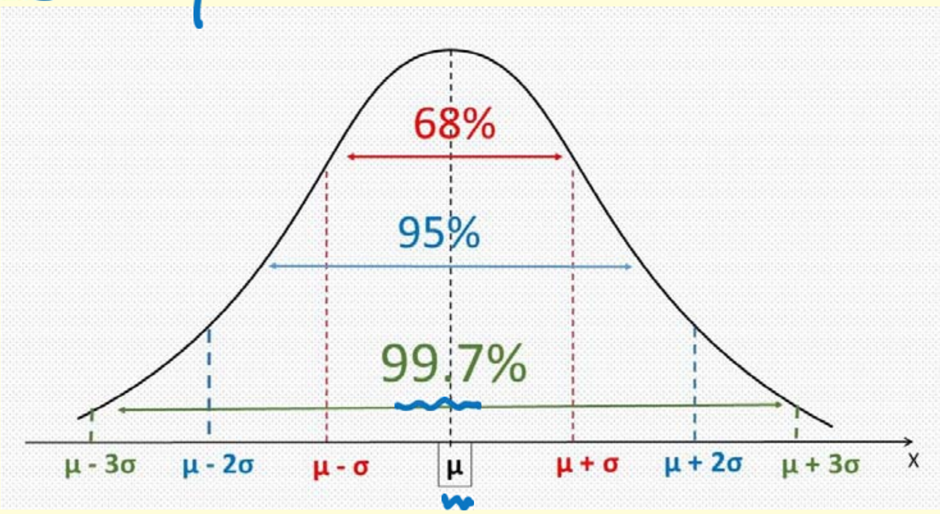


Normal Distn. Properties: Properties

68 - 95 - 99.7

- *Symmetric around mean* μ :
  $Prob[\widetilde{X} \leq \mu] = Prob[\mu \leq \widetilde{X}] = 0.5$
- $Prob[\mu - \sigma \leq \widetilde{X} \leq \mu + \sigma] = 0.68$
- $Prob[\mu - 2\sigma \leq \widetilde{X} \leq \mu + 2\sigma] = 0.95$
- $Prob[\mu - 3\sigma \leq \widetilde{X} \leq \mu + 3\sigma] = 0.997$

$\mu - 3\sigma$ $\quad$ μ $\quad$ $\mu + 3\sigma$



Normal Distn. Properties: Properties

- *Symmetric around mean* μ :

  $Prob[\widetilde{X} \leq \mu] = Prob[\mu \leq \widetilde{X}] = 0.5$
- $Prob[\mu - \sigma \leq \widetilde{X} \leq \mu + \sigma] = 0.68$
- $Prob[\mu - 2\sigma \leq \widetilde{X} \leq \mu + 2\sigma] = 0.95$
- $Prob[\mu - 3\sigma \leq \widetilde{X} \leq \mu + 3\sigma] = 0.997$
- *Example* (quality control): A light-bulb's life is normally distributed with $\mu = 1000$ hrs and $\sigma = 50$.

  $\Rightarrow$ 99.7% of the bulbs last between 850 and 1150 hrs

$1000 - 3\sigma$
$= 850$

1000

$1000 + 3 \times 50$
$\parallel$
1150

Normal Distn. An example

- Suppose a Doordash driver's delivery time, $\widetilde{t}$, is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$

Normal Distn. An example

- Suppose a Doordash driver's delivery time, $\widetilde{t}$, is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$

