

# Unit 3: Combinatorics, Probability, and Statistics

Charles Rambo

UCLA Anderson

2024

# Table of Contents I

## 1 Combinatorics

- Permutations and Combinations
- Stars-and-Bars
- The Pigeonhole Principle
- Inclusion–Exclusion Principle

## 2 Probability

- Conditional Probabilities
- Random Variables
- Discrete Random Variables
- Continuous Random Variables

# Table of Contents II

- Expected Values
- Moments
- Functions of a Random Variable
- Joint Distributions
- Covariance
- Efficient Frontier

## 3 Statistics

- Inequalities
- Law of Large Numbers
- Empirical CDFs

# Table of Contents III

- Central Limit Theorem
- Hypothesis Testing

# Combinatorics

# Combinatorics

## Definition

**Combinatorics** is the branch of mathematics concerned with counting.

The fundamental counting principle states:

*If there are  $m_k$  ways to make the  $k$ -th decision for  $k = 1, 2, \dots, n$ ,  
then there are a total of*

$$\prod_{k=1}^n m_k = m_1 \cdot m_2 \cdot \dots \cdot m_n$$

*ways to make the  $n$  independent decisions.*

# Combinatorics Example

## Example

Let's say that the personal code for your bank pin consists of six characters, each of which can be any letter or numerical digit.

$$26 \quad 10 \quad \Rightarrow \quad 36 \text{ options for each of the six characters}$$

- (a) How many possible personal codes are there?

$$36^6$$

- (b) Suppose you want your pin to have no repeated characters. How many possible pins would there be now?

$$36 \cdot 35 \cdot 34 \cdot 33 \cdot 32 \cdot 31$$

# Permutations and Combinations

## Definition

If the order in which the objects are chosen matters, each complete selection of  $k$  objects from an original  $n$  is called a **permutation**. If the order does not matter, then each complete selection is called a **combination**.

# Permutations and Combinations without Repetition

The number of *permutations* of  $k$  elements from a selection of  $n$  *without repetition* is

$${}_n P_k = \frac{n!}{(n - k)!}.$$

The number of *combinations* of  $k$  elements from a selection of  $n$  *without repetition* is

$${}_n C_k = \binom{n}{k} = \frac{n!}{k!(n - k)!}.$$

# Permutations and Combinations without Repetition

## Example

- (a) Among eight comparable investment funds, *ex ante*, how many different ways are there to list the top three performing funds?
- (b) How many three-element subsets does a set of nine elements have?

Sol

(a) The order in which the funds are listed is important

$\Rightarrow$  Permutation

$$n P_k = \frac{n!}{(n-k)!} = \frac{8!}{(8-3)!} = \frac{8!}{5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{5!} = 336$$

(b) In sets, the order doesn't matter.  $\Rightarrow$  Combination

$$n C_k = \frac{n!}{k!(n-k)!} = \frac{9!}{3! 6!} = \frac{9 \cdot 8 \cdot 7 \cdot 6!}{3! \cdot 6!} = \frac{9 \cdot 8 \cdot 7}{6!} = 84$$

# $n$ Elements of $k$ Types

Suppose there are  $n_k$  indistinguishable elements of type  $k$ , where  $k = 1, 2, \dots, m$ . Let

$$n = n_1 + n_2 + \dots + n_m.$$

Then the number of *unique* rearrangements of the  $n$  elements is

$$\binom{n}{n_1, n_2, \dots, n_m} = \frac{n!}{n_1! n_2! \dots n_m!}.$$

# $n$ Elements of $k$ Types

## Example

How many unique rearrangements of the letters in the word BANANA are there?

Sol

# of B's: 1

# of A's: 3

# of N's: 2

Total of six letters

Hence, the number of unique rearrangements  
is

$$\frac{6!}{1! \cdot 3! \cdot 2!}$$
$$= \frac{\cancel{6}^3 \cdot 5 \cdot 4 \cdot \cancel{3}^2}{1! \cdot \cancel{3}^1 \cdot \cancel{2}^1}$$

$= 60$

# Stars-and-Bars

## Theorem (Stars-and-Bars)

*Suppose there are  $k$  types of items. If the order in which items are selected does not matter, the number of ways to select  $n$  items is*

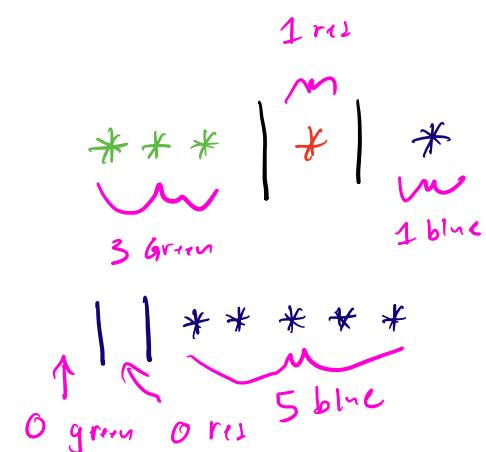
$$\binom{n+k-1}{k-1} = \binom{n+k-1}{n}.$$

$k$  types ← bars for the  $k$  types ← we  $k-1$  bars to separate  $k$  types

$n$  Selections ← stars for the section

Types: Green, red, blue

We can select 5



How to think about this?

$$3-1 + 5 = 7$$

Places. How many ways can we select  $3-1=2$  bars or

5 stars

$$\binom{7}{3-1} = \binom{7}{5}$$

In general, we have  $k$  types and we select  $n$ . The order in which we select elements doesn't matter.

$\Rightarrow k-1$  bars

$\Rightarrow n+k-1$  places

How many ways to select  $k-1$  bars or  $n$  stars?

$$\boxed{\binom{n+k-1}{k-1} = \binom{n+k-1}{n}}.$$

# Stars-and-Bars Example

## Example

In how many ways can we write the number 11 as the sum of five positive integers?

Sol

$$\square^{\cancel{1}} + \square^{\cancel{1}} + \square^{\cancel{1}} + \square^{\cancel{1}} + \square^{\cancel{1}} = 11$$

$$\rightarrow \square^{\cancel{0}} + \square^{\cancel{0}} + \square^{\cancel{0}} + \square^{\cancel{0}} + \square^{\cancel{0}} = 6 \quad \leftarrow \text{The amount that is free to be in and from}$$

5 types, 6 Selection  $\rightarrow$  4 bars and 6 stars

$$\binom{6+4}{4} = \binom{10}{4} = \binom{10}{6} = \frac{10!}{4! 6!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3 \cdot 4} = 210$$

Oops! I used the permutation formula during lecture!!

# Stars-and-Bars on YouTube

Watch Ken Ribet explain Stars-and-Bars (<https://youtu.be/UTCScjoPymA>).



Stars and Bars (and bagels) - Numberphile



Numberphile  
4.32M subscribers



Subscribed



5.3K



1



Share



Download



# The Pigeonhole Principle

## Theorem (Pigeonhole Principle)

Suppose there are  $\underline{n}$  objects distributed among  $\underline{m}$  sets, where  $n \geq m$ .

Then one set must contain at least  $\lceil \underline{n/m} \rceil$  elements.

# The Pigeonhole Principle Example

## Example

Suppose we have a set of  $n + 1$  whole numbers. Show that we can always select two elements  $a$  and  $b$  such that  $a - b$  is divisible by  $n$ .

Sol  $n + 1$  elements Note: There are  $n$  remainders when you divide a number by  $n$ ,  $0, 1, 2, 3, \dots, n-1$ .

There are  $n + 1$  elements. So, there must

$$\lceil \frac{n+1}{n} \rceil = \lceil 1 + \frac{1}{n} \rceil = 2$$

We know there must be at least two elements of the same remainder. If two elements have the same remainder, their difference has a remainder of  $0$ . Hence, their difference is divisible by  $n$ .

n=3

Reminder:

$$\begin{array}{r} 2, 5 \\ 2 \end{array}, \quad \begin{array}{r} 4, 7 \\ 1 \end{array}$$

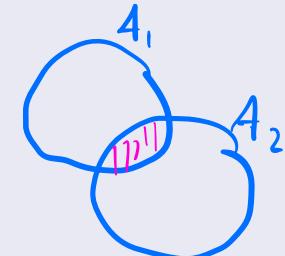
$$5 - 2 = 3$$

$$7 - 4 = 3$$

# Inclusion–Exclusion Principle

## Theorem (Inclusion–Exclusion Principle)

Suppose  $|A|$  denotes the number of elements in  $A$ . Consider finite sets  $A_1$  and  $A_2$ . The number of elements in  $A_1 \cup A_2$  is

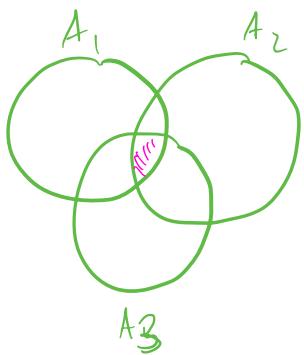


$$\underbrace{|A_1 \cup A_2|}_{=} = \underbrace{|A_1|}_{=} + \underbrace{|A_2|}_{=} - \underbrace{|A_1 \cap A_2|}_{=}$$

More generally, for  $A_1, A_2, \dots, A_n$ , the number of elements in  $\bigcup_{i=1}^n A_i$  is

$$\left| \bigcup_{i=1}^n A_i \right| = \sum_{i=1}^n |A_i| - \sum_{1 \leq i < j \leq n} |A_i \cap A_j| + \sum_{1 \leq i < j < k \leq n} |A_i \cap A_j \cap A_k| - \dots + (-1)^{n+1} |A_1 \cap A_2 \cap \dots \cap A_n|.$$

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$



# Inclusion–Exclusion Principle Example

## Example

There are 100 students enrolled in the UCLA MFE class of 2025. Sixty-five students are taking statistical arbitrage, 43 are taking credit markets, and 37 are taking behavioral finance. If 23 students are taking statistical arbitrage and credit markets, 15 are taking statistical arbitrage and behavioral finance, and 20 are taking credit markets and behavioral finance, how many students are taking all three classes?

