

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.) $p(x) = \text{Prob}(\tilde{X} = x)$

Cumulative distribution fn. (c.d.f.):

$$\text{Prob}(\tilde{X} \leq x) = \sum_{y: y \leq x} p(y)$$

Review: Prob. Module 3

- Random variables: Discrete
- Probability mass function (p.m.f.) $p(x) = \text{Prob}(\tilde{X} = x)$

Cumulative distribution fn. (c.d.f.):

$$\text{Prob}(\tilde{X} \leq x) = \sum_{y: y \leq x} p(y)$$

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

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

Review: Prob. Module 3

- Common discrete distributions:

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

Binomial (number of successes in n trials):

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



Review: Prob. Module 3

- Common discrete distributions:

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

Binomial (number of successes in n trials):

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

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

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

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
 - (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

Covariance and Correlation

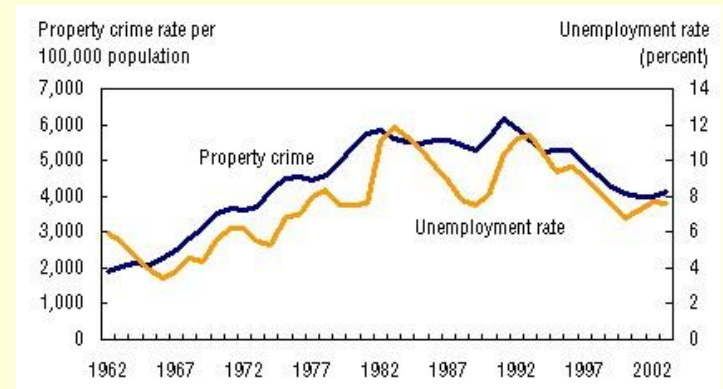
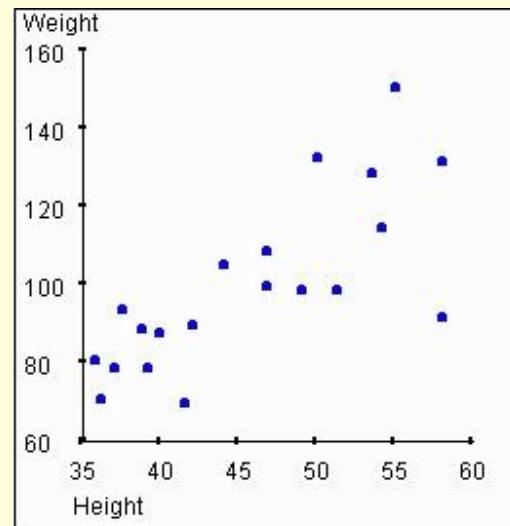
- **Expectation** $\mu_X = E\tilde{X}$: mean/average value of \tilde{X}
Variance $Var(\tilde{X})$: dispersion of \tilde{X}

Covariance and Correlation

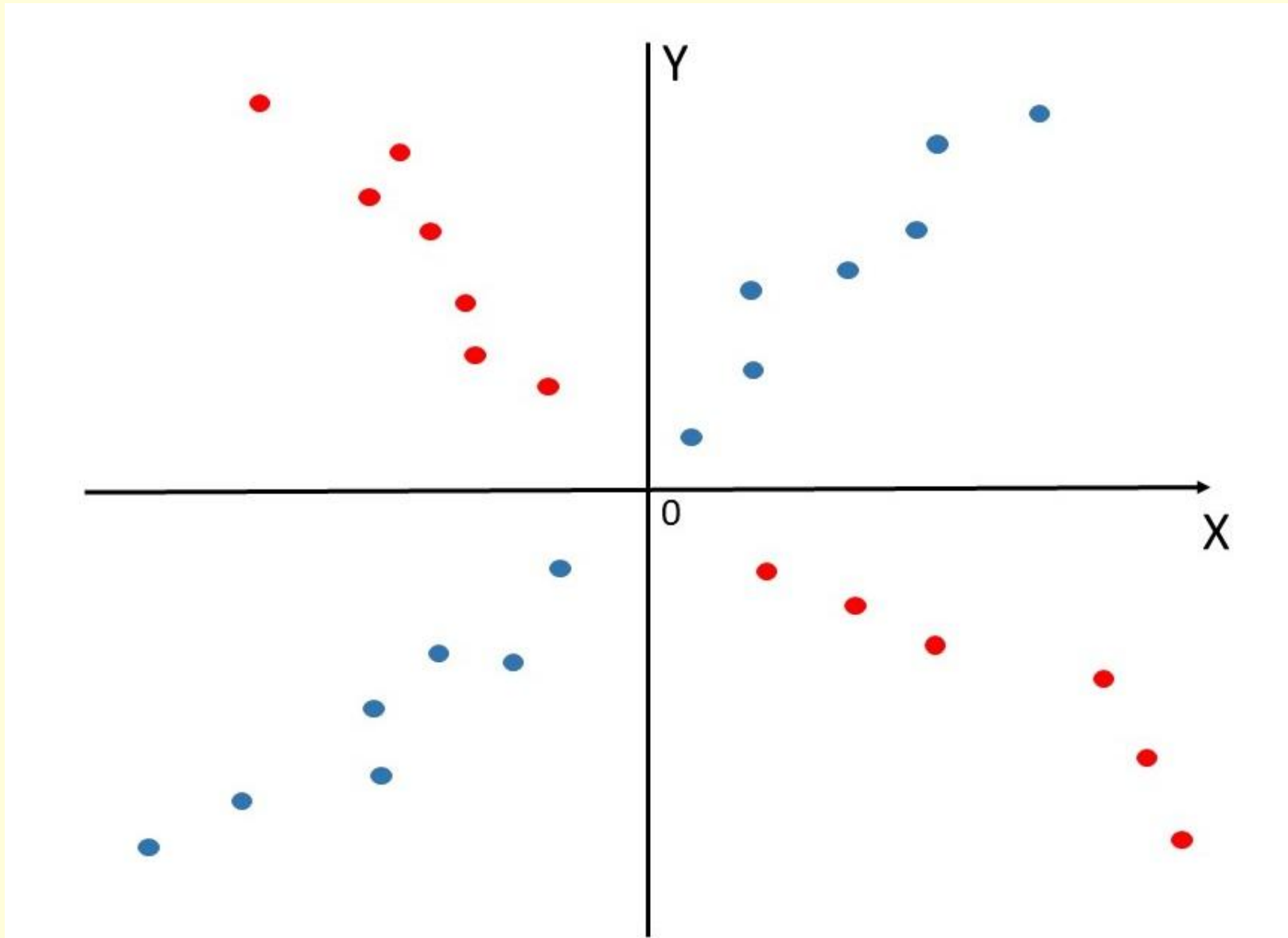
- **Expectation** $\mu_X = E\tilde{X}$: mean/average value of \tilde{X}
Variance $Var(\tilde{X})$: dispersion of \tilde{X}
- **Covariance** (\tilde{X}, \tilde{Y}) : relationship between \tilde{X} and \tilde{Y}

Covariance and Correlation

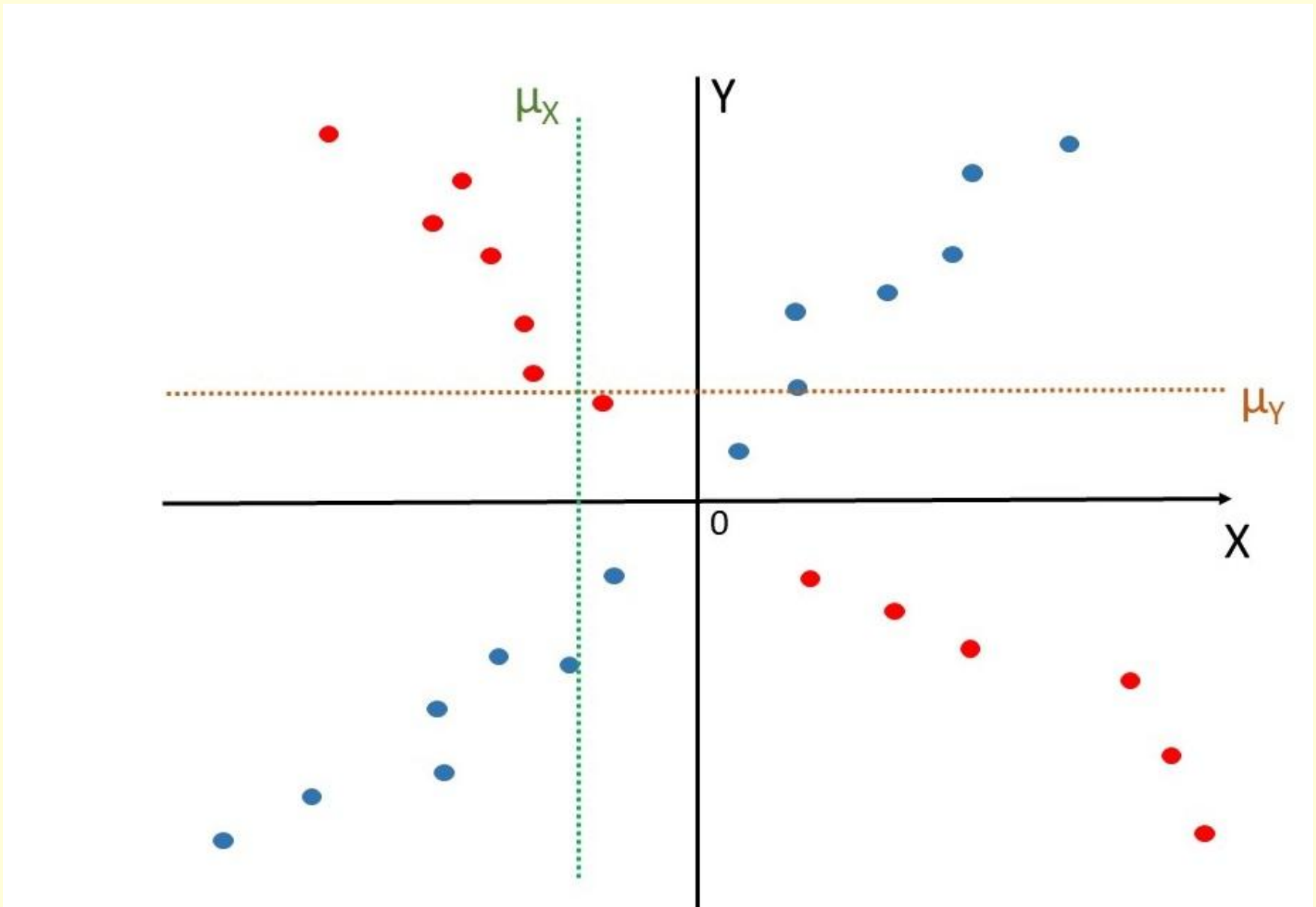
- **Expectation** $\mu_X = E\tilde{X}$: mean/average value of \tilde{X}
- **Variance** $Var(\tilde{X})$: dispersion of \tilde{X}
- **Covariance** (\tilde{X}, \tilde{Y}) : relationship between \tilde{X} and \tilde{Y}
- **Examples**



Covariance: Basic idea



Covariance: Basic idea



Covariance: definition

- $Cov(\tilde{X}, \tilde{Y}) = \sigma_{XY} = E(\tilde{X} - \mu_X)(\tilde{Y} - \mu_Y)$
 $= \sum_{\text{all } x,y} (x - \mu_X)(y - \mu_Y)p(x, y)$

Covariance: an easier formula

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

Covariance: an easier formula

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

Computing covariance: an example

- Joint pmf of \tilde{X} and \tilde{Y} :

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

- $Cov(\tilde{X}, \tilde{Y}) = \sigma_{XY} = E(\tilde{X}\tilde{Y}) - E(\tilde{X})E(\tilde{Y}) =$

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
 - (b) The dispersion of a random variable
 - (c) The maximum value of a random variable
 - (d) Relationship between two random variables

Covariance: some properties

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

Covariance: some properties

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

Covariance: some properties

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

Covariance: some properties

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

Covariance: some properties

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

Correlation coefficient

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

Correlation coefficient

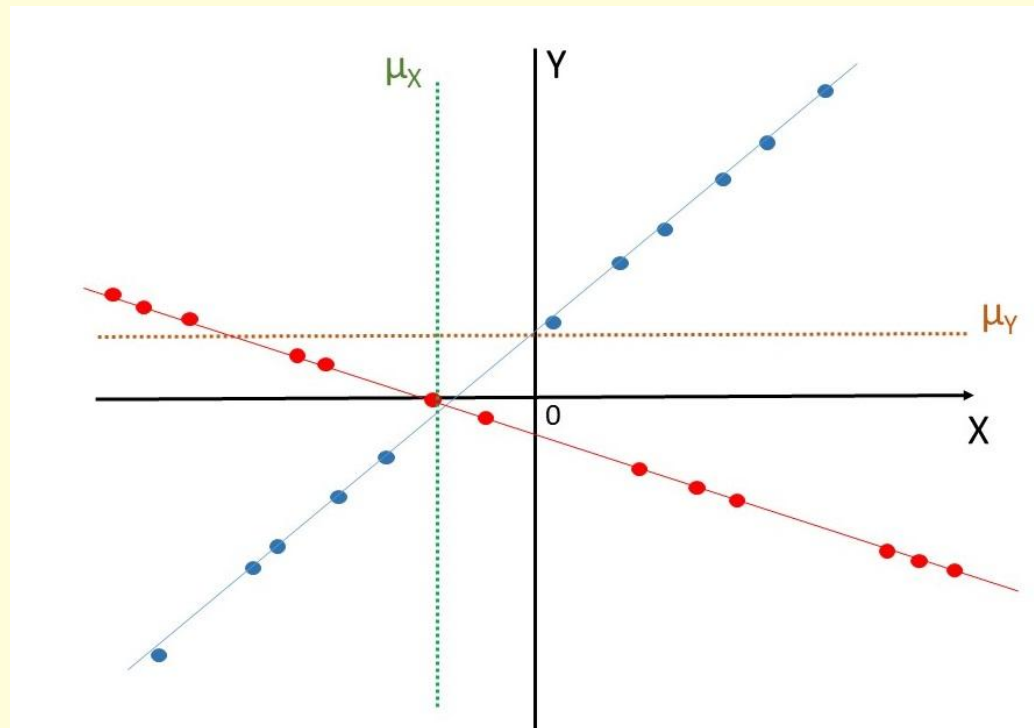
- $\text{Corr}(\tilde{X}, \tilde{Y}) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\text{Cov}(\tilde{X}, \tilde{Y})}{\sqrt{\text{Var}(\tilde{X}) \text{Var}(\tilde{Y})}}$
- $-1 \leq \rho_{XY} \leq 1$