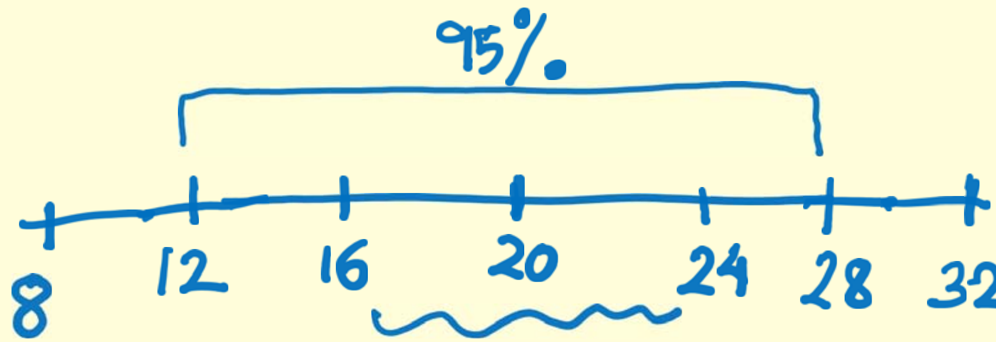
s.d.

- $Prob[16 \le \widetilde{t} \le 24] =$ 0.68

  $Prob[12 \le \widetilde{t} \le 28] =$ 0.95

  $Prob[8 \le \widetilde{t} \le 32] =$ 0.997

68-95-99.7 rule

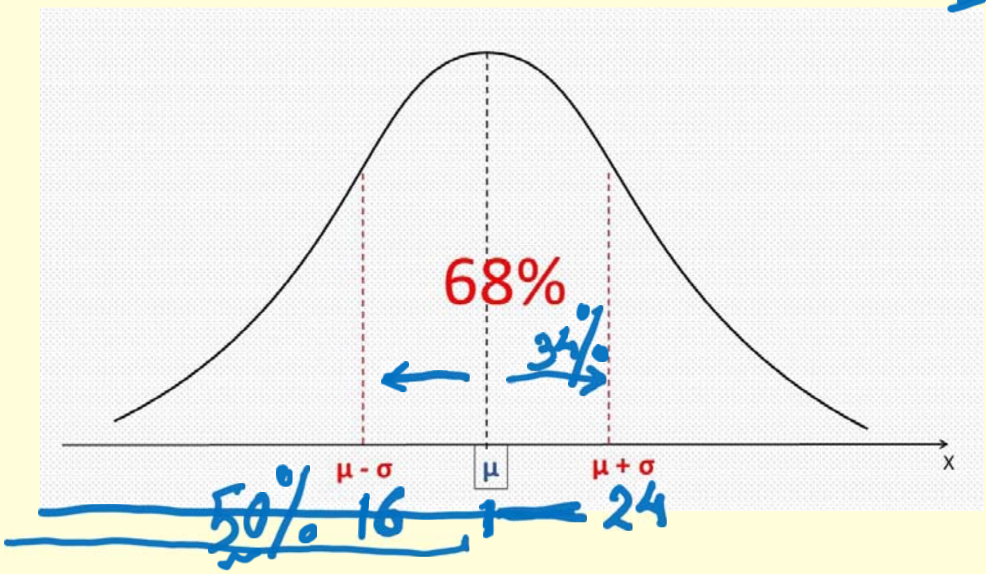


Normal Distn. An example

- Suppose a Doordash driver's delivery time, $\tilde{t}$, is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$
- $Prob[16 \leq \tilde{t} \leq 24] =$

  $Prob[12 \leq \tilde{t} \leq 28] =$

  $Prob[8 \leq \tilde{t} \leq 32] =$
- Other probabilities?: $Prob[\tilde{t} \leq 24] =$ $0.5 + 0.34 = 0.84$



Normal Distn. An example

- Suppose a Doordash driver's delivery time, $\widetilde{t}$, is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$
- $Prob[16 \leq \widetilde{t} \leq 24] =$