Sol

$A_1$  = Students in stat. arb.

$A_2$  = Students in credit markets

$A_3$  = Students in behavioral finance

We want to find

$$|A_1 \cap A_2 \cap A_3|$$

i.e. # Students taking all 3.

$$\begin{aligned} |A_1 \cup A_2 \cup A_3| &= |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| \\ &\quad + |A_1 \cap A_2 \cap A_3| \end{aligned}$$

$$\Rightarrow 100 = 65 + 43 + 37 - 23 - 15 - 20 + x$$

$$\Rightarrow 100 = 89 + x$$

$$\Rightarrow x = 11 \rightarrow \text{Eleven students are taking all three classes!}$$

# Probability

# $\sigma$ -Algebra

Set of all subsets  
of  $\Omega$ . People denote it  $2^{\Omega}$  because for  $\Omega$  finite,  
then the set of all subsets has  $2^{|\Omega|}$  elements.

## Definition

Let  $\Omega$  be some set, and let  $2^{\Omega}$  represent the set of all subsets of  $\Omega$ . Then a subset  $\mathcal{F}$  of  $2^{\Omega}$  is called a  **$\sigma$ -algebra** if and only if it satisfies the following three properties:

SA.1  $\Omega$  is in  $\mathcal{F}$ .

SA.2 If  $A \in \mathcal{F}$  then  $A^c = \{x \in \Omega : x \notin A\} \in \mathcal{F}$ .

SA.3 If  $A_k \in \mathcal{F}$  for  $k = 1, 2, \dots$ , then  $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ .

# Probability Definition

## Definition

Let  $\Omega$  be a set and let  $\mathcal{F}$  be a  $\sigma$ -algebra of  $\Omega$ . A **probability** is defined as a function  $P : \mathcal{F} \rightarrow [0, 1]$  if the following hold:

P.1  $P(\emptyset) = 0.$

P.2  $P(\Omega) = 1.$

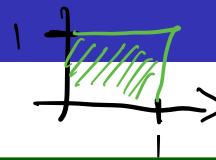
P.3 For  $E_k \in \mathcal{F}$  and  $E_k \cap E_\ell = \emptyset$  for  $k \neq \ell$ , we have

$$P\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} P(E_k).$$

*Sets are disjoint*

We typically call  $\Omega$  the **sample space**. The elements of  $\mathcal{F}$  are called **events**. Altogether,  $(\Omega, \mathcal{F}, P)$  forms a **probability space**.

# Probability Example



## Example

Consider the sample space  $\underline{[0, 1]} \times \underline{[0, 1]}$ . We define  $P : \mathcal{F} \rightarrow [0, 1]$  such that  $\underline{P}(E) = \text{area}(E)$ . In this cases, we would have

$$\underline{\mathcal{F}} = \{E \subseteq \underline{[0, 1]} \times \underline{[0, 1]} : \text{area of } E \text{ exists}\}.$$

We need  $\mathcal{F}$  because there are subsets of  $[0, 1] \times [0, 1]$  whose area can't be measured! *← Math thing. It turns out existence is equivalent to the axiom of choice*

# Properties of Probability Measure

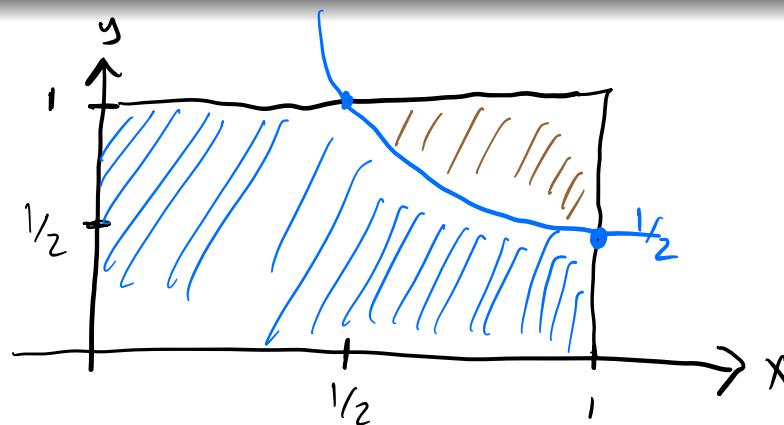
Consider the probability space  $(\Omega, \mathcal{F}, P)$ . Suppose  $A$  and  $B$  are in  $\mathcal{F}$ .

- $\underline{\underline{A}} \subseteq \underline{\underline{B}}$  implies  $P(\underline{\underline{A}}) \leq P(\underline{\underline{B}})$
- $P(\underline{\underline{A}^c}) = 1 - P(\underline{\underline{A}})$
- $P(\underline{\underline{A \cup B}}) = P(\underline{\underline{A}}) + P(\underline{\underline{B}}) - P(\underline{\underline{A \cap B}})$

# Probability Example

## Example

The point  $(x, y)$  is selected at random from  $[0, 1] \times [0, 1]$ . Assume that all points within the sample space are equally likely to be selected. What is the probability that  $xy$  is less than  $1/2$ ?



$$xy \leq \frac{1}{2}$$
$$y \leq \frac{1}{2x}$$

$$y=1 \quad y=\frac{1}{2x}$$
$$\downarrow \quad \downarrow$$

$$\begin{aligned} \text{Sol} \quad P(xy \leq \frac{1}{2}) &= 1 - P(xy > \frac{1}{2}) = 1 - \int_{\frac{1}{2}}^1 \left(1 - \frac{1}{2x}\right) dx \\ &= 1 - \left[x - \frac{1}{2} \ln x\right]_{\frac{1}{2}}^1 \end{aligned}$$

$$\begin{aligned}
 &= 1 - \left[ x - \frac{1}{2} \ln x \right]_{1/2}^1 \\
 &= 1 - \left[ 1 - \frac{1}{2} \ln 1 - \frac{1}{2} + \frac{1}{2} \ln \frac{1}{2} \right] \\
 &= 1 - \left( \frac{1}{2} - \frac{1}{2} \ln 2 \right) \\
 &= 1 - \frac{1}{2} + \frac{1}{2} \ln 2 \\
 &= \boxed{\frac{1}{2} + \frac{1}{2} \ln 2}
 \end{aligned}$$

$$\ln \frac{1}{2} = \ln 1 - \ln 2 \\ = -\ln 2$$

# Conditional Probabilities

## Definition

If the occurrence of the event  $A$  depends on the occurrence of  $B$  then the conditional probability is  $A \text{ given } B$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

whenever  $P(B) > 0$ . If  $P(B) = 0$  the conditional probability is undefined.

# Conditional Probability Example

Credit	Excellent	Good	Poor
Defaults	930	1,350	5,856
Non-Defaults	10,321	18,007	13,536

## Example

A loan originator has a 50,000 observation loan data set. This originator places applicants' credit into one of three categories: excellent, good, and poor. Details about this data set are shown above. What is the probability of default given that an applicant's credit is placed into the category good?

$$\begin{aligned} \text{Sol } P(\text{default} | \text{good}) &= \frac{P(\text{default and good})}{P(\text{good})} \\ &= \frac{\frac{1,350}{50,000} \times 50,000}{\frac{1,350 + 18,007}{50,000}} \\ &= \frac{1,350}{1,350 + 18,007} \end{aligned}$$

$\approx 6.97\%$

# Bayes' Rule

## Theorem (Bayes)

Suppose  $P(B) > 0$ . Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

## Theorem (Law of Total Probabilities)

Suppose  $B_i \cap B_j = \emptyset$  for  $i \neq j$  and  $\bigcup_{i=1}^n B_i = \Omega$ . Then

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n).$$

# Bayes' Rule Example

## Example

Assume the probability that an unskilled asset manager beats the market is 50%, and the probability that a skilled asset manager beats the market is 60%. If 10% of asset managers are skilled, what is the probability that an asset manager is skilled given that she beats the market?

Sol

$$\begin{aligned} P(\text{skilled} \mid \text{beats}) &= \frac{P(\text{beats} \mid \text{skilled}) P(\text{skilled})}{P(\text{beats})} \xleftarrow{\text{Bayes' Rule}} \\ &= \frac{P(\text{beats} \mid \text{skilled}) P(\text{skilled})}{P(\text{beats} \mid \text{skilled}) P(\text{skilled}) + P(\text{beats} \mid \text{unskilled}) P(\text{unskilled})} \xleftarrow{\text{use law of total probability}} \\ &= \frac{0.60 \times 0.10}{0.60 \times 0.10 + 0.50 \times 0.90} \end{aligned}$$

Assumption: either skilled or unskilled

$$\Rightarrow = \frac{0.06}{0.06 + 0.45}$$

$$= \frac{0.06}{0.51}$$

$$= \frac{6}{51}$$

$$\approx 11.76\%$$

Prob. of being skilled, given a beat is  
11.76%.