Correlation coefficient

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

Correlation coefficient: Properties

- $-1 \leq \rho_{XY} \leq 1$
- If $\rho_{XY} > 0$: \tilde{X} and \tilde{Y} are positively correlated
- If $\rho_{XY} = 1$: \tilde{X} and \tilde{Y} are *perfectly, linearly, positively correlated*



Correlation coefficient: Properties

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

Correlation coefficient: Properties

- $-1 \leq \rho_{XY} \leq 1$
- If $\rho_{XY} > 0$: \tilde{X} and \tilde{Y} are positively correlated
If $\rho_{XY} = 1$: \tilde{X} and \tilde{Y} are *perfectly, linearly, positively correlated*
- If $\rho_{XY} < 0$: \tilde{X} and \tilde{Y} are negatively correlated
If $\rho_{XY} = -1$: \tilde{X} and \tilde{Y} are *perfectly, linearly, negatively correlated*
- $\text{Corr}(\tilde{X}, \tilde{Y}) = \text{Corr}(a + c\tilde{X}, b + d\tilde{Y})$

Correlation: an example

- Joint pmf of \tilde{X} and \tilde{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

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

$$\text{Cov}(\tilde{X}, \tilde{Y}) = -2.8, \text{Var}(\tilde{X}) = \quad, \text{Var}(\tilde{Y}) =$$

Clicker question 3

- **Q:** Which one of the following correlation coefficients measures the strongest relationship between random variables \tilde{X} and \tilde{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 \tilde{X} and \tilde{Y} ?
 - (a) -0.8
 - (b) -0.4
 - (c) 0.3
 - (d) 0.6

Random Variables: Continuous

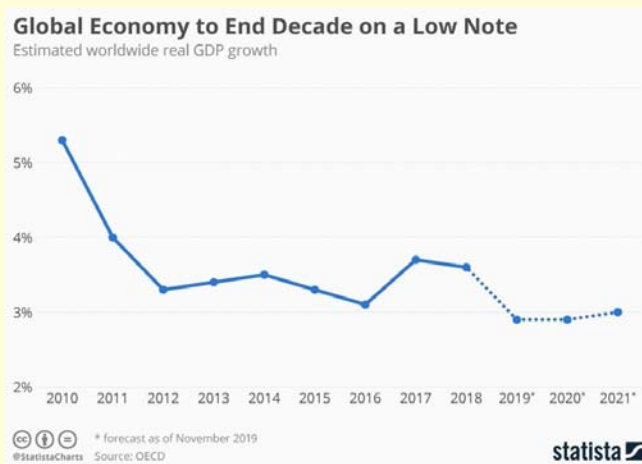
- *Random variable \tilde{X} : A variable taking on numerical values based on the outcome of some random event*

Random Variables: Continuous

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

Random Variables: Continuous

- *Random variable \tilde{X}* : A variable taking on *numerical values* based on the outcome of some random event
- It is *continuous* if \tilde{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 \tilde{X}* : A variable taking on *numerical values* based on the outcome of some random event
- It is *continuous* if \tilde{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

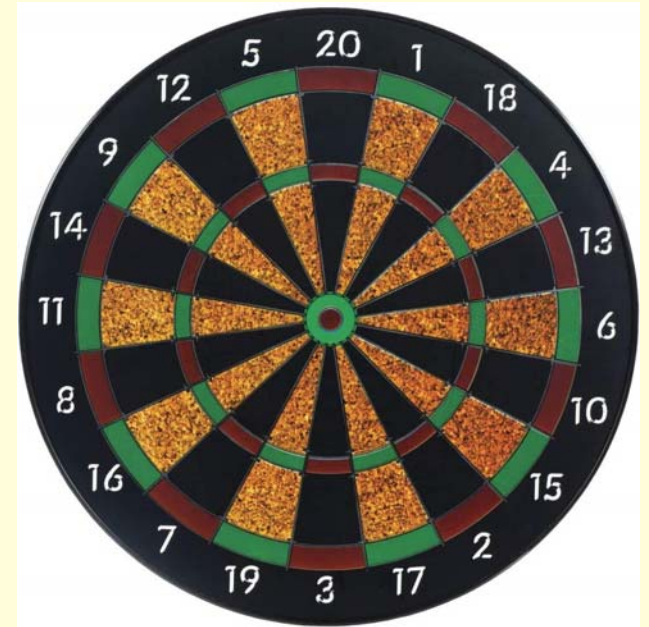
- $M =$ maximum possible value of \tilde{X}

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

$Range = M - m$

Continuous Random Variables: probabilities

- $M =$ maximum possible value of \tilde{X}
 $m =$ minimum possible value of \tilde{X}
 $Range = M - m$
- Prob $[\tilde{X} = x] = ?$



Continuous Random Variables: probabilities

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

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

$$\text{Range} = M - m$$

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

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

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

Continuous Random Variables: c.d.f.

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

Continuous Random Variables: c.d.f.

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

Continuous Random Variables: c.d.f.

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

Continuous Random Variables: c.d.f.

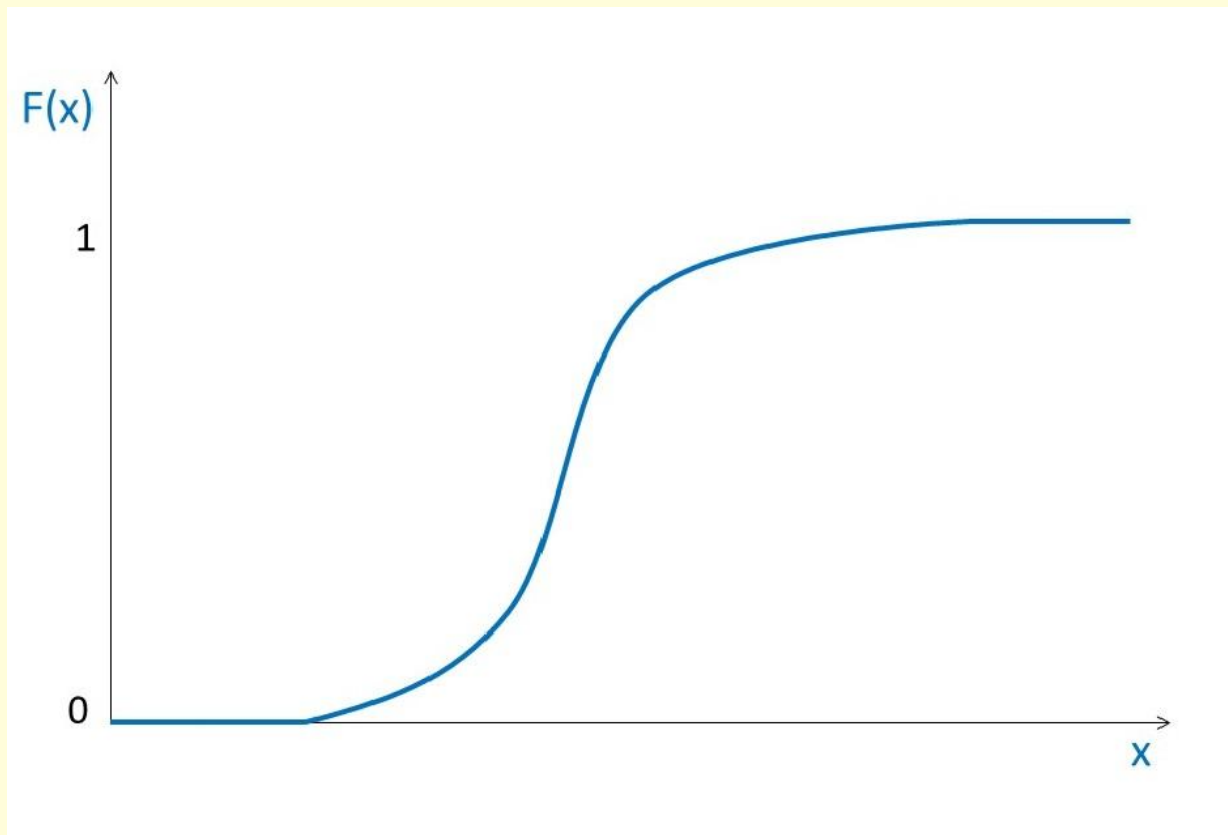
- *Recall:* $\text{Prob} [\tilde{X} \leq x] = F(x)$: c.d.f. of \tilde{X}
- **Prob** $\{\tilde{X} \in [a, b]\} = ?$
- **Prob** $\{\tilde{X} \in [a, b]\} = P[\tilde{X} \leq b] - P[\tilde{X} \leq a] = F(b) - F(a)$
- How to get an idea of which points are more/less likely?

Continuous r.v.s: prob. density function

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

Continuous r.v.s: prob. density function

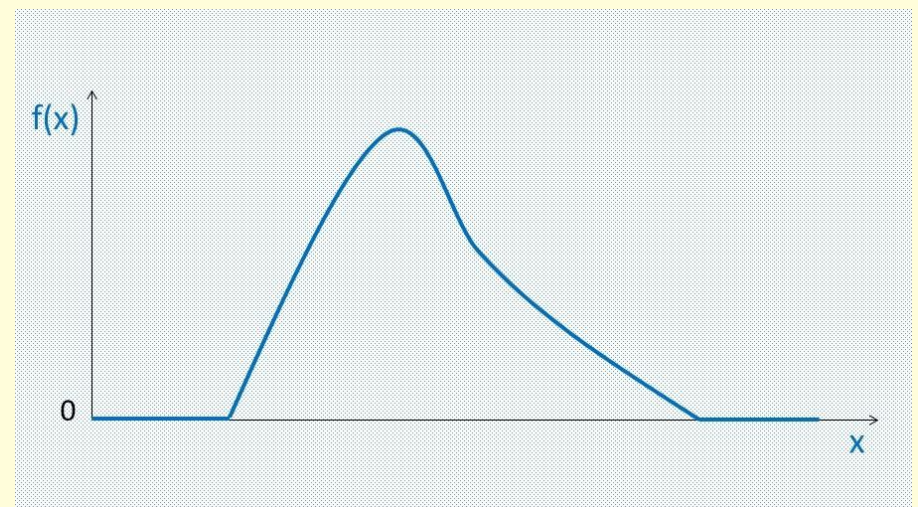
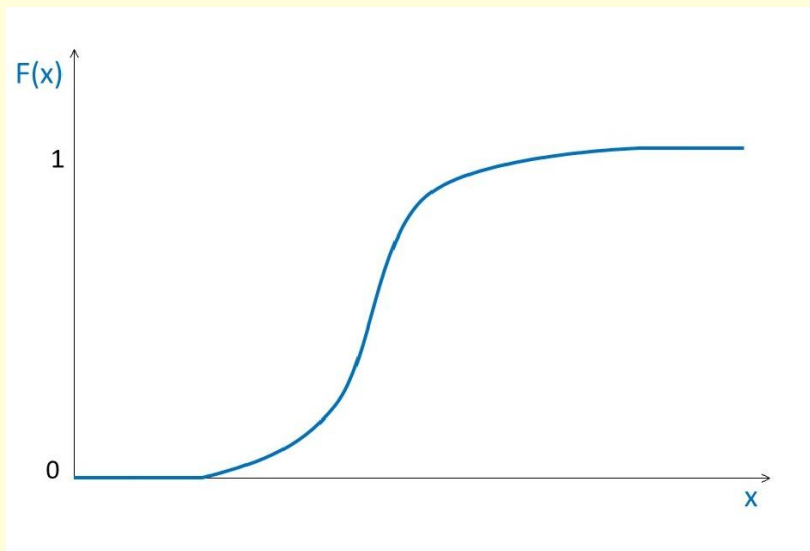
- Recall in discrete case: $F(x_1) - F(x_0) = p(x_1)$
- *Continuous r.v.:* Look at rate of change in cdf