  $Prob[12 \leq \widetilde{t} \leq 28] =$

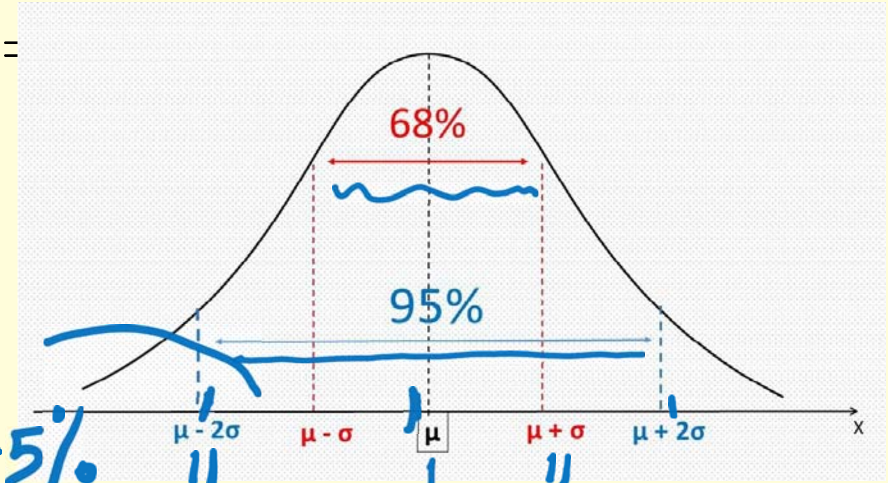
  $Prob[8 \leq \widetilde{t} \leq 32] =$
- Other probabilities?: $Prob[\widetilde{t} \leq 24] =$
- $Prob[12 \leq \widetilde{t} \leq 24] =$

$0 \cdot 34 + 0 \cdot 475$

$= 0 \cdot 815$

$\frac{95}{2}$

$= 47 \cdot 5\%$



12 ⊢ 24

20 ⊢ 34%

Normal Distn. An example

- Suppose a Doordash driver's delivery time, $\widetilde{t}$, is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$
- $Prob[16 \leq \widetilde{t} \leq 24] = 0.68$,......
- Other probabilities?: $Prob[\widetilde{t} \leq 22] = F(22)$

  $Prob[22 \leq \widetilde{t} \leq 26] = F(26) - F(22) =$

Normal Distn. An example

- Suppose a Doordash driver's delivery time, $\widetilde{t}$, is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$
- $Prob[16 \leq \widetilde{t} \leq 24] = 0.68$,......
- Other probabilities?: $Prob[\widetilde{t} \leq 22] = F(22)$ $= 0.69$

  $Prob[22 \leq \widetilde{X} \leq 26] = F(26) - F(22) =$ $0.93 - 0.67 = 0.24$
- Use *Excel* : $= NORM.DIST(x, mean, std.dev., TRUE)$

  GoogleSheets : $NORMDIST$

Clicker question 7

- Suppose a baby's birth weight, $\widetilde{w}$, is normally distributed with $\mu = 3.5$ kg and $\sigma = 0.5$.

  A family is expecting a baby. What is the probability that the baby's weight will be between 3 and 4 kg?

(a) 0.5

(b) 0.68

(c) 0.95

(d) 600

Solution to clicker question 7

- Suppose a baby's birth weight, $\widetilde{w}$, is normally distributed with $\mu = 3.5$ kg and $\sigma = 0.5$.

  A family is expecting a baby. What is the probability that the baby's weight will be between 3 and 4 kg?

  $3-0.5$ $3+0.5$

  (a) 0.5

  (b) 0.68

  (c) 0.95

  (d) 600

Normal Distribution: Properties

- If $\widetilde{X} \sim N(\mu, \sigma^2)$, then $a\widetilde{X} + b \sim N(a\mu + b, a^2\sigma^2)$

mean $\underline{a\mu + b}$ ✓

Var : $a^2\sigma^2$ ✓

Normal Distribution: Properties

- If $\widetilde{X} \sim N(\mu, \sigma^2)$, then $a\widetilde{X} + b \sim N(a\mu + b, a^2\sigma^2)$
- *Example:* If $\widetilde{X} \sim N(3.5, 0.25)$, then