# Bayes' Rule on YouTube

Watch the StatQuest video about Bayes' theorem (a.k.a. Bayes' rule) on YouTube (<https://youtu.be/9wCnvr7Xw4E>).



## Bayes' Theorem: Clearly Explained!!!

A screenshot of a YouTube video player. The video title is "Bayes' Theorem: Clearly Explained!!!!". The player shows a progress bar at 0:12 / 13:59, a description "Awesome song and introduction &gt;", and standard video controls (play, volume, settings, etc.).

Bayes' Theorem, Clearly Explained!!!!



StatQuest with Jo...

952K subscribers

Join

Subscribed

5K



Share

Download

...



# Independent Events

## Definition

Events  $A$  and  $B$  are independent if  $P(A|B) = P(A)$ .

If two events are independent then it's easy to prove

$$P(A \cap B) = P(A)P(B).$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

↓  
indep.

$$P(A) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A)P(B)$$

# Independent Events Example

## Example

One urn contains four red balls and six blue balls. A second urn contains sixteen red balls and  $x$  blue balls. A single ball is drawn from each urn. The probability that both balls are the same color is 0.44. Calculate  $x$ .

Sol Draws from urns are indep.

$$\begin{aligned} P(C_1 = C_2) &= P(C_1 = c_2 = \text{red}) + P(C_1 = c_2 = \text{blue}) \\ &\stackrel{\substack{\uparrow \\ \text{color} \\ \text{of ball} \\ \text{in first} \\ \text{urn}}}{=} P(C_1 = \text{red}) P(C_2 = \text{red}) + P(C_1 = \text{blue}) P(C_2 = \text{blue}) \\ &= \frac{4}{4+6} \cdot \frac{16}{16+x} + \frac{6}{4+6} \cdot \frac{x}{16+x} = 0.44 \end{aligned}$$

given ↴

$$\Rightarrow = \frac{4}{10} \cdot \frac{16}{16+x} + \frac{6}{10} \cdot \frac{x}{16+x} = 0.44$$

$$\Rightarrow 5(16+x) \left( \frac{2}{5} \cdot \frac{16}{16+x} + \frac{3}{5} \cdot \frac{x}{16+x} \right) = 0.44 \cdot 5(16+x)$$

$$\Rightarrow 2 \cdot 16 + 3 \cdot x = 2 \cdot 20 (16+x)$$

$$\Rightarrow 32 + 3x = 35.2 + 2.2x$$

-2.2x                          -2.2x

$$\Rightarrow 32 + 0.8x = 35.2$$

-32                            -32

$$\Rightarrow \frac{0.8x}{0.8} = \frac{3.2}{0.8}$$

$$\Rightarrow x = \frac{3.2}{0.8}$$

$$\Rightarrow x = 4$$

There are 4 blue balls in the second urn.

# Random Variables

## Definition

A random variable  $X$  is a function  $X : \Omega \rightarrow \mathbb{R}$  where  $P(X^{-1}(A))$  exists for each  $A \subseteq \mathbb{R}$ .

The probability that  $X$  takes on values in the subset  $S$  of  $\mathbb{R}$  is written

$$P(\underbrace{X \in S}_{\text{underline}}) = P(\{\omega \in \Omega : X(\omega) \in S\}).$$

The probability that  $X$  takes on value  $x$  in  $\mathbb{R}$  is written

$$P(\underbrace{X = x}_{\text{underline}}) = P(\{\omega \in \Omega : X(\omega) = x\}).$$

# Types of Random Variables

There are three types of random variables: *discrete* random variables, *continuous* random variables, and *mixed* random variables.

- A discrete random variable is a random variable whose range is either finite or countably infinite.
- A continuous random variable is a random variable whose range is uncountable. Un countable means list elements one-by-one,  
e.g.  $[0, 1]$  is uncountable
- A mixed random variable is partially discrete and partially continuous.

## Example

State whether the random variables are discrete, continuous, or mixed.

- (a) A coin is tossed ten times. The random variable  $X$  is the number of tails. Discrete.  $X = 0, 1, 2, \dots, 10$

- (b) The random variable  $Y$  measures the time until a firm defaults on its debt. Cont. The range is  $[0, \infty)$  ↪ uncountable set

- (c) The random variable  $Z$  measures the time, in years, until a firm defaults on a particular bond that matures in  $T$  years. If the firm meets all of its payment obligations for the bond,  $Z$  is given a value of  $T$ . Mixed

↑  
Stuff like zero filling

# Probability Mass Function

## Definition

For a discrete random variable  $X$ , we define the **probability mass function** or **pmf** by the equation

$$p(x) = P(X = x).$$

# Probability Mass Function Example

## Example

A coin is flipped three times. Let the random variable  $X$  denote the number of heads. Find the pmf of  $X$ .

Sol

$$X = 0$$

TTT

$$P(X=0) = \frac{1}{8}$$

$$X = 1$$

HTT, THT, TTH

$$P(X=1) = \frac{3}{8}$$

$$X = 2$$

HTH, THH, HHT

$$P(X=2) = \frac{3}{8}$$

$$X = 3$$

HHH

$$P(X=3) = \frac{1}{8}$$

The pmf

$$P(X) = \begin{cases} \frac{1}{8}, & X = 0, 3 \\ \frac{3}{8}, & X = 1, 2 \\ 0, & \text{else.} \end{cases}$$

# Cumulative Distribution Function

## Definition

For a discrete random variable  $X$ , we define the **cumulative distribution function** or **cdf** as

$$F(x) = P(X \leq x) = \sum_{t \leq x} P(X = t).$$

# Probability Mass Function Example

## Example

The random variable  $X$  denotes the number selected from the set  $\{1, 2, 3, 4\}$ . If each number is equally likely, find the cdf.

Sol

cdf  $P(X=k) = \frac{1}{4}$  for  $k=1, 2, 3, \text{ or } 4$

$$F(2) = P(X \leq 2) = P(X=1) + P(X=2) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2}$$

$$F(x) = \frac{x}{4} \leftarrow \text{almost true}$$

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{x+1}{4}, & 1 \leq x \leq 4 \\ 1, & x > 4 \end{cases}$$

e.g.  $x = 2.5$

$$F(2.5) = \frac{1-2.5+1}{4} = \frac{2}{4} = \frac{1}{2}$$

# Discrete Uniform Distribution

## Definition

The **discrete uniform distribution** is a symmetric probability distribution.

In this distribution, a finite number of values are equally likely to be observed. Every one of the  $n$  values has equal probability  $1/n$  of being observed.

# Binomial Distribution

## Definition

The **binomial distribution** with parameters  $n$  and  $p$  is the discrete probability distribution of the number of successes in a sequence of  $n$  independent experiments, each ending in either success with probability  $p$  or failure with probability  $q = 1 - p$ .

To denote that  $X$  is a binomial random variable with parameters  $n$  and  $p$ , we

write  $X \sim \mathcal{B}(n, p)$ . The probability

$$P(X = x) = \begin{cases} \frac{\# \text{ of ways}}{\text{to choose } x \text{ of } n} \cdot \binom{n}{x} p^x q^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{otherwise.} \end{cases}$$

Example: # heads in 3 flips of a coin.

# Binomial Distribution Example

## Example

Suppose a portfolio manager has a portfolio of 25 high yield bonds. She assumes that each bond has a one-year default probability of 5%. Her portfolio is well diversified so defaults are independent. What is the probability that more than one bond defaults?

$$\text{SOL}$$
$$P(N > 1) = \sum_{k=2}^{25} \binom{25}{k} (0.05)^k (1 - 0.05)^{25-k}$$

↑  
# of defaults

↑ "successes" is  
↓ "failure" is  
↑ "default" is  
↓ "non-default"

hard to calculate

$$= 1 - P(N \leq 1)$$

using the complement

$$= 1 - P(N=0) - P(N=1)$$

Binomial distribution formula

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= 1 - \left( \frac{1}{0.95} \right)^{25} (0.05)^0 - \left( \frac{25}{1} \right) (0.05)^1 (0.95)^{24}$$

$$= 1 - 0.95^{25} - 25 \times 0.05 \times 0.95^{24}$$

$$\approx 35.76\%$$

# Binomial Distribution Python Example

## Example

Use Python to check the previous result by simulating the event 100,000 times.

```
import numpy as np
from scipy.stats import binom
# Generate 100,000 simulations
defaults = binom.rvs(n = 25, p = 0.05, loc = 0, size = 100_000, random_state = 0)
# Calculate probability
prob = np.mean(defaults > 1)
print(f'The probability that more than one bond defaults is {prob:.3f}.')
```

Class for binomial distribution  
25 bonds  
Prob of default  
Method to generate random numbers  
Sets random seed so each time you run it it's the same

The output is 0.357.