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.) $f(x)$:**

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



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.) $f(x)$:**

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

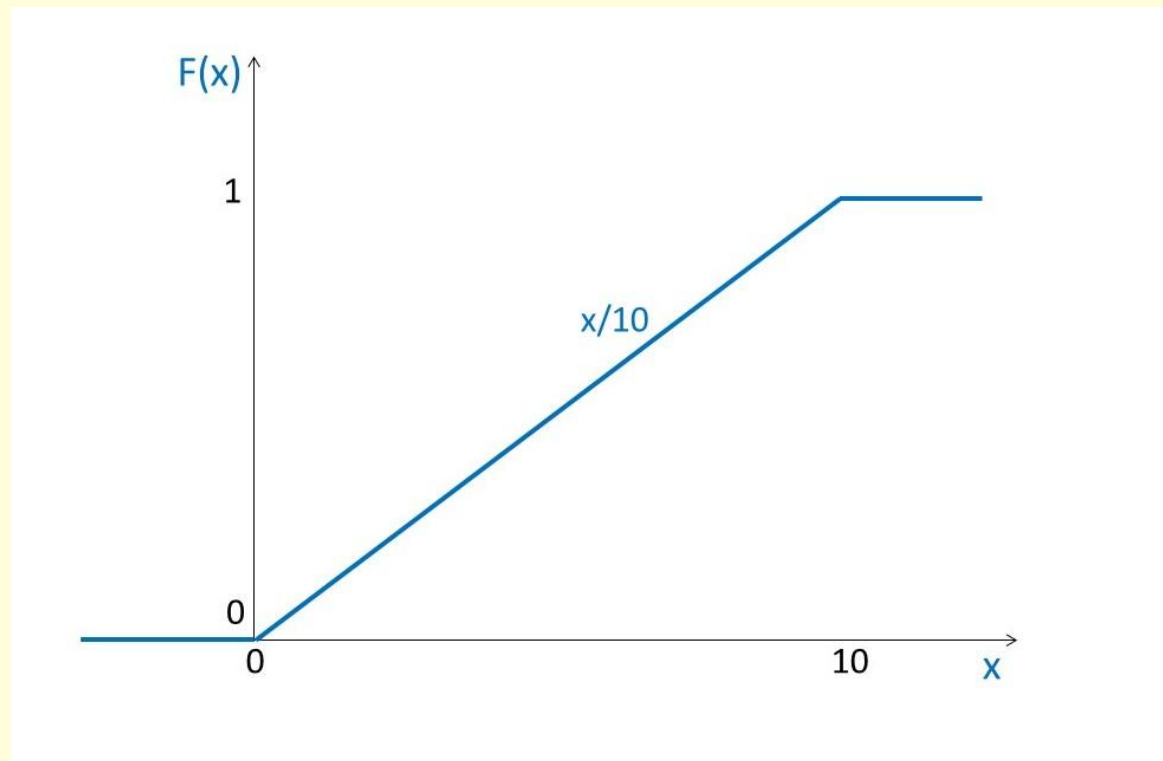
- Interpretation of pdf $f(x)$:

$f(x)$ high: relative likelihood of occurrence 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** $x > 10$



Cdf and pdf: an example

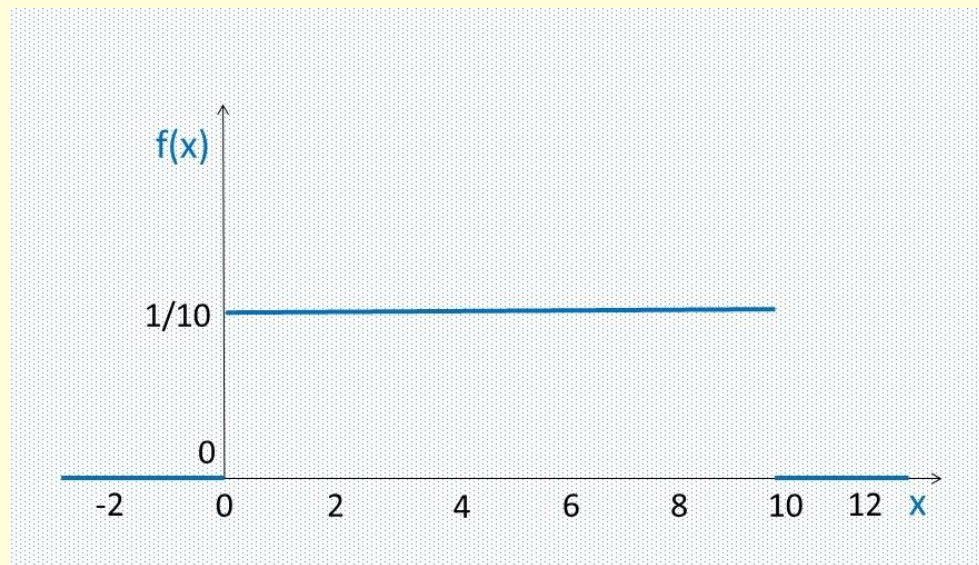
- *Uniform distribution over $[0, 10]$* : $F(x) = 0$ **for** $x < 0$,
 $F(x) = \frac{x}{10}$ **for** $x \in [0, 10]$, $F(x) = 1$ **for** $x > 10$
- **Relative likelihoods of different values?**

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

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

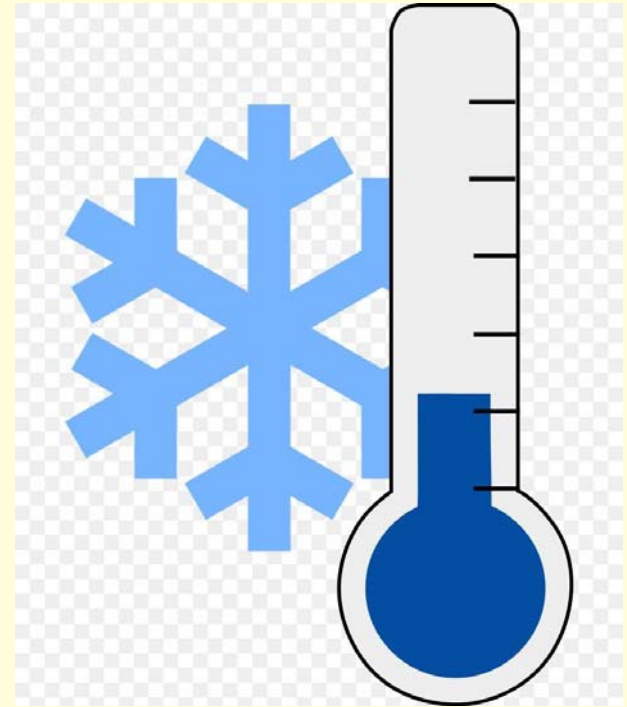
- **Q:** What does c.d.f. and p.d.f. stand for? Pick the option that has both terms correct.
 - (a)** concentrated density function, probability density function
 - (b)** cumulative density function, probability density function
 - (c)** cumulative density function, probability distribution function
 - (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