$\boxed{2\widetilde{X} + 10} \sim N(\ 17\ ,\ 1\ )$

$2 \times 3.5 + 10$

var : $4 \times 0.25 = 1$

Normal Distribution: Properties

- If $\widetilde{X} \sim N(\mu, \sigma^2)$, then $a\widetilde{X} + b \sim N(a\mu + b, a^2\sigma^2)$
- *Example:* If $\widetilde{X} \sim N(3.5, 0.25)$, then

  $2\widetilde{X} + 10 \sim N(\quad , \quad)$
- <u>*Sum*</u>: If $\widetilde{X} \sim N(\mu_X, \sigma_X^2)$ and $\widetilde{Y} \sim N(\mu_Y, \sigma_Y^2)$ and $\widetilde{X}$ and $\widetilde{Y}$ are independent,

  $\widetilde{Z} = \widetilde{X} + \widetilde{Y} \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Normal Distribution: Properties

- If $\widetilde{X} \sim N(\mu, \sigma^2)$, then $a\widetilde{X} + b \sim N(a\mu + b, a^2\sigma^2)$
- *Example:* If $\widetilde{X} \sim N(3.5, 0.25)$, then

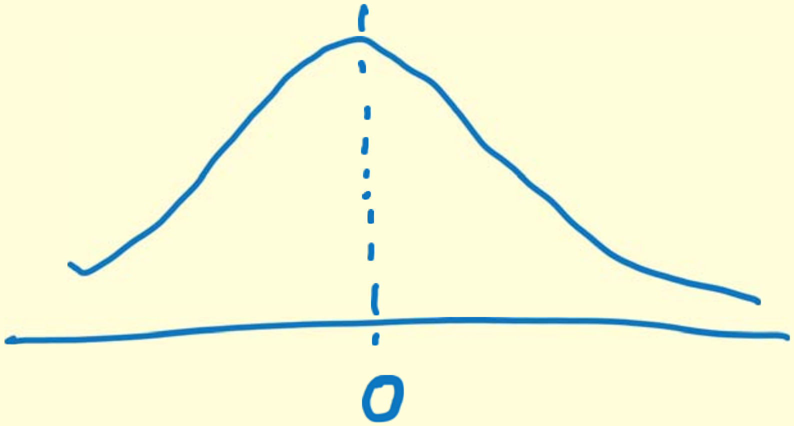
  $2\widetilde{X} + 10 \sim N(\quad , \quad)$
- <u>*Sum*</u>: If $\widetilde{X} \sim N(\mu_X, \sigma_X^2)$ and $\widetilde{Y} \sim N(\mu_Y, \sigma_Y^2)$ and $\widetilde{X}$ and $\widetilde{Y}$ are independent,

  $\widetilde{Z} = \widetilde{X} + \widetilde{Y} \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- *Example:* If husband's income $\sim N(2000, 100)$, and wife's income $\sim N(2500, 200)$, then

  Joint income $\sim N(4500, 300)$

Standard Normal Distribution

- $\widetilde{X} \sim N(0, 1)$: normal distn. with $\mu = 0, \sigma^2 = 1$



Standard Normal Distribution

- $\widetilde{X} \sim N(0,1)$: normal distn. with $\mu = 0, \sigma^2 = 1$
- *Standardization:* If $\widetilde{X} \sim N(\mu, \sigma^2)$, then

$$\text{Z-score: } Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

$$\tilde{X} - \mu \sim N(0, \sigma^2)$$

$$\frac{\tilde{X} - \mu}{\sigma} \sim N\left(\underset{\substack{\| \\ 0}}{\frac{0}{\sigma}}, \underset{\substack{\| \\ 1}}{\frac{1}{\sigma^2}\sigma^2}\right)$$

Standard Normal Distribution

- $\widetilde{X} \sim N(0,1)$: normal distn. with $\mu = 0, \sigma^2 = 1$
- *Standardization:* If $\widetilde{X} \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

- Uses of Z-score: *For comparing across different situations*

Standardization: An example

- *Standardization:* If $\widetilde{X} \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Standardization: An example

- *Standardization:* If $\widetilde{X} \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
- *Example*: (i) *Pam* scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

  (ii) *Tam* scored 85 in a SAT exam in which marks were normally distributed with mean $\mu = 75$ and $\sigma = 40$.