# Poisson Distribution

## Definition

The **Poisson distribution** is a discrete probability distribution that expresses the probability of a given number of events  $x$  occurring in a fixed interval if these events occur independently with a constant rate  $\lambda$ .

To denote that  $X$  is a Poisson random variable with parameter  $\lambda$ , we write  $X \sim \mathcal{P}(\lambda)$ . The probability

$$P(X = x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases}$$

# Poisson Distribution Example

## Example

After graduation from the MFE program, you become a famous and successful portfolio manager. To help younger men and women succeed in finance, you begin writing a memoir of your life. In the first draft of your book, there are an average of three spelling errors per page. Suppose that the number of errors per page follows a Poisson distribution. What is the probability of having no errors on a page?

Sol     $\lambda = 3$

$$P(N=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$P(N=0) = \frac{3^0}{0!} e^{-3}$$

$$\approx 4.98\%$$

# Continuous Random Variables

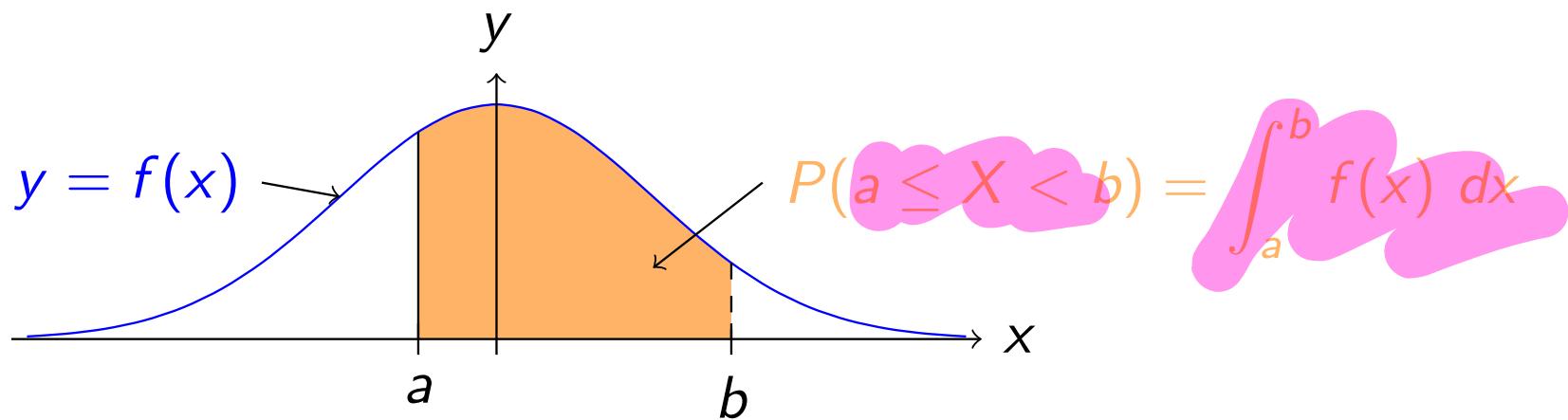
## Definition

We say that a random variable  $X$  is **continuous** if there exists a non-negative function  $f$  defined for all real numbers and having the property that for any set  $B$  of real numbers we have

$$P(X \in B) = \int_B f(x) dx.$$

The function  $f$  is called the **probability density function** or **pdf** of the random variable  $X$ .

# Diagram



Note: Area under a point is zero for a cont. distribution

so  $P(a \leq X < b) = P(a < X < b)$

# Cumulative Distribution Function

## Definition

The **cumulative distribution function** or **cdf** of a random variable  $X$  is defined to be

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

where  $f$  is the pdf of  $X$ .

# Properties of CDFs

The cumulative distribution function  $F$  of a continuous random variable  $X$  with probability density function  $f$  satisfies the following properties.

(a)  $0 \leq F(x) \leq 1$

(b) If  $x_1 < x_2$ , then  $F(x_1) < F(x_2)$

(c)  $\lim_{x \rightarrow -\infty} F(x) = 0$

(d)  $\lim_{x \rightarrow \infty} F(x) = 1$

*and (f)*

(e)  $P(a < X \leq b) = F(b) - F(a)$

(f)  $F$  is continuous

(g)  $F'(x) = f(x)$  whenever  $F'$  exists

All of these, except (g), hold for discrete random variables as well.

# PDF and CDF Example

## Example

Suppose a random variable  $X$ 's pdf is proportional to  $x^2$  on the interval  $[-1, 2]$  and 0 elsewhere. Find the pdf and cdf of  $X$ .

Sol Note:  $\int_{-1}^2 f(x) dx = 1$  pdf of  $X$

Given:  $f(x) = Cx^2$ .

$$\Rightarrow \int_{-1}^2 Cx^2 dx = 1 \Rightarrow \int_{-1}^2 Cx^2 dx = \frac{Cx^3}{3} \Big|_{-1}^2 = \frac{C \cdot 8}{3} - \frac{C \cdot (-1)}{3}$$
$$= \frac{8C}{3} + \frac{C}{3}$$
$$= \frac{9C}{3}$$
$$= 3C$$
$$\Rightarrow C = 1/3$$

$\Rightarrow$  P.d.f:

$$f(x) = \frac{1}{3} x^2$$

$$f(x) = \begin{cases} 0, & x \notin [-1, 2] \\ \frac{1}{3} x^2, & x \in [-1, 2] \end{cases}$$

$\Rightarrow$  C.d.f:

$$\begin{aligned} F(x) &= \int_{-1}^x \frac{1}{3} t^2 dt \\ &= \frac{1}{9} t^3 \Big|_{-1}^x \\ &= \frac{1}{9} x^3 - \frac{1}{9} (-1) \\ &= \frac{1}{9} x^3 + \frac{1}{9}, \quad -1 \leq x \leq 2 \end{aligned}$$

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{9} x^3 + \frac{1}{9}, & -1 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$$

Note: The Support of the RV  $X$  is  $[-1, 2]$ . The values of  $X$  where the p.d.f is nonzero.

# The Continuous Uniform Distribution

## Definition

The **continuous uniform distributions** describes an experiment where there is an arbitrary outcome that lies between certain bounds. The bounds are defined by the parameters  $a$  and  $b$ , which are the minimum and maximum values. The interval can either be closed or open. The distribution is often abbreviated

$$\mathcal{U}(a, b).$$



The pdf and cdf of the continuous uniform distribution on  $[a, b]$  are

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

and

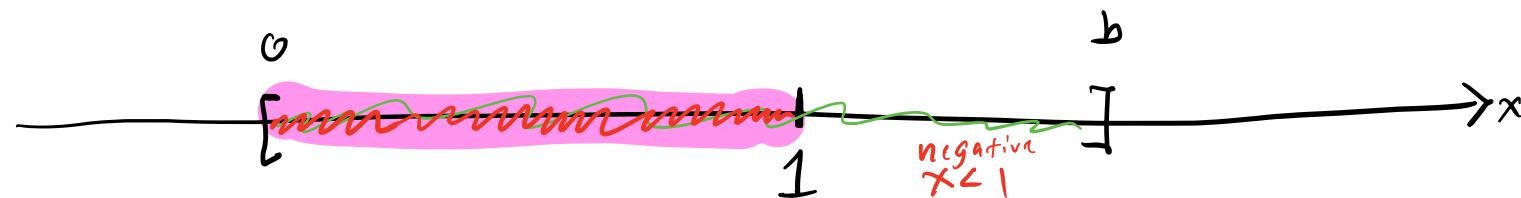
$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b. \end{cases}$$

# Continuous Uniform Distribution Example

## Example

Suppose that  $\underline{X} \sim U(0, b)$ , where  $b > 1$ . Find  $P(\underline{X^2} < X)$ .

Sol



$$P(X^2 < X) = P(X^2 - X < 0) = P(X(X-1) < 0)$$

always  
positive  
in  $[0, b]$

$$= \frac{1-0}{b-0}$$

$$= \frac{1}{b}$$

# Normal Distribution

## Definition

The **normal or Gaussian distribution** has probability density function

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

The parameters of the distribution are  $\mu$  and  $\sigma^2$ . To denote that  $X$  follows a normal distribution, we write  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

Be Careful sometimes  
People use  $\sigma$  instead  
of  $\sigma^2$

# Normal Python Example

## Example

Use Python to illustrate how  $\mu$  and  $\sigma$  affect the graph of the pdf.

```
import numpy as np, matplotlib.pyplot as plt
from scipy.stats import norm

# Use LaTeX
plt.rcParams['text.usetex'] = True

# Use Seaborn style
plt.style.use('seaborn')

# Mu- and sigma-values
mu_vals, sigma_vals = [-2, 0, 1], [0.5, 1, 1.5]

# Set up subplots
fig, ax = plt.subplots(1, 2, sharey = True, figsize = (10, 6))

# Get x-values
x_vals = np.linspace(-4.5, 3.5, 100)
```