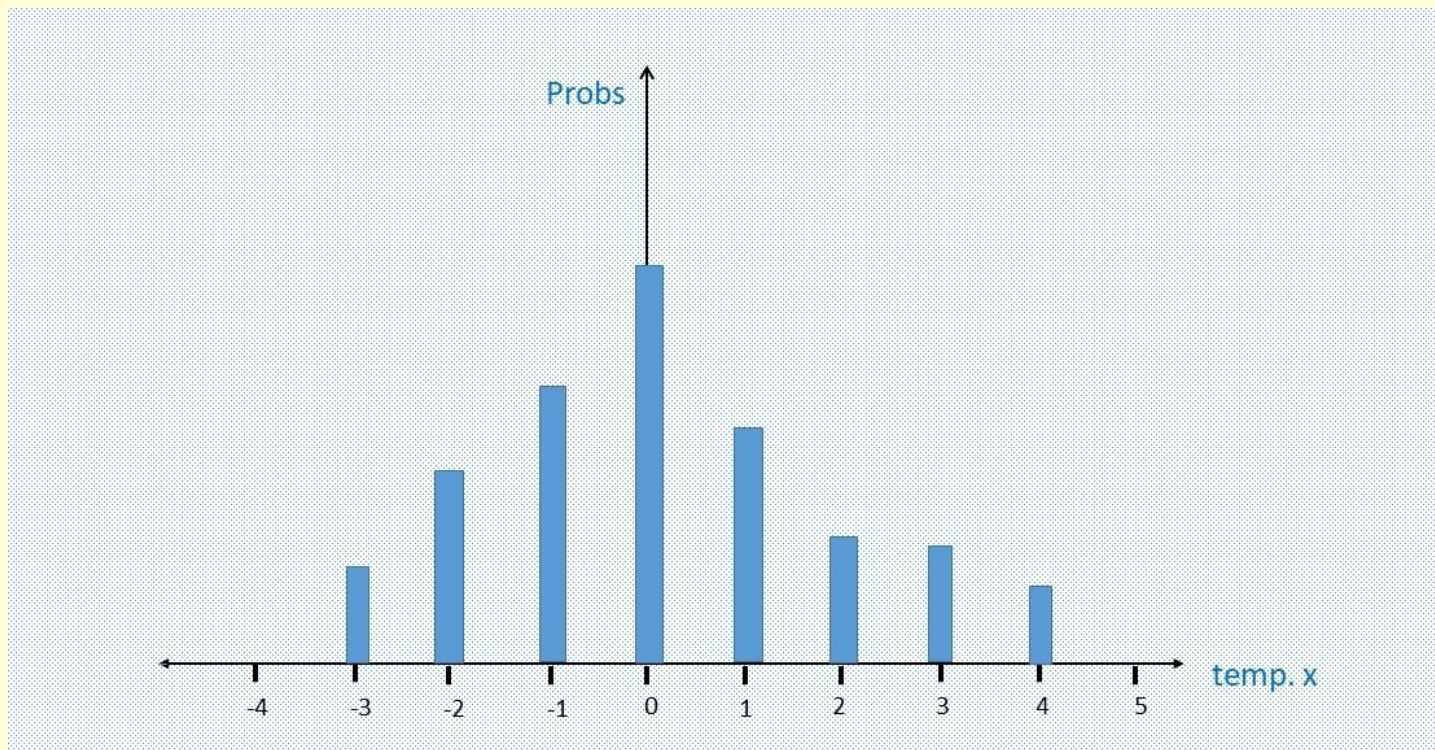
Prob. density fn.: An alternative view

- \tilde{X} = temp. in a city



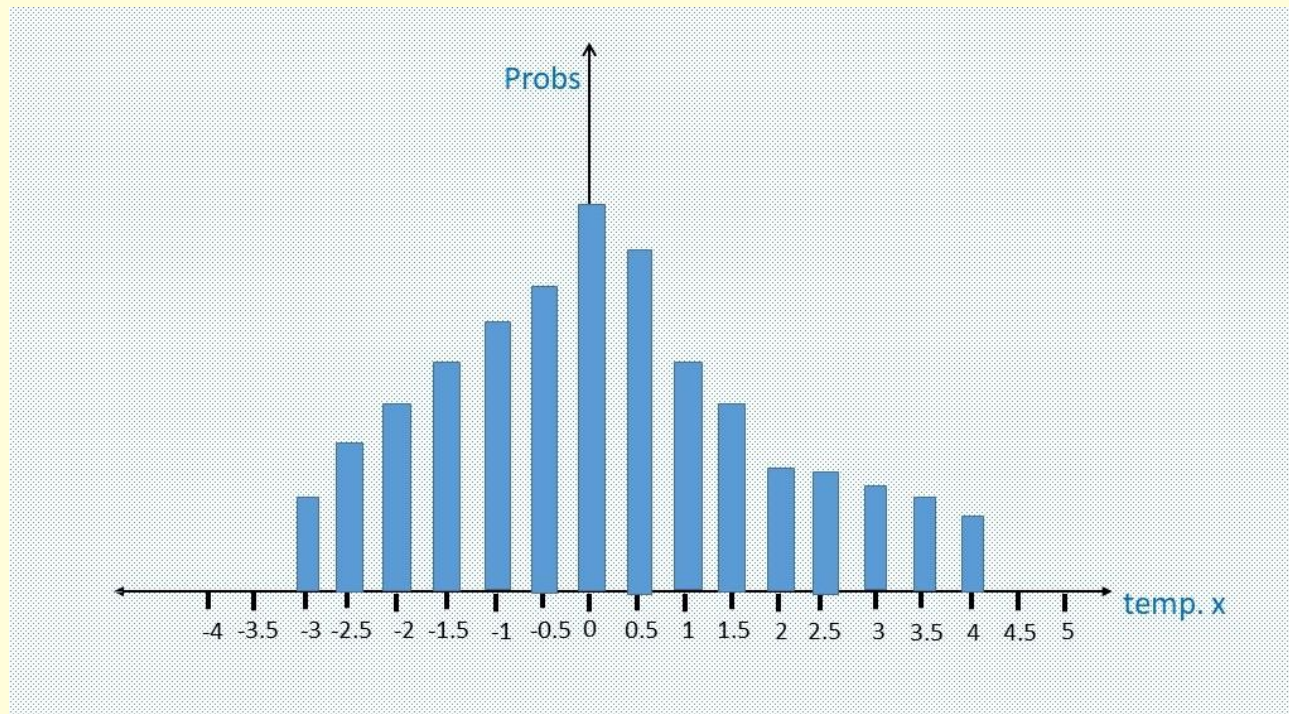
Prob. density fn.: An alternative view

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



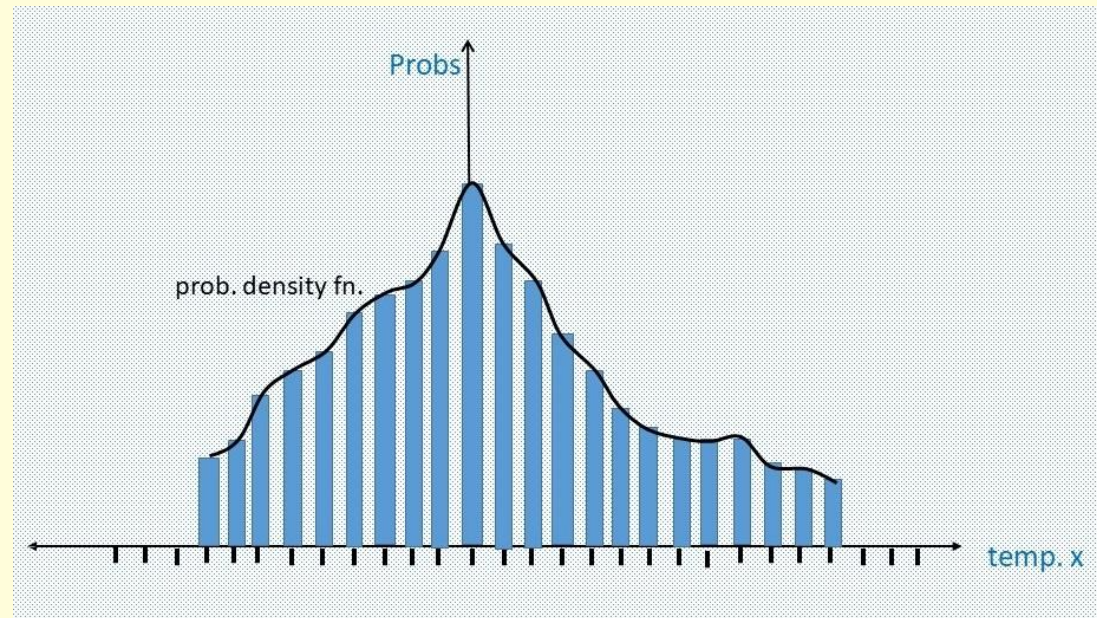
Prob. density fn.: An alternative view

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



Prob. density fn.: An alternative view

- \tilde{X} = temp. in a city
- Measure \tilde{X} in whole degrees: probs. = $p(x)$
- Measure \tilde{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.: $\text{Prob} [\tilde{X} \leq x] = F(x)$

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

Clicker question 5

- For a random variable \tilde{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 \tilde{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