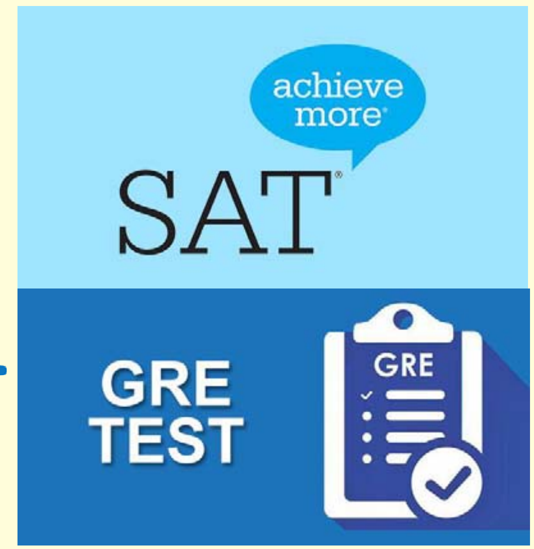
  Who did better?

Standardization: An example

- *Standardization:* If $\widetilde{X} \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
- *Example*: (i) *Pam* scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

  (ii) *Tam* scored 85 in a SAT exam in which marks were normally distributed with mean $\mu = 75$ and $\sigma = 40$.

  Who did better?



- ✓ *Pam's Z*-score $= \dfrac{75-70}{10} = \dfrac{5}{10} = \dfrac{1}{2}$

  Tam's Z-score $= \dfrac{85-75}{40} = \dfrac{10}{40} = \dfrac{1}{4}$

Percentile

- Pam scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

  Q: What fraction of the class did worse than her?

Percentile

- Pam scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

  Q: What fraction of the class did worse than her?

- $Prob\ [\widetilde{X} \leq 75] = F(75) =$?

  Use $\underline{Excel} := NORM.DIST(x, mean, std.dev., TRUE)$

Percentile

- Pam scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

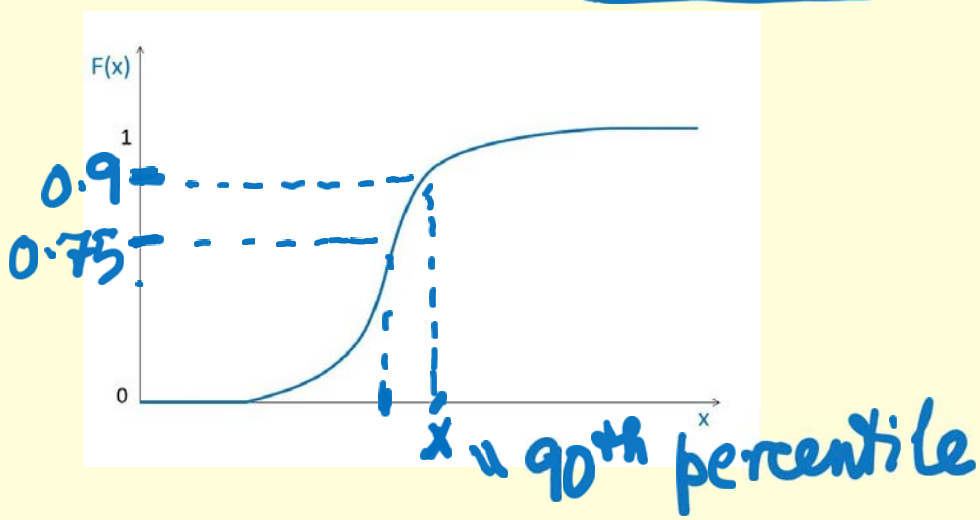
  Q: What fraction of the class did worse than her?