# Normal Python Example

```
# Loop over mu-values
for mu in mu_vals:
    # Note norm.pdf vectorized
    y_vals = norm.pdf(x_vals, loc = mu, scale = 1)
    ax[0].plot(x_vals, y_vals, label = r'$\mu = $' + str(mu))

# Create legend for first subplot
ax[0].legend()

# Loop over sigma-values
for sigma in sigma_vals:
    # Note norm.pdf vectorized; also scale is std not var
    y_vals = norm.pdf(x_vals, loc = 0, scale = sigma)
    ax[1].plot(x_vals, y_vals, label = r'$\sigma = $' + str(sigma))

# Create legend for second subplot
ax[1].legend()

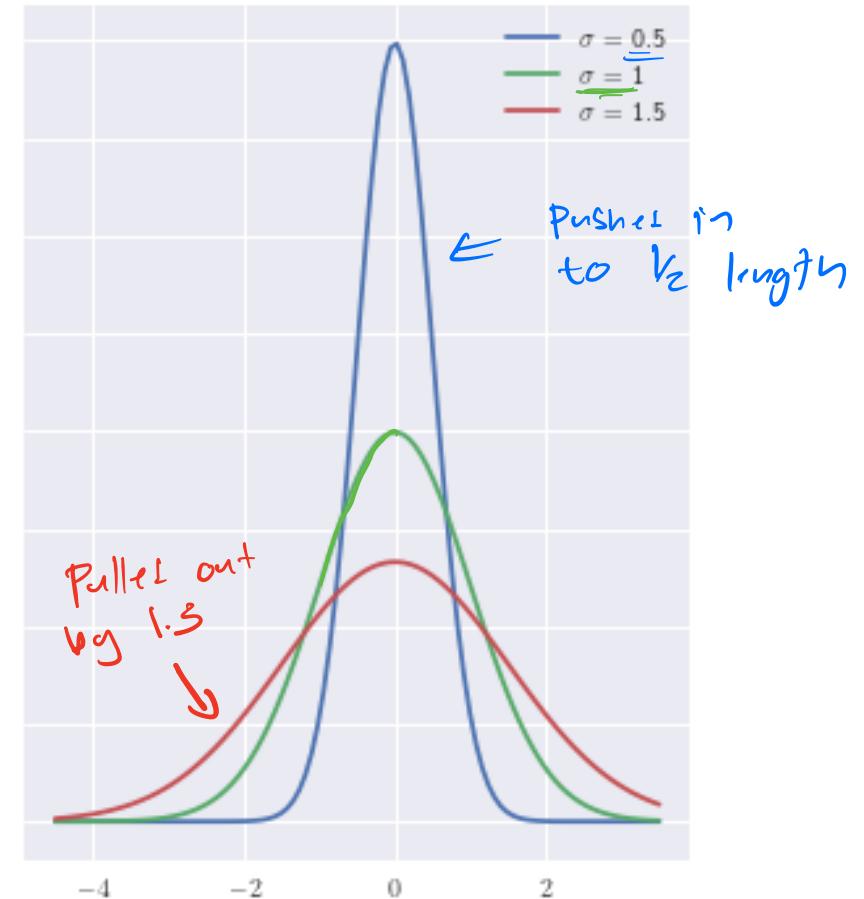
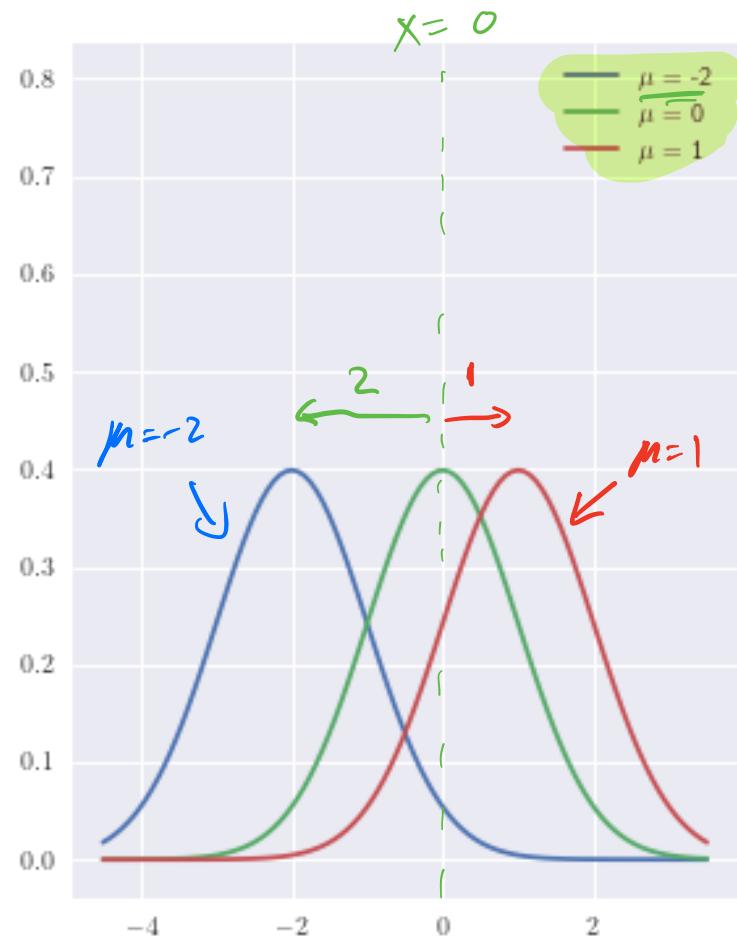
# Save the figure
plt.savefig(path + r'ex3-1.png')

# Show plot
plt.show()
```

mean  
not  $\sigma^2$ .  
this is  $\sigma$ .

✓  
✓

# Normal Python Example Result



# Standard Normal Distribution

## Definition

The **standard normal distribution** is  $\mathcal{N}(0, 1^2)$ , i.e. it is a normal distribution with  $\mu = 0$  and  $\sigma^2 = 1$ . The pdf and cdf are sometimes denoted  $\phi$  and  $\Phi$ , respectively.

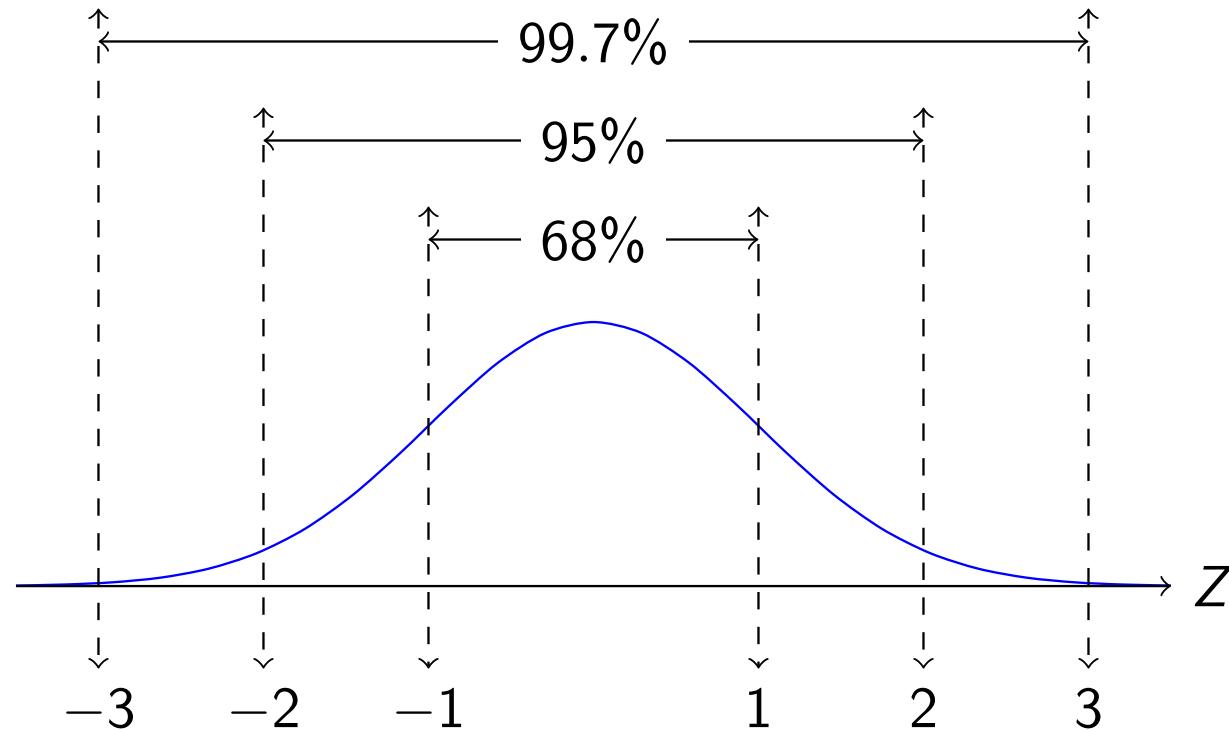
You can translate a normal random variable into a standard normal variable by calculating its Z-score:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

implied

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1^2).$$

# Empirical Rule



If  $Z \sim \mathcal{N}(0, 1^2)$ , then

$P(|Z| < 1) \approx 68\%$ ,       $P(|Z| < 2) \approx 95\%$ ,      and       $P(|Z| < 3) = 99.7\%$ .

# Normal Distribution Example

## Example

Suppose that the daily returns for a stock index  $R \sim \mathcal{N}(0, 0.01^2)$ . What is the probability of this index dropping more than 3%?