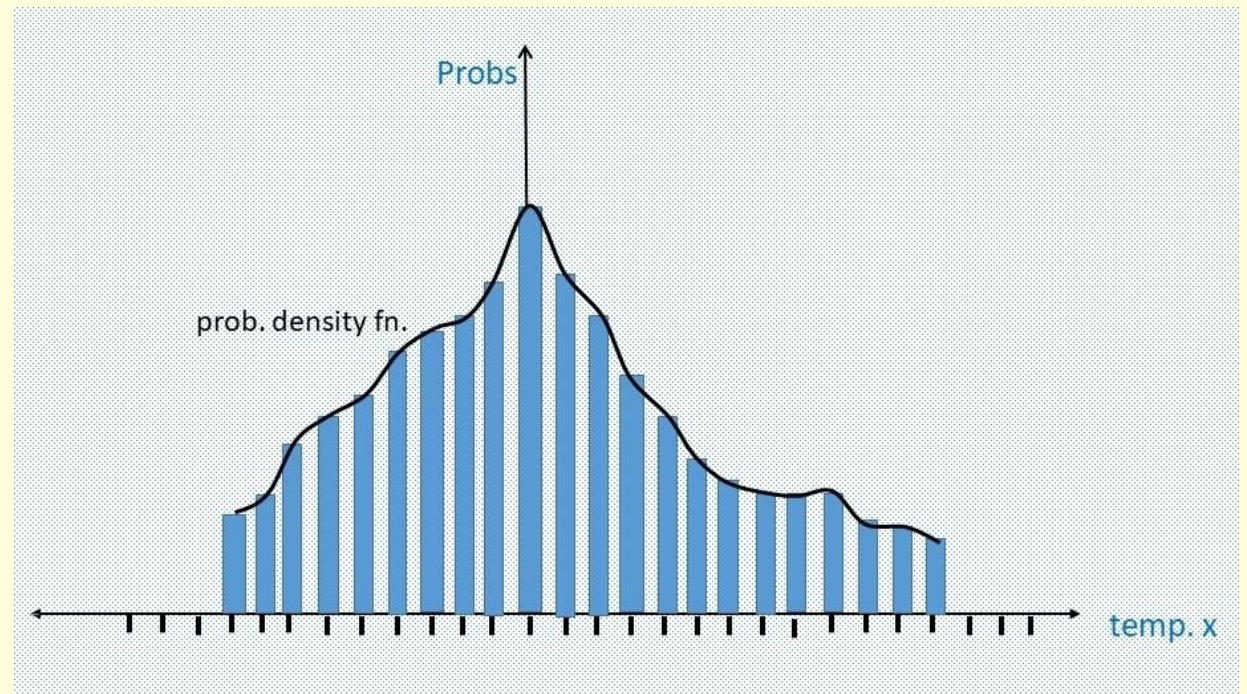
(b) $2x$

(c) x^2

(d) $2x^2$

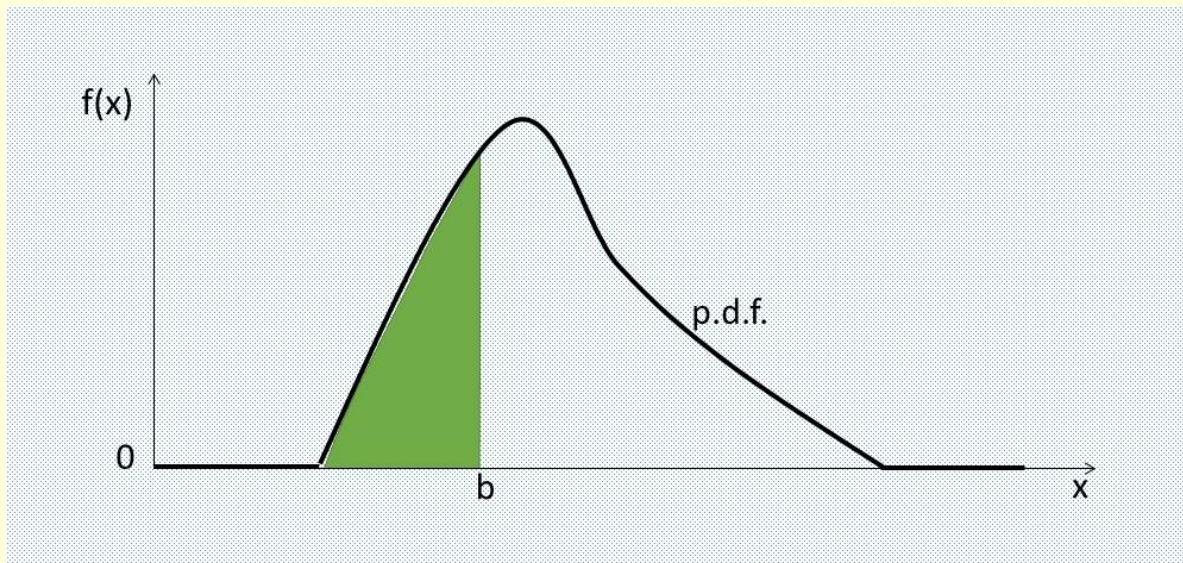
C.d.f. and p.d.f.: Connections

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



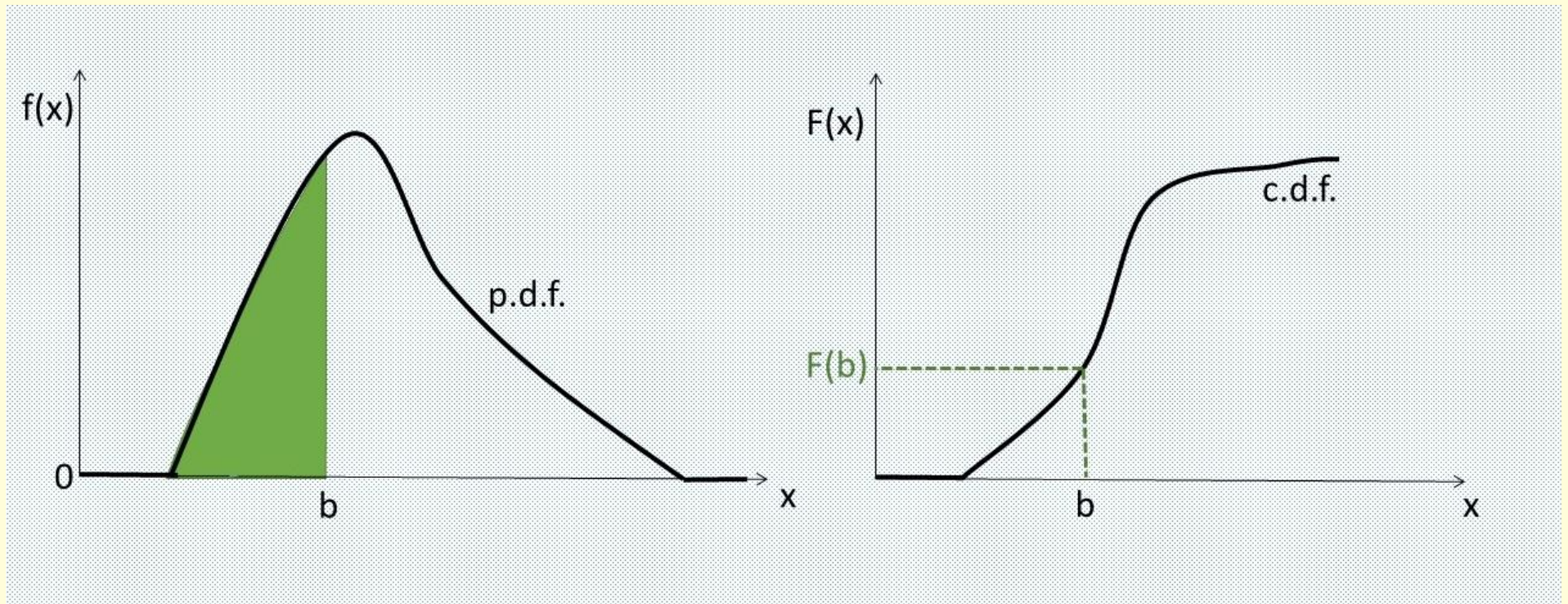
C.d.f. and p.d.f.: Connections

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



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

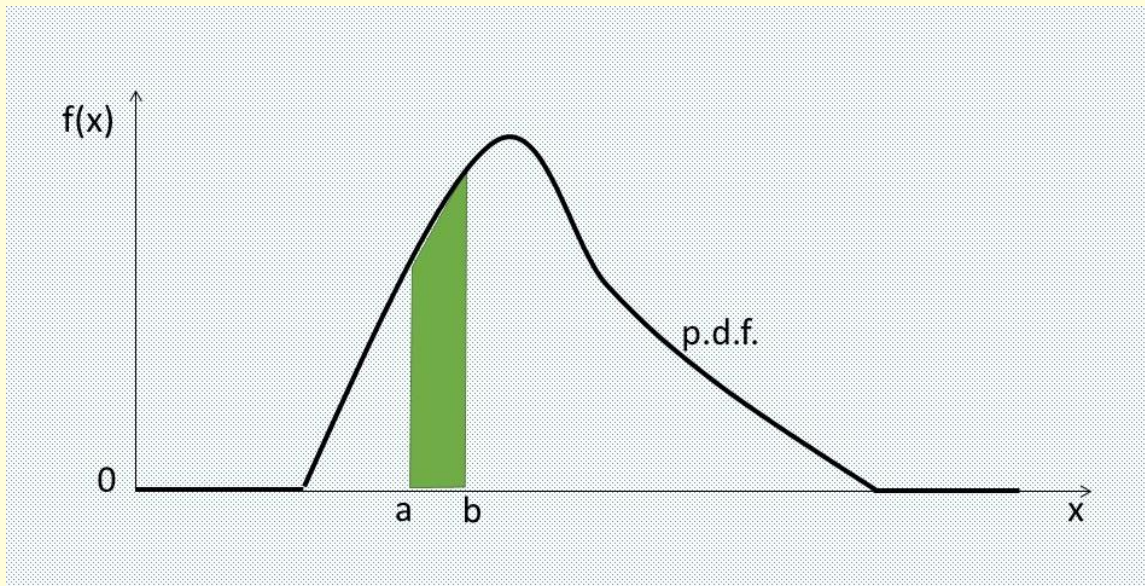
- $F(b) = \int_{-\infty}^b f(x) dx$



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

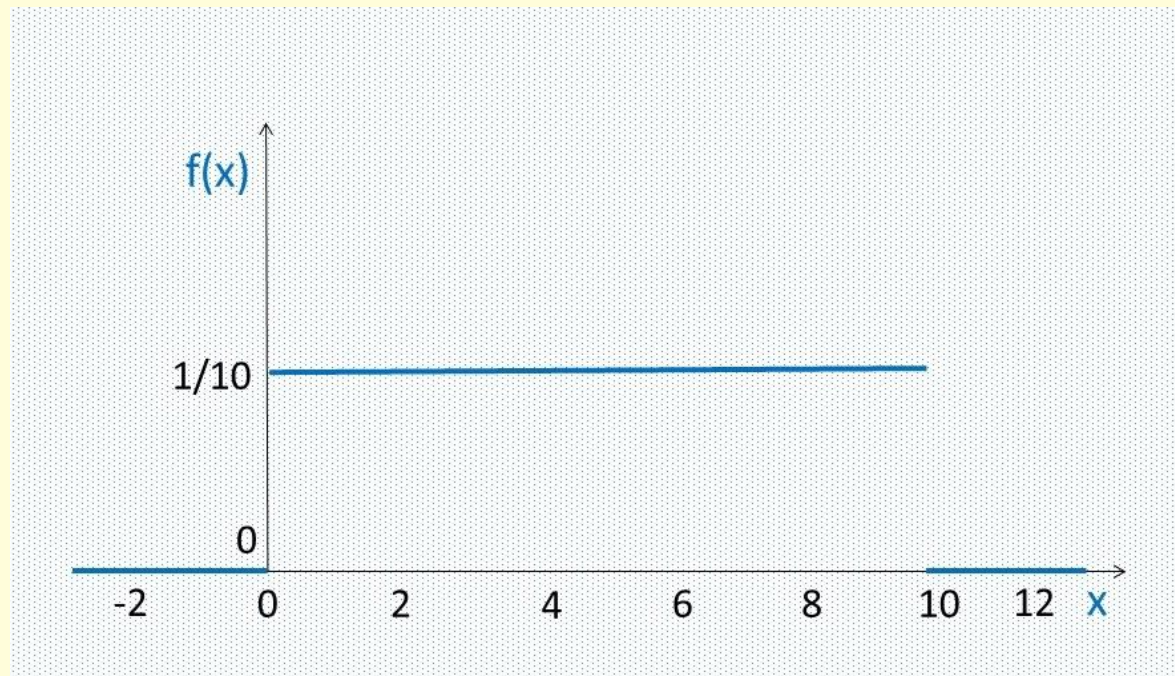
- $F(b) = \int_{-\infty}^b f(x) dx$
- **Prob** $\{\tilde{X} \in [a, b]\} = Prob[\tilde{X} \leq b] - Prob[\tilde{X} \leq a] =$

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



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

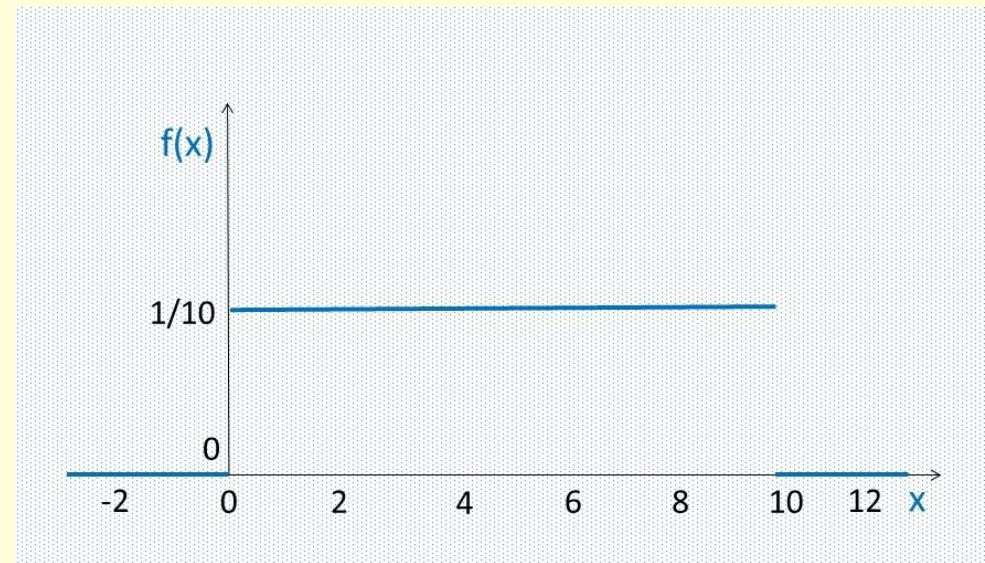
- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,
 $f(x) = 0$ elsewhere
- cdf $F(x)$?

Cdf and pdf: an example

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

Cdf and pdf: an example

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



Cdf and pdf: an example

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

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 $[\tilde{X} \leq 5] = F(5) =$*

Cdf and pdf: an example

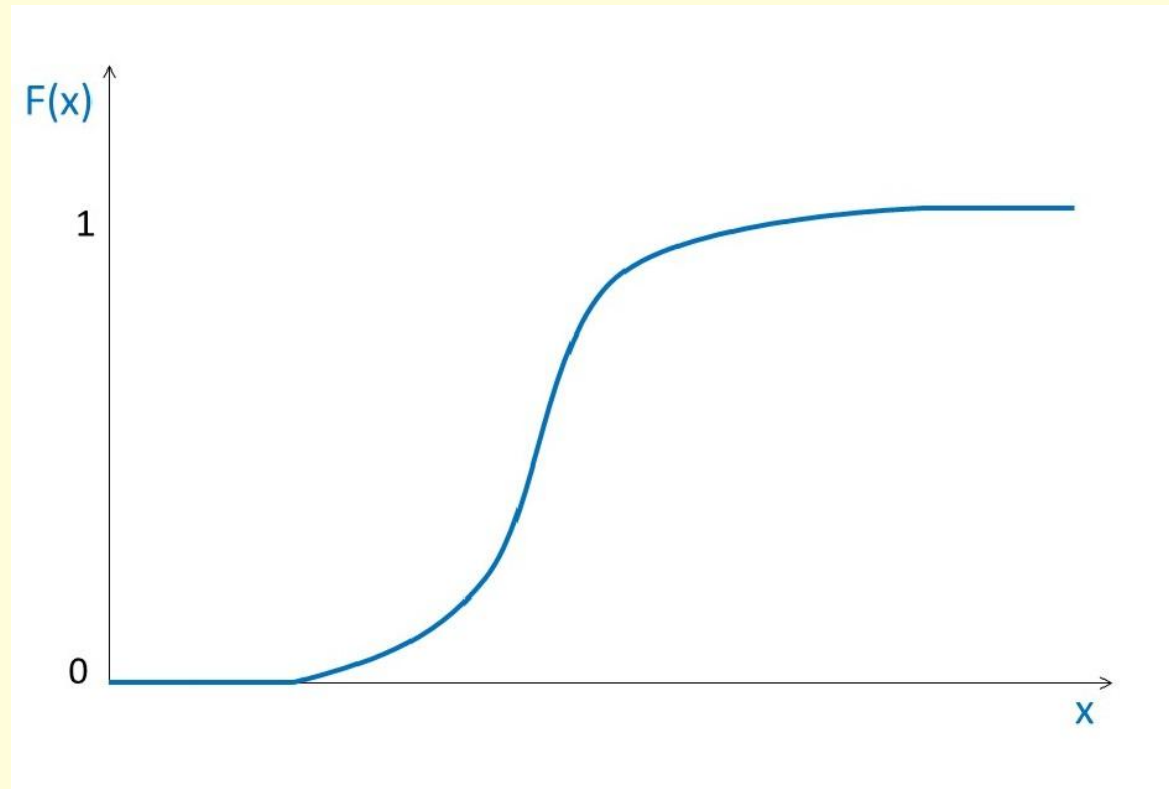
- *Uniform distribution pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,
 $f(x) = 0$ elsewhere
- cdf $F(b) = \int_{-\infty}^b f(x) dx =$
- Prob $[\tilde{X} \leq 5] = F(5) =$
- Prob $\{\tilde{X} \in [3, 5]\} = F(5) - F(3) =$

P.d.f.: properties

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

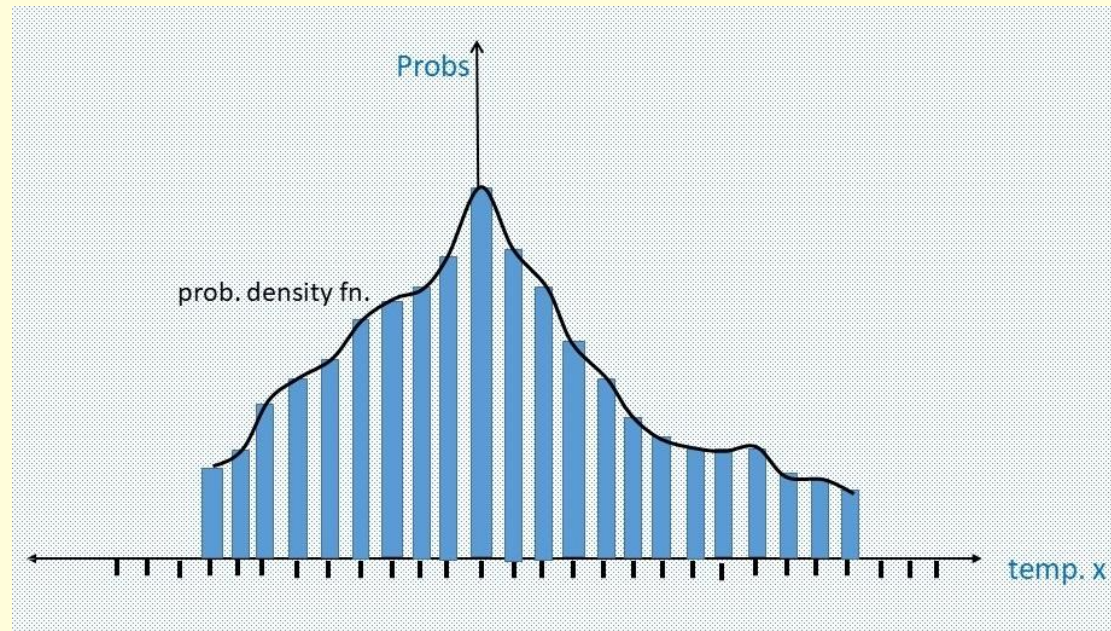
P.d.f.: properties

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

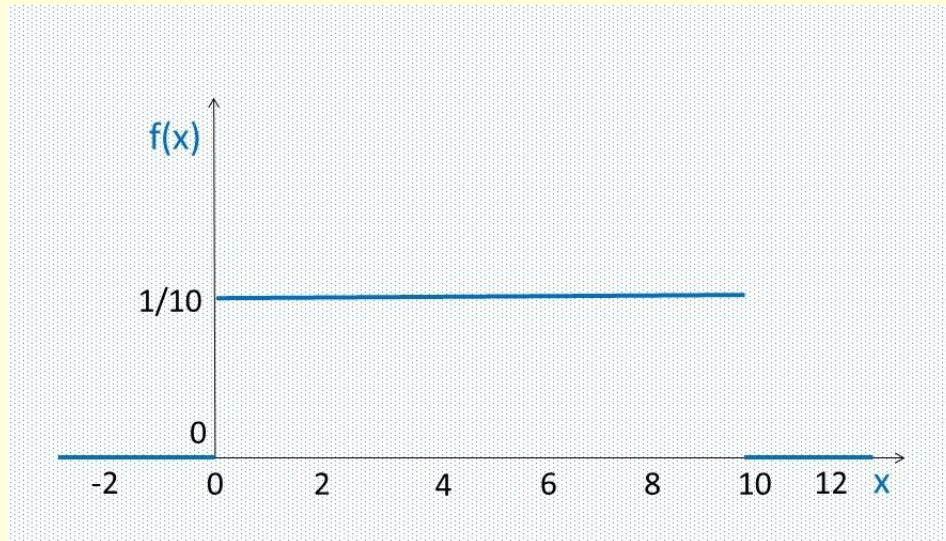


P.d.f.: properties

- 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 = $\int_{-\infty}^{\infty} f(x) dx = 1$
- Example: Uniform distn pdf: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,
 $f(x) = 0$ elsewhere

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 = $\int_{-\infty}^{\infty} f(x)dx = 1$
- Example: *Uniform distn pdf*: $f(x) = \frac{1}{10}$ **for** $x \in [0, 10]$,
 $f(x) = 0$ elsewhere
- Total area =



Expectation and Variance

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

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

Expectation and Variance

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

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

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

Expectation and Variance

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

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

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

- **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 \tilde{X} :

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

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

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

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

Expectation and Variance: An example

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

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

Expectation and Variance: An example

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

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

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

Expectation and Variance: An example

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

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

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

- Variance of $\tilde{X} = \sigma_X^2 = E(X^2) - (EX)^2$

Common Continuous Distributions 1: Uniform Distn.