- $Prob\,[\widetilde{X} \leq 75] = F(75) =$

  Use <u>*Excel*</u> $: = NORM.DIST(x, mean, std.dev., TRUE)$

- What is the 90^{th} percentile score?

  i.e. x such that $F(x) = 0.9$?



GRE SCORING SCALE

	25th PERCENTILE	50th PERCENTILE	75th PERCENTILE	90th PERCENTILE
Verbal Reasoning	145	151	157	162
Quantitative Reasoning	147	153	160	167
Analytical Writing	3.5	4.0	4.5	5.0

Percentile

- Pam scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

  Q: What fraction of the class did worse than her?

- $Prob\,[\widetilde{X} \leq 75] = F(75) =$

  Use $\underline{Excel} := NORM.DIST(x, mean, std.dev., TRUE)$

- What is the 90^{th} percentile score?

  i.e. x such that $F(x) = 0.9$? $\longrightarrow$ 82.8

- Use $\underline{Excel} := NORMINV(p, mean, std.dev.)$

Percentile

- Pam scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

  Q: What fraction of the class did worse than her?

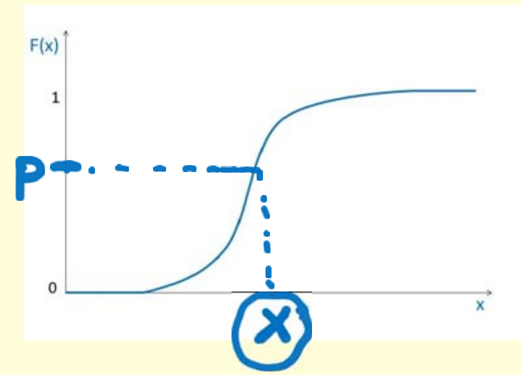
- $Prob\,[\widetilde{X} \leq 75] = F(75) =$

  Use <u>*Excel*</u> : $= NORM.DIST(x, mean, std.dev., TRUE)$

- What is the 90^{th} percentile score?

  i.e. x such that $F(x) = 0.9$? 82·8

- Use <u>*Excel*</u> : $= NORMINV(p, mean, std.dev.)$



- More general **percentile**: Find x s.t. $F(x) = p$

Percentile

- Pam scored 75 in a SAT exam in which marks were normally distributed with mean $\mu = 70$ and $\sigma = 10$.

  Q: What fraction of the class did worse than her?

- $Prob\ [\widetilde{X} \leq 75] = F(75) =$

  Use $\underline{Excel}$ $: = NORM.DIST(x, mean, std.dev., TRUE)$

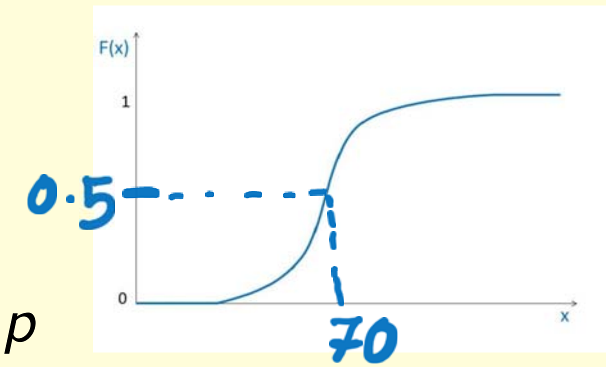
- What is the 90^{th} percentile score?

  i.e. x such that $F(x) = 0.9$?

- Use $\underline{Excel}$ $: = NORMINV(p, mean, std.dev.)$

- More general **percentile**: Find x s.t. $F(x) = p$

- 50^{th} percentile: **median** i.e x s.t. $F(x) = 0.5$ → Half of the class is below this.



Module recap

- Covariance and Correlation
- *Continuous r.v.s*: c.d.f. and p.d.f.
- Expectation and Variance
- Common continuous distributions: *Uniform* and *Normal*
- Many other continuous distributions (not covered here):

  Exponential, *t*, *F*, *Beta*, *Pareto*,......