Sol  $R \sim \mathcal{N}(0, 0.01^2)$

$$R \mapsto \frac{R - 0}{0.01} \sim \mathcal{N}(0, 1^2) \quad \text{Z-score}$$

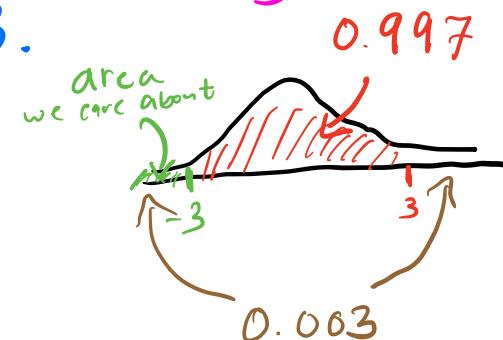
If  $R = -0.03$ , then  $Z = \frac{-0.03 - 0}{0.01} = -3$ .

$$P(R < -0.03) = P(Z < -3)$$

$$= 0.003/2$$

$$= 0.0015$$

$$= 0.15\%$$



# Exponential Distribution

## Definition

The **exponential distribution** describes an experiment where events occur continuously and independently at a constant average rate  $\lambda$ . Exponential random variables are often used to model arrival times, waiting times, and equipment failure times.

To denote that the random variable  $X$  follows an exponential distribution, we write  $X \sim \mathcal{E}(\lambda)$ . The pdf and cdf of an exponential distribution are

$$f(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$$

and

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

# Exponential Distribution Example

## Example

Suppose that the time until default, measured in years, of a firm  $T$  follows an exponential distribution with  $\lambda = \underline{\underline{0.10}}$ . What is the probability that  $2 < T < 5$ ?

Sol We want to know

$$P(2 < T < 5).$$

Recall cdf

$$F(t) = 1 - e^{-\lambda t} = 1 - e^{-0.10t} \quad P(T \leq t)$$

$$\Rightarrow P(2 < T < 5) = F(5) - F(2) = (1 - e^{-0.10 \cdot 5}) - (1 - e^{-0.10 \cdot 2})$$

$$\Rightarrow (1 - e^{-0.5}) - (1 - e^{-0.2}) = e^{-0.2} - e^{-0.5} \approx 21.22\%$$

# Expected Values

## Definition

The **expected value** of a discrete random variable  $X$  is

$$E[X] = \mu_X = \sum_x xP(x),$$

where we sum over the range of  $X$ . If  $X$  is a continuous random variable, then the expected value is

$$E[X] = \mu_X = \int_{-\infty}^{\infty} xf(x) \, dx.$$

# Property of Expected Values

For random variables  $X$  and  $Y$ , and scalars  $\alpha$  and  $\beta$ , we have

$$E[\alpha \underline{X} + \beta \underline{Y}] = \alpha E[X] + \beta E[Y].$$

Follows from linearity of integral or sum.

# Expected Values Example

## Example

Suppose that the pmf of  $N$  is

$$p(n) = \begin{cases} \left(\frac{1}{2}\right)^n, & n = 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases}$$

Calculate  $E[N]$ .

Sol

$$\begin{aligned} E[N] &= \sum_{n=1}^{\infty} n p(n) = \sum_{n=1}^{\infty} n \left(\frac{1}{2}\right)^n \\ &= \frac{1/2}{(1-1/2)^2} \end{aligned}$$

$$\begin{aligned} \frac{1}{1-x} &= \sum_{n=0}^{\infty} x^n, |x| < 1 \\ \Rightarrow \frac{d}{dx} \left( \frac{1}{1-x} \right) &= \sum_{n=0}^{\infty} \frac{d}{dx} (x^n) \\ \Rightarrow \frac{1}{(1-x)^2} &= \sum_{n=1}^{\infty} n x^{n-1} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1/2}{(1/2)^2} \\
 &= \frac{1/2}{1/4} \\
 &= \frac{1}{2} \cdot \frac{4}{1}^2 \\
 \Rightarrow & \boxed{E[N] = 2}
 \end{aligned}
 \quad \Rightarrow \quad \frac{x}{(1-x)^2} = \sum_{n=1}^{\infty} n \times^n$$

# Variance and Standard Deviation

## Definition

The **variance** of a discrete random variable  $X$  is

$$\text{Var}(X) = \sigma_X^2 = \sum_x (x - \mu_X)^2 p(x).$$

If  $X$  is a continuous random variable

$$\text{Var}(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx.$$

The **standard deviation** of  $X$  is the square root of its variance and is denoted by  $\sigma_X$ .

# Variance Example

## Example

Prove the identity

Super useful!



$$\text{Var}(X) = E[\underline{X^2}] - (E[X])^2.$$

Sol  $\text{Var}(x) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$

↑  
expected  
value of  $X$

↑  
pdf of  $X$

$$= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x) dx$$

Constant  
value

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \int_{-\infty}^{\infty} x f(x) dx + \mu^2 \int_{-\infty}^{\infty} f(x) dx \stackrel{1}{=} 1$$

$$= \underbrace{\int_{-\infty}^{\infty} x^2 f(x) dx}_{E[X^2]} - 2\mu \underbrace{\int_{-\infty}^{\infty} x f(x) dx}_{\mu} + \mu^2 \underbrace{\int_{-\infty}^{\infty} f(x) dx}_{1}$$

$$\begin{aligned}\rightarrow &= E[x^2] - 2\mu \cdot \mu + \mu^2 \cdot 1 \\ &= E[x^2] - 2\mu^2 + \mu^2 \\ &= E[x^2] - \mu^2 \leftarrow \mu = E[x]\end{aligned}$$

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

# Property of Variance

For random variable  $X$  and scalars  $\alpha$  and  $\beta$

$$\text{Var}(\underbrace{\alpha X + \beta}_{\text{= =}}) = \underbrace{\alpha^2}_{\text{=}} \text{Var}(X).$$

Note:  $E[\alpha X + \beta] = \alpha E[X] + \beta$

$$\begin{aligned}\text{Var}(\alpha X + \beta) &= \int_{-\infty}^{\infty} (\alpha X + \beta - \overbrace{E[\alpha X + \beta]}^{= E[X] + \beta})^2 f(x) dx \\ &= \int_{-\infty}^{\infty} (\alpha X + \cancel{\beta} - \alpha E[X] - \cancel{\beta})^2 f(x) dx \\ &= \int_{-\infty}^{\infty} (\alpha X - \alpha E[X])^2 f(x) dx\end{aligned}$$

$$\begin{aligned} &= \int_{-\infty}^{\infty} \sigma^2 (x - E[x])^2 f(x) dx \\ &= \sigma^2 \int_{-\infty}^{\infty} (x - E[x])^2 f(x) dx \\ &= \sigma^2 \text{Var}(X) \end{aligned}$$

# Expected Value and Variance Formulas

Distribution	Expected Value	Variance
Discrete Uniform	$\frac{1}{n} \sum_{k=1}^n x_k$	$\frac{1}{n} \sum_{k=1}^n x_k^2 - \left( \frac{1}{n} \sum_{k=1}^n x_k \right)^2$
$\mathcal{B}(n, p)$	$np$	$np(1 - p)$
$\mathcal{P}(\lambda)$	$\lambda$	$\lambda$
$\mathcal{U}(a, b)$	$\frac{b + a}{2}$	$\frac{(b - a)^2}{12}$
$\mathcal{N}(\mu, \sigma^2)$	$\mu$	$\sigma^2$
$\mathcal{E}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

# $k$ -th Moment

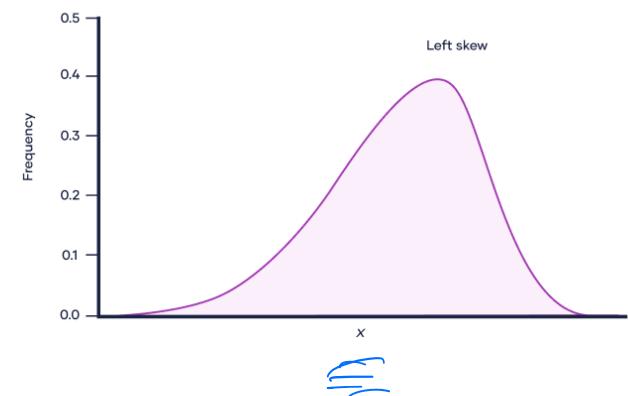
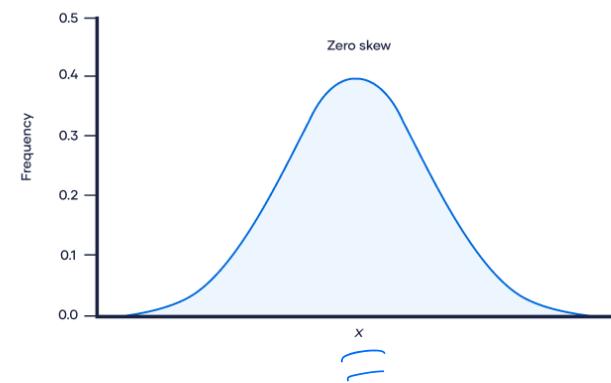
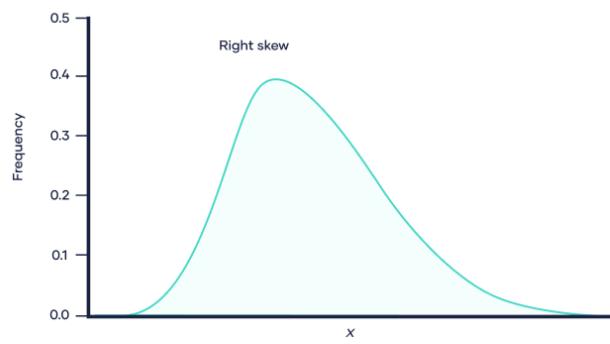
## Definition

- If  $X$  is a random variable, then its  **$k$ -th moment** is  $E[X^k]$ .
- If  $\mu = E[X]$  and  $\sigma^2 = \text{Var}(X)$ , then the **standardized  $k$ -th moment** is  $E\left[\left(\frac{X-\mu}{\sigma}\right)^k\right]$ .