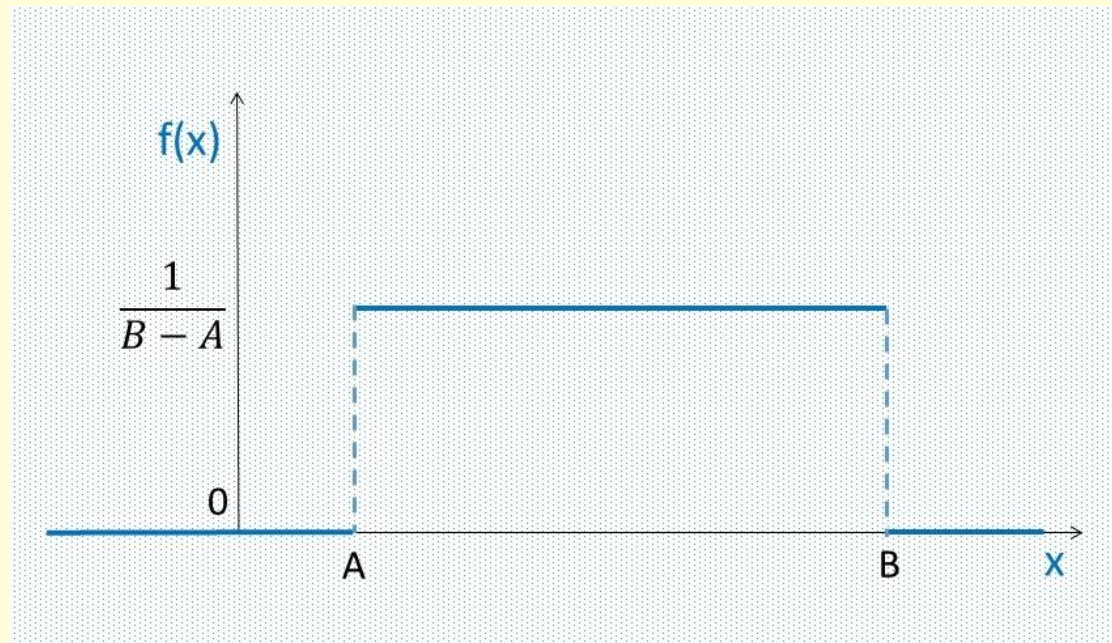
- \tilde{X} distributed over an interval $[A, B]$

Common Continuous Distributions 1: Uniform Distn.

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

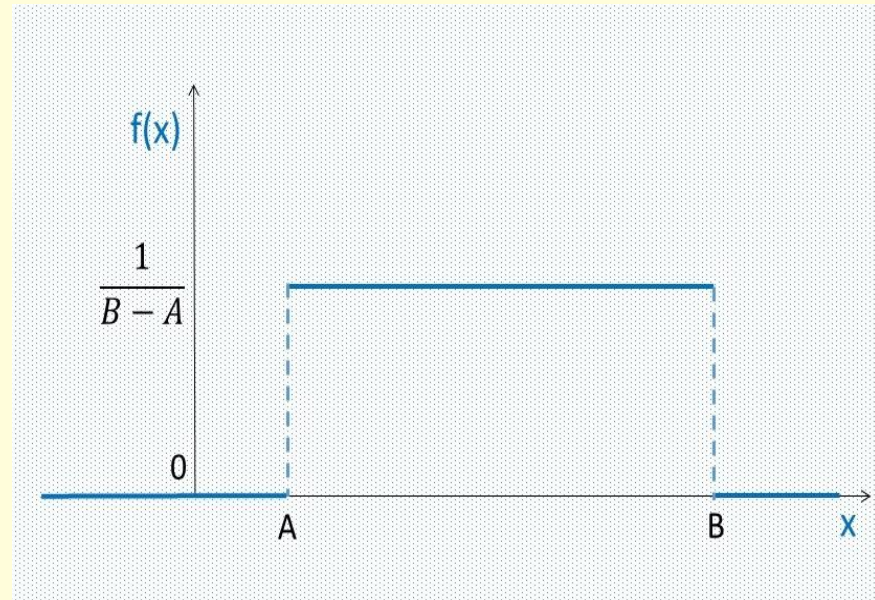
Common Continuous Distributions 1: Uniform Distn.

- \tilde{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.

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



Common Continuous Distributions 1: Uniform Distn.

- \tilde{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. \tilde{X} = its price at the end of the day.
Uniformly distributed over $[10, 50]$

Uniform Distribution: Example

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

Uniform Distribution: Example

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

Uniform Distribution: Example

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

Clicker question 6

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

(a) 100
(b) 500
(c) 550
(d) 600

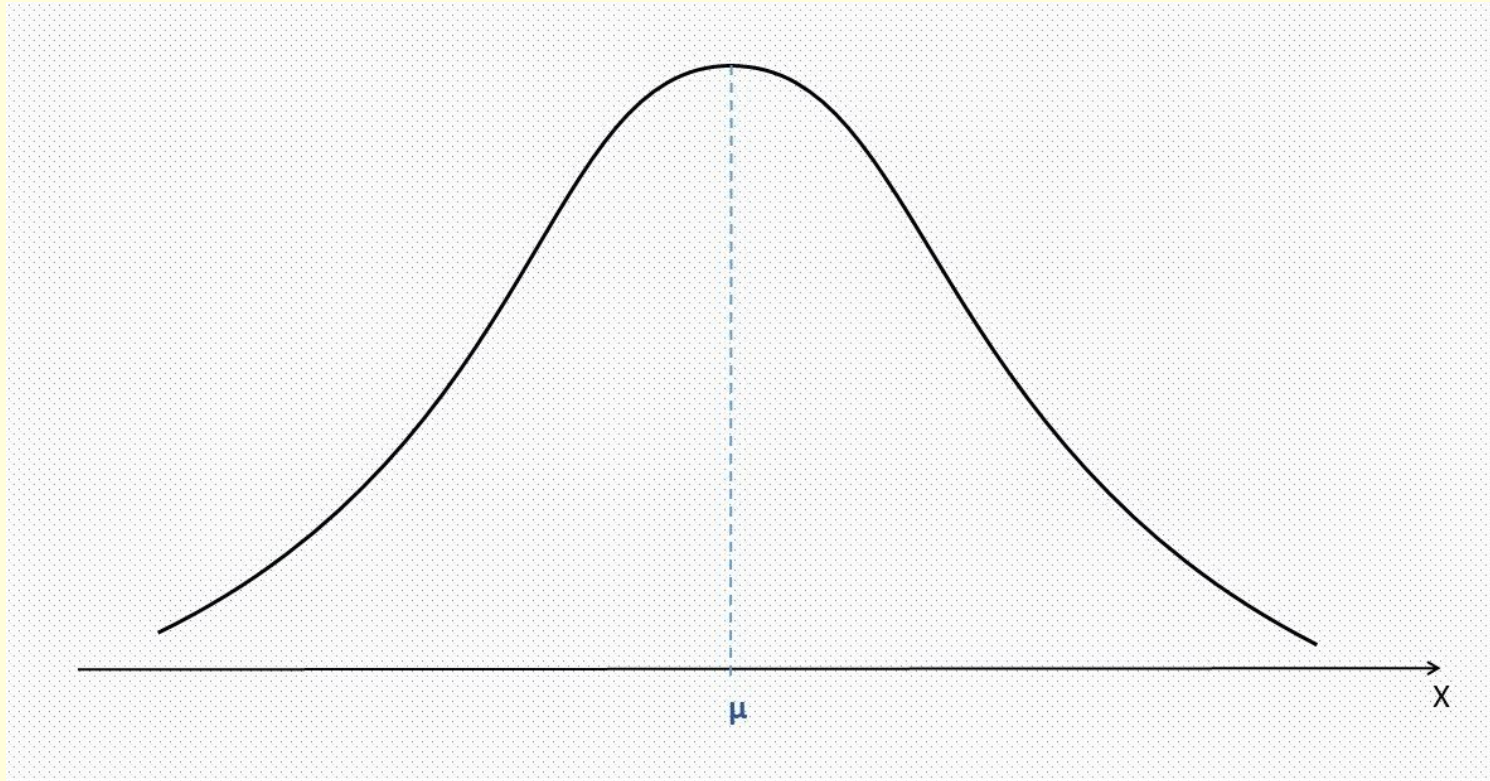
Solution to clicker question 6

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

(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.....



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
- $\tilde{X} \sim N(\mu, \sigma^2)$: *p.d.f:*

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

Common Continuous Distributions 2: Normal Distn.

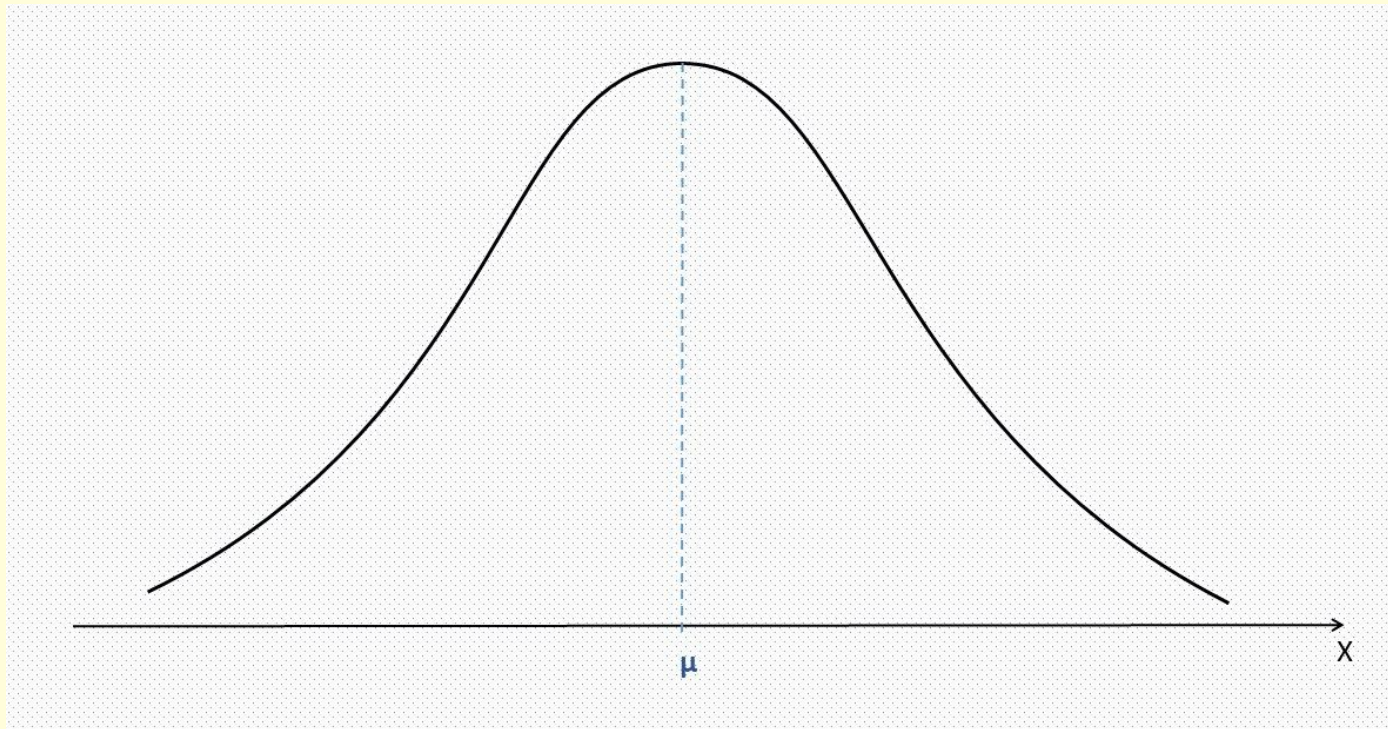
- *Most commonly used distribution:* Test scores, weight, height, IQ, stock returns.....
- Has 2 parameters: *mean* μ , *variance* σ^2
- $\tilde{X} \sim N(\mu, \sigma^2)$: *p.d.f:*

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

- *cdf* $F(b) = \int_{-\infty}^b f(x)dx$: Use software or prob. tables
Excel : = *NORM.DIST*(*x*, *mean*, *std.dev.*, *TRUE*)
GoogleSheets : *NORMDIST*

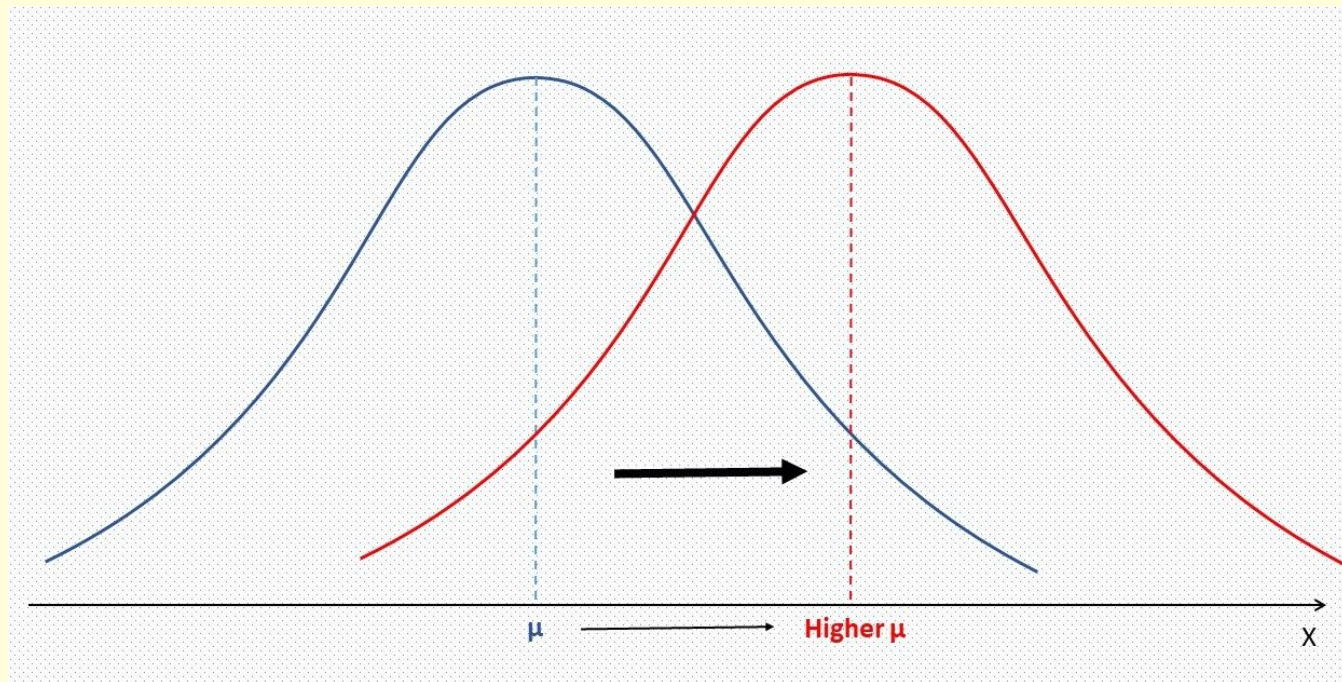
Normal Distn.: Shape

- Symmetric around mean μ :



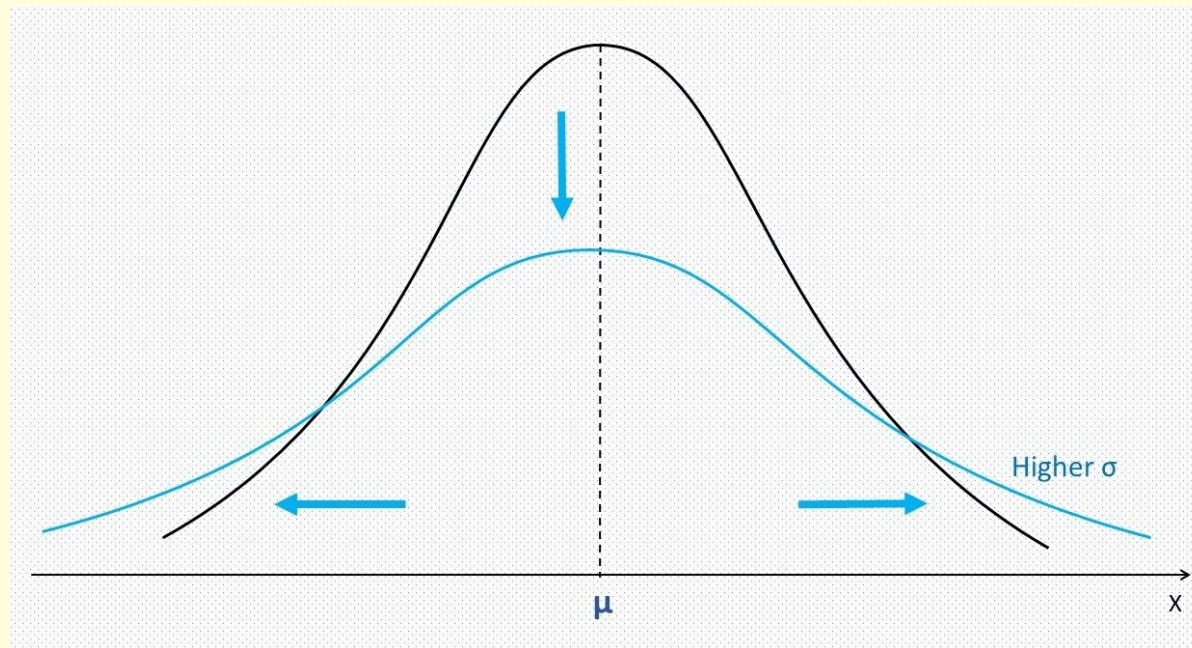
Normal Distn.: Shape

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



Normal Distn.: Shape

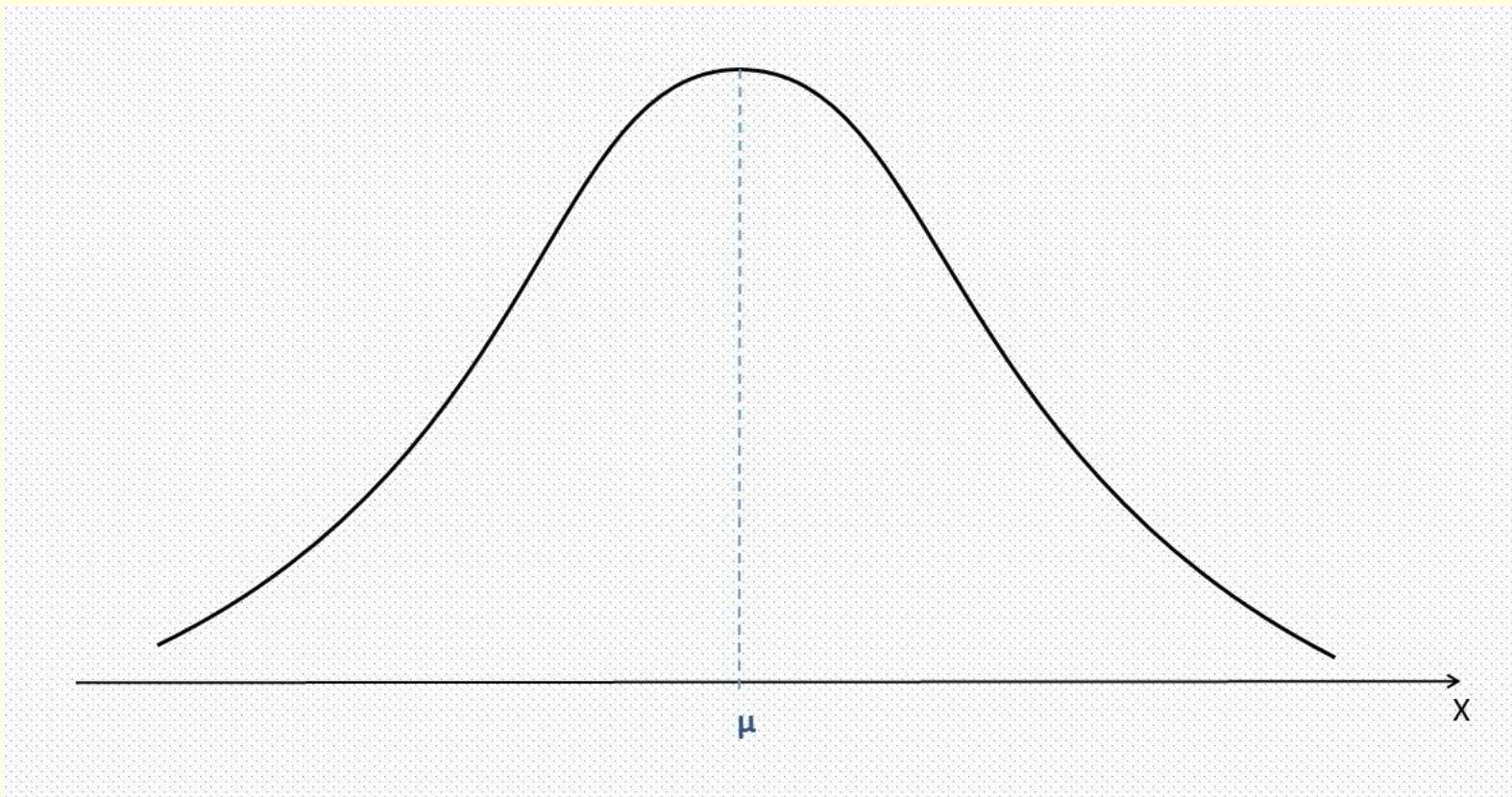
- Symmetric around mean μ :
- As μ changes: move left or right
- As σ changes: flatten or become steeper



Normal Distn. Properties: 68 - 95 - 99.7 rule

- *Symmetric around mean μ :*

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

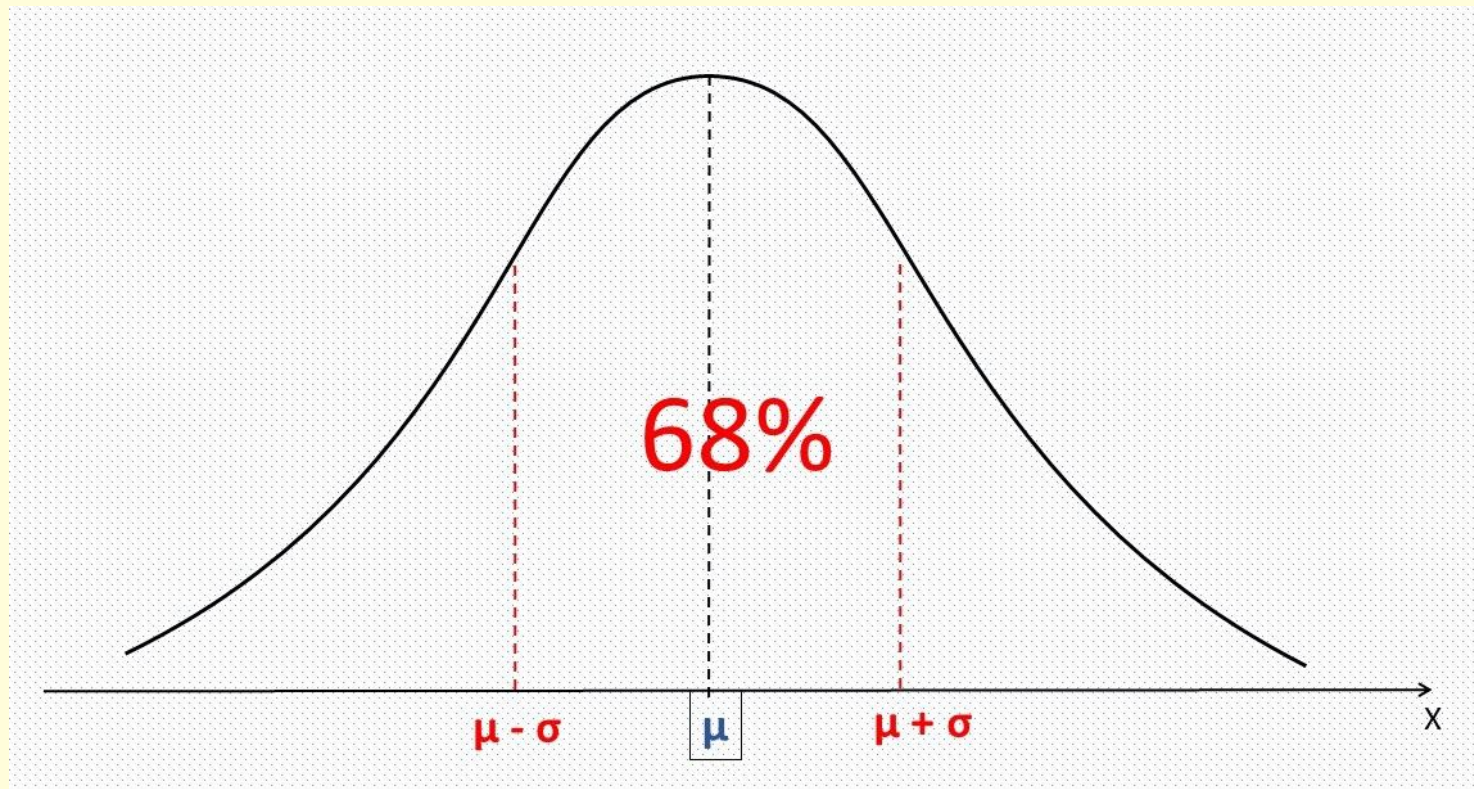


Normal Distn. Properties: 68 - 95 - 99.7 rule

- *Symmetric around mean μ :*

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

- $\text{Prob}[\mu - \sigma \leq \tilde{X} \leq \mu + \sigma] = 0.68$



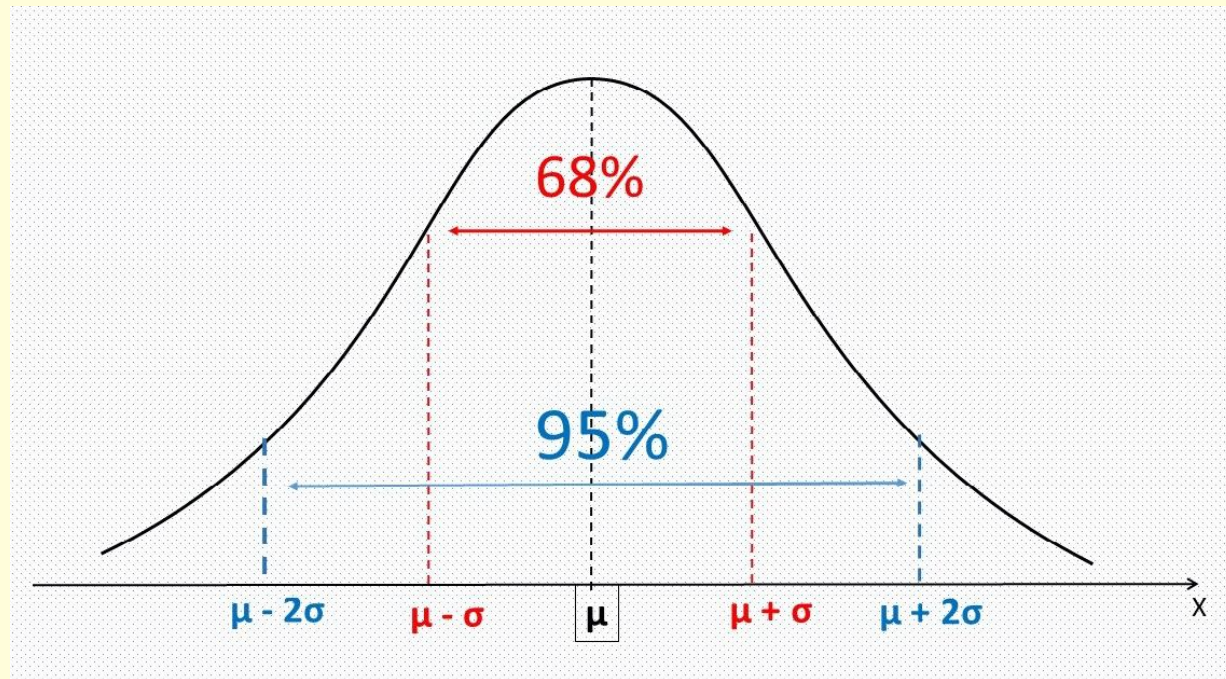
Normal Distn. Properties: 68 - 95 - 99.7 rule