# Skew

## Definition

The standardized third moment is the **skew**.

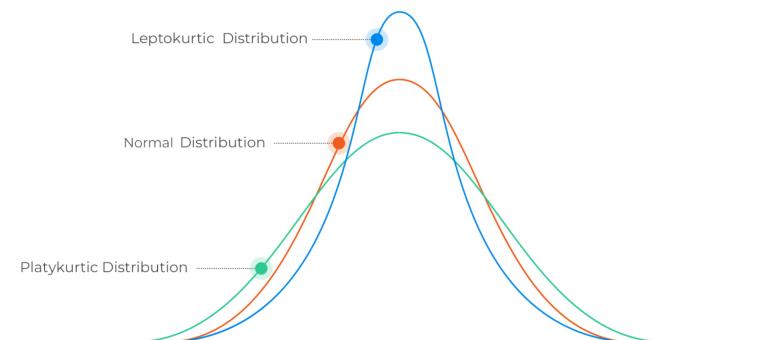


$$\hat{E} \left[ \left( \frac{x - \mu}{\sigma} \right)^3 \right]$$

# Kurtosis

## Definition

The standardized fourth moment is the **kurtosis**. A normal distribution has kurtosis of 3. If the kurtosis is greater than 3, it is leptokurtic, and if the kurtosis is less than 3 it platykurtic.



$$E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right]$$

Note: People are a little sloppy about kurtosis and think bigger kurtosis  $\Leftrightarrow$  bigger tails

The **kurtosis** is typically considered a measure of tail thickness, though the true meaning is more nuanced:

<https://stats.stackexchange.com/questions/172467/what-is-the-meaning-of-tail-of-kurtosis>.

# Moment Generating Function

## Definition

The **moment generating function** of a random variable  $X$  is

$$\psi(t) = E[e^{\underline{tX}}].$$

We call  $\psi$  the moment generating function of  $X$  because

$$E[X^k] = \psi^{(k)}(\underline{0}).$$

To see why this is the case, it is helpful to note

$$e^{\underline{tX}} = \underline{1} + \underline{tX} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \dots$$

# Moment Generating Function Example

## Example

Find the moment generating function of  $X$ , the discrete uniform distribution with support  $\{1, 2, 3\}$ .

Sol  $\Psi(t) = E[e^{tx}] = e^{t \cdot 1} \cdot \frac{1}{3} + e^{t \cdot 2} \cdot \frac{1}{3} + e^{t \cdot 3} \cdot \frac{1}{3}$

$$= \frac{1}{3} (e^t + e^{2t} + e^{3t})$$

Solution

Let's check out the derivatives.

$$\Psi'(t) = \frac{1}{3} (e^t + 2e^{2t} + 3e^{3t}) \xrightarrow{t=0} \Psi'(0) = \frac{1}{3} (1 + 2 + 3) = E[X]$$

$$\Psi''(t) = \frac{1}{3} (e^t + 2^2 e^{2t} + 3^2 e^{3t}) \xrightarrow{t=0} \Psi''(0) = \frac{1}{3} (1^2 + 2^2 + 3^2) = E[X^2]$$

# Function of a RV Python Example

## Example

Suppose  $Z \sim \mathcal{N}(0, 1)$ . Consider  $X = Z^2$ . Use Python to approximate the graph of the pdf of  $X$  using a histogram.

```
# Import modules
import numpy as np, matplotlib.pyplot as plt
from scipy.stats import norm

# Use LaTeX
plt.rcParams['text.usetex'] = True

# Use Seaborn style
plt.style.use('seaborn')

# Generate Z
Z = norm.rvs(size = 100_000) ← Standard normal, so default params are fine
# Calculate X
X = Z**2 ← loc=0, scale=1 ← std of normal distribution

# Generate histogram; make sure density is True
plt.hist(X, bins = 100, density = True)

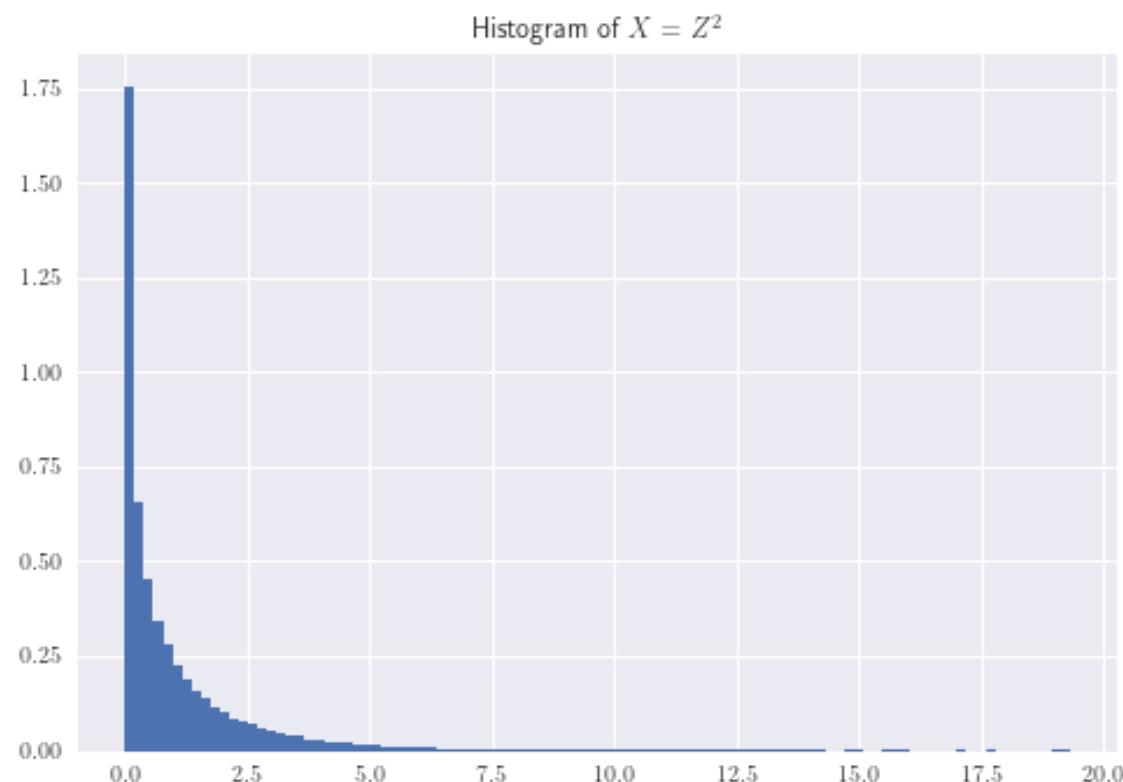
# Get title
plt.title(r'Histogram of $X = Z^2$')

# Save the figure
plt.savefig(path + r'ex3-2.png')

plt.show()
```

# Function of a Random Variable Python Result

This is a known distribution  $\chi^2(1)$ . See Wikipedia for more details.



# Functions of a Random Variable

This result can be obtained analytically by considering the cdf and differentiating.

# Function of a Random Variable Example

## Example

Suppose  $Z \sim \mathcal{N}(0, 1)$ . Consider  $X = Z^2$ . Calculate the pdf of  $X$ .

Sol Consider

cdf of  $X$

$$F_X(x) = P(X \leq x) = P(\sqrt{Z^2} \leq \sqrt{x}) = P(|Z| \leq \sqrt{x})$$

Assume  
 $x \geq 0$

$$= P(-\sqrt{x} \leq Z \leq \sqrt{x})$$

pdf of  $X$

$$= \Phi(\sqrt{x}) - \Phi(-\sqrt{x}), \quad \Phi \text{ is cdf of } Z \sim \mathcal{N}(0, 1^2)$$

$$\Rightarrow f_X(x) = \frac{d}{dx} (\Phi(\sqrt{x}) - \Phi(-\sqrt{x})), \quad \Phi \text{ is cdf of } Z \sim \mathcal{N}(0, 1^2)$$

$$= \phi(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} - \phi(-\sqrt{x}) \cdot \frac{-1}{2\sqrt{x}}$$

$$= \phi(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} + \phi(-\sqrt{x}) \cdot \frac{1}{2\sqrt{x}}$$

$\phi$  is even, i.e. symmetric  
about the y-axis  
 $\phi(-t) = \phi(t)$

$$\begin{aligned}
 & \Rightarrow = \phi(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} + \phi(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} \\
 & = \phi(\sqrt{x}) \cdot \frac{1}{\sqrt{x}} \\
 & = \frac{1}{\sqrt{x}} \phi(\sqrt{x}), \quad \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \\
 & = \frac{1}{\sqrt{x}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\sqrt{x}^2/2} \\
 & = \frac{1}{\sqrt{2\pi x}} e^{-x/2}, \quad x \geq 0
 \end{aligned}$$

$$\Rightarrow f_x(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\sqrt{2\pi x}} e^{-x/2}, & x \geq 0 \end{cases}$$

# Functions of a Random Variable Theorem

## Theorem

Let  $X$  be a random variable with pdf  $f$ . Suppose  $P(a < x < b) = 1$ . Let  $Y = r(X)$  and suppose  $r$  is differentiable and one-to-one for  $a < x < b$ .