- *Symmetric around mean μ :*

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

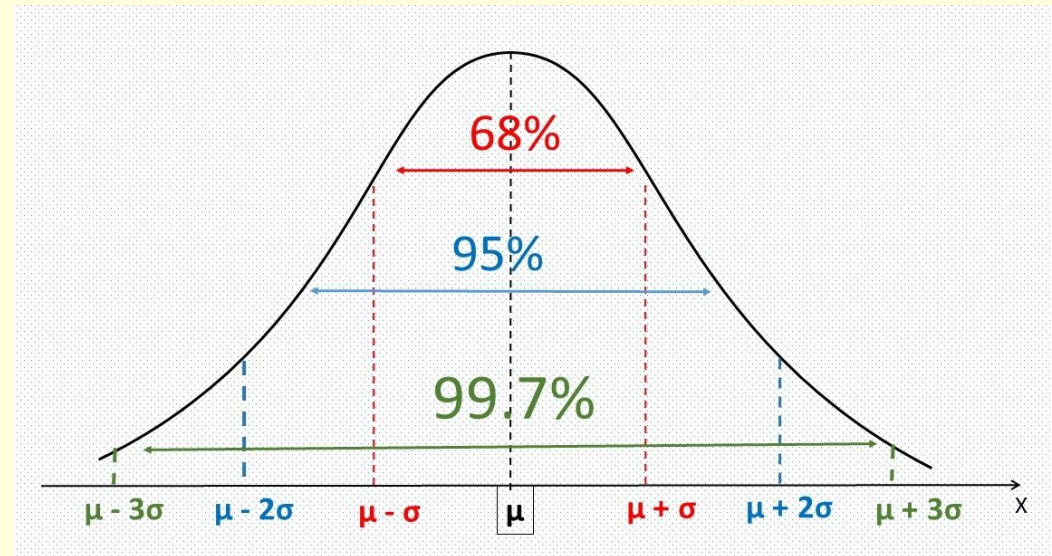
- $Prob[\mu - \sigma \leq \tilde{X} \leq \mu + \sigma] = 0.68$

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



Normal Distn. Properties: 68 - 95 - 99.7 rule

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



Normal Distn. Properties: 68 - 95 - 99.7 rule

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

Normal Distn. An example

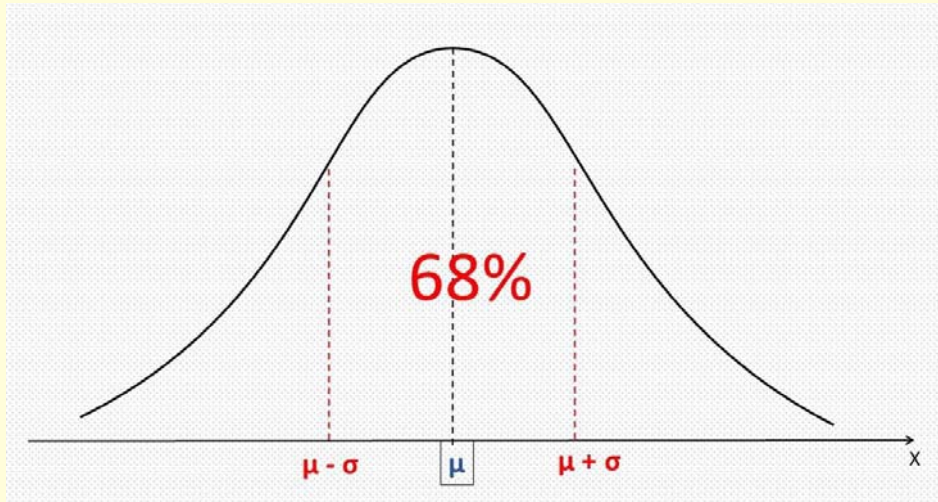
- Suppose a Doordash driver's delivery time, \tilde{t} , is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$

Normal Distn. An example

- Suppose a Doordash driver's delivery time, \tilde{t} , is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$
- $Prob[16 \leq \tilde{t} \leq 24] =$
- $Prob[12 \leq \tilde{t} \leq 28] =$
- $Prob[8 \leq \tilde{t} \leq 32] =$

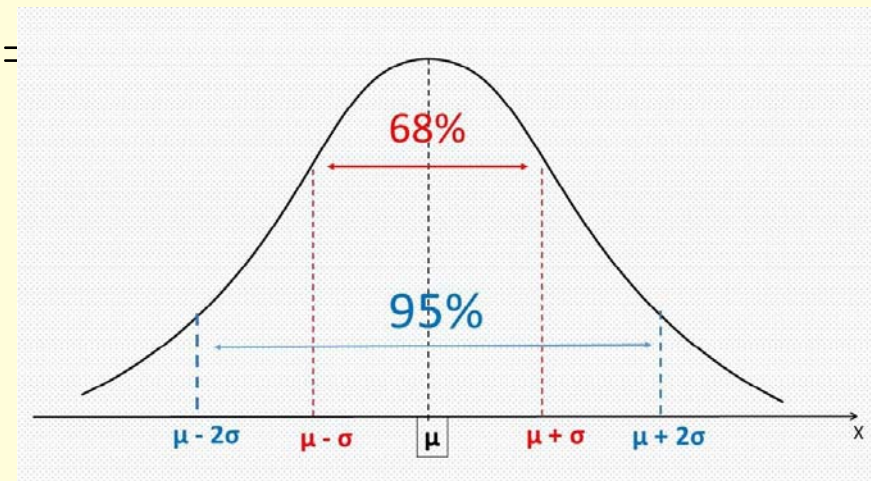
Normal Distn. An example

- 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?: $Prob[\tilde{t} \leq 24] =$



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?: $Prob[\tilde{t} \leq 24] =$
- $Prob[12 \leq \tilde{t} \leq 24] =$



Normal Distn. An example

- Suppose a Doordash driver's delivery time, \tilde{t} , is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$
- $Prob[16 \leq \tilde{t} \leq 24] = 0.68, \dots$
- Other probabilities?: $Prob[\tilde{t} \leq 22] = F(22)$
 $Prob[22 \leq \tilde{X} \leq 26] = F(26) - F(22) =$

Normal Distn. An example

- Suppose a Doordash driver's delivery time, \tilde{t} , is normally distributed with $\mu = 20$ minutes, and $\sigma = 4$
- $Prob[16 \leq \tilde{t} \leq 24] = 0.68, \dots$
- Other probabilities?: $Prob[\tilde{t} \leq 22] = F(22)$
 $Prob[22 \leq \tilde{X} \leq 26] = F(26) - F(22) =$
- Use Excel : = *NORM.DIST*(*x*, *mean*, *std.dev.*, *TRUE*)
GoogleSheets : *NORMDIST*

Clicker question 7

- Suppose a baby's birth weight, \tilde{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, \tilde{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



Normal Distribution: Properties

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

Normal Distribution: Properties

- If $\tilde{X} \sim N(\mu, \sigma^2)$, then $a\tilde{X} + b \sim N(a\mu + b, a^2\sigma^2)$
- *Example:* If $\tilde{X} \sim N(3.5, 0.25)$, then
 $2\tilde{X} + 10 \sim N(\quad , \quad)$

Normal Distribution: Properties

- If $\tilde{X} \sim N(\mu, \sigma^2)$, then $a\tilde{X} + b \sim N(a\mu + b, a^2\sigma^2)$
- *Example:* If $\tilde{X} \sim N(3.5, 0.25)$, then
 $2\tilde{X} + 10 \sim N(\quad , \quad)$
- Sum: If $\tilde{X} \sim N(\mu_X, \sigma_X^2)$ and $\tilde{Y} \sim N(\mu_Y, \sigma_Y^2)$ and \tilde{X} and \tilde{Y} are independent,
 $\tilde{Z} = \tilde{X} + \tilde{Y} \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Normal Distribution: Properties

- If $\tilde{X} \sim N(\mu, \sigma^2)$, then $a\tilde{X} + b \sim N(a\mu + b, a^2\sigma^2)$
- *Example:* If $\tilde{X} \sim N(3.5, 0.25)$, then
 $2\tilde{X} + 10 \sim N(\quad , \quad)$
- Sum: If $\tilde{X} \sim N(\mu_X, \sigma_X^2)$ and $\tilde{Y} \sim N(\mu_Y, \sigma_Y^2)$ and \tilde{X} and \tilde{Y} are independent,
 $\tilde{Z} = \tilde{X} + \tilde{Y} \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
- *Example:* If husband's income $\sim N(2000, 100)$, and wife's income $\sim N(2500, 200)$, then
Joint income $\sim N(\quad , \quad)$



Standard Normal Distribution

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

Standard Normal Distribution

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

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

Standard Normal Distribution

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

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

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

Standardization: An example

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

Standardization: An example

- *Standardization:* If $\tilde{X} \sim N(\mu, \sigma^2)$, 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 $\tilde{X} \sim N(\mu, \sigma^2)$, 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 =
Tam's Z-score =



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* $[\tilde{X} \leq 75] = F(75) =$

Use 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 $\sigma = 10$.

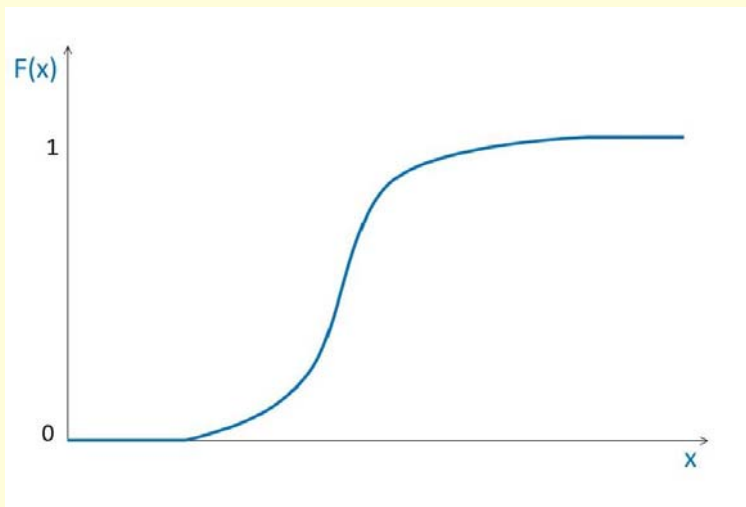
Q: What fraction of the class did worse than her?

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

Use Excel : = *NORM.DIST*(*x*, *mean*, *std.dev.*, *TRUE*)

- What is the 90th 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?

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

Use Excel : = *NORM.DIST*(x , *mean*, *std.dev.*, *TRUE*)

- What is the 90th percentile score?

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

- Use Excel : = *NORM.INV*(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?

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

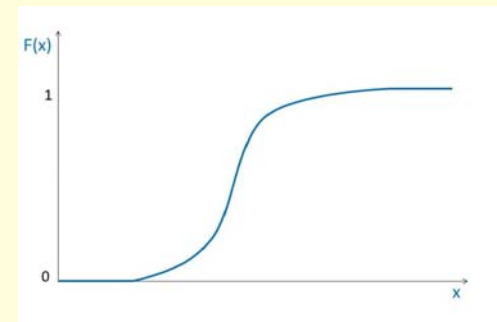
Use Excel : = *NORM.DIST*(x , *mean*, *std.dev.*, *TRUE*)

- What is the 90th percentile score?

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

- Use Excel : = *NORM.INV*(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?

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

Use Excel : = *NORM.DIST*(*x*, *mean*, *std.dev.*, *TRUE*)

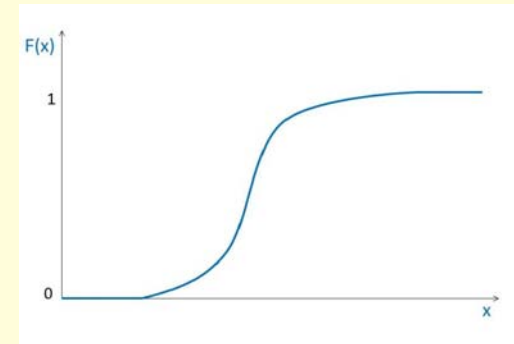
- What is the 90th percentile score?

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

- Use Excel : = *NORM.INV*(*p*, *mean*, *std.dev.*)

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

- 50th percentile: **median** i.e x s.t. $F(x) = 0.5$



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,.....