Let the image of  $(a, b)$  under  $r$  be  $(\alpha, \beta)$ . If  $s$  is the inverse of  $r$ , then the pdf of  $Y$  is

$$g(y) = \begin{cases} f(s(\underline{y})) \left| \frac{ds}{dy} \right|, & \alpha < y < \beta \\ 0, & \text{otherwise.} \end{cases}$$

# Functions of a Random Variable Example

## Example

Suppose  $X \sim U(0, 1)$  and let  $Y = \frac{1}{27}X^3$ . Find the pdf of  $Y$ .

Support is  
finite length

One-to-one

sol Pdf of  $X$

$$f_x(x) = \frac{1}{1} = 1.$$

$$r(x) = \frac{1}{27}x^3 \Rightarrow s(y) = 3\sqrt[3]{y} = 3y^{1/3}$$

So, from the formula on the last page

$$\begin{aligned} f_y(y) &= f_x(s(y)) \cdot \frac{d}{dy}(s(y)) \\ &= 1 \cdot \frac{1}{3} y^{-2/3} = \frac{1}{3y^{2/3}}, \quad 0 < y < 1 \end{aligned}$$

# Joint Distributions

## Definition

Two random variables  $X$  and  $Y$  are said to be **jointly continuous** if there exists a function  $f_{XY}(x, y) \geq 0$  with the property that for every subset  $C$  of  $\mathbb{R}^2$ , we have

$$P((X, Y) \in C) = \iint_C f_{XY}(x, y) \, dA.$$

The function  $f_{XY}(x, y)$  is called the **joint probability density function** of  $X$  and  $Y$ , and the **joint cumulative distribution function** is

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(s, t) \, ds \, dt.$$

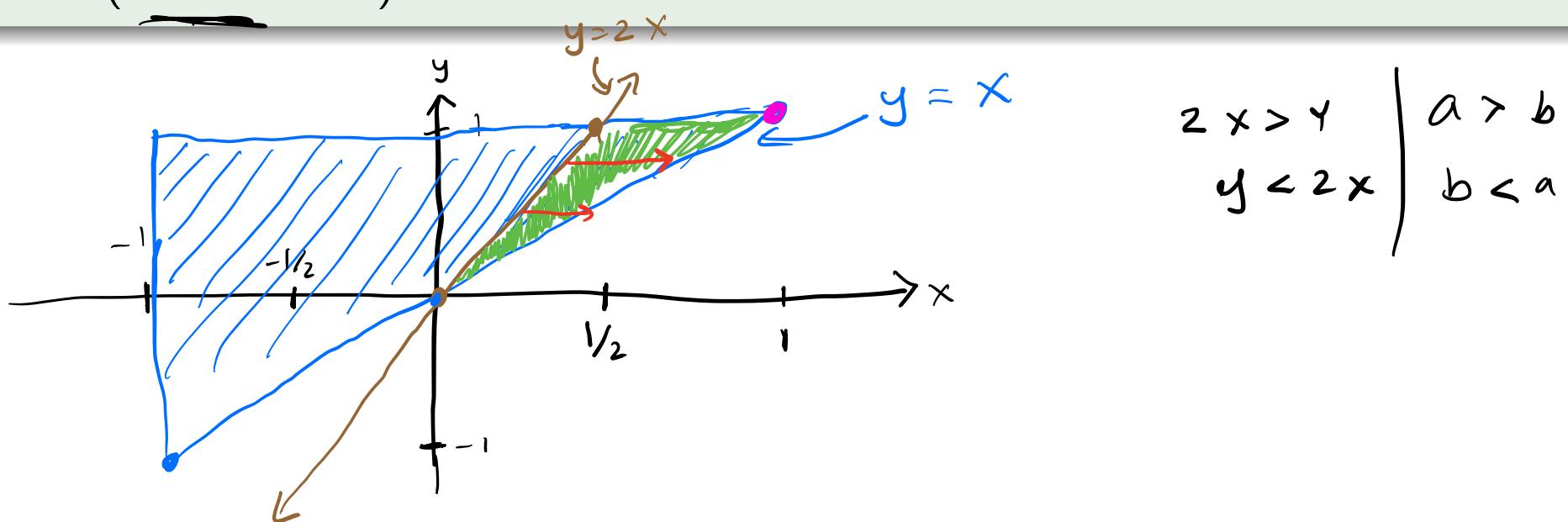
# Joint Distributions Example

## Example

Suppose

$$f_{XY}(x, y) = \begin{cases} \frac{1}{2}, & -1 \leq x \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then  $P(2X - Y > 0) =$



$$\underline{Sol} \quad P(2x - y > 0) = \int_0^1 \int_{y/2}^y \frac{1}{2} dx dy$$

$$= \frac{1}{2} \int_0^1 \int_{y/2}^y dx dy$$

$$= \frac{1}{2} \int_0^1 x \left[ \begin{matrix} y \\ x=y/2 \end{matrix} \right] dy$$

$$= \frac{1}{2} \int_0^1 y - y/2 dy$$

$$= \frac{1}{4} \int_0^1 y dy$$

$$= \frac{1}{8} y^2 \Big|_{y=0}^1$$

$$= \frac{1}{8} (1^2 - 0^2)$$

$$\boxed{P(2x - y > 0) = \frac{1}{8}}$$

# Marginal PDF

The marginal pdfs of  $X$  and  $Y$  given joint pdf  $f_{XY}$  are

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

# Independent Events

If the random variables  $X$  and  $Y$  are independent, then

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

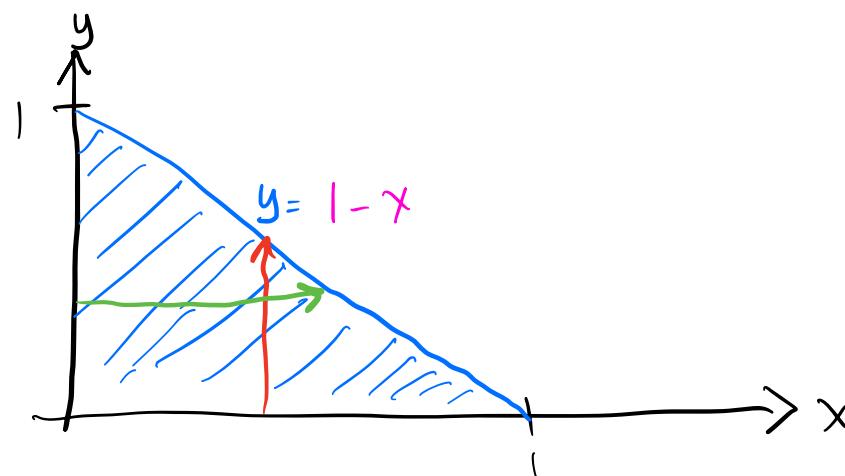
# Independent Events Example

## Example

The joint pdf of  $X$  and  $Y$  is given by

$$f_{XY}(x, y) = \begin{cases} 3(x + y), & x + y \leq 1, x, y \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Are  $X$  and  $Y$  independent?



In class, I wrote  $y = x$  instead of  $y = 1 - x$ !

Sol

$$f_x(x) = \int_{y=0}^{1-x} 3(x+y) dy = 3\left(xy + \frac{y^2}{2}\right) \Big|_{y=0}^{1-x} = 3\left[x(1-x) + \frac{(1-x)^2}{2}\right] \\ = 3(1-x)\left(x + \frac{1-x}{2}\right) \\ = \frac{3}{2}(1-x)(1+x)$$

$$f_y(y) = \int_{x=0}^{1-y} 3(x+y) dx = \frac{3}{2}(1-y)(1+y) \quad \text{Symmetry tells us same as before but with } y$$

To see if independent, check if

$$f_{xy}(x,y) = f_x(x) \cdot f_y(y).$$

In this case,

$$3(x+y) = \underbrace{\frac{3}{2}(1-x)(1+x)}_{f_x(x)} \cdot \underbrace{\frac{3}{2}(1-y)(1+y)}_{f_y(y)}$$

$$\Leftrightarrow 3(x+y) \neq \frac{9}{4}(1-x^2)(1-y^2)$$

Not independent!

# Change of Variables

## Theorem

Let  $X$  and  $Y$  be continuous random variables with joint density  $f_{XY}$ . Assume that there is a subset  $S$  of  $\mathbb{R}^2$  such that  $P((X, Y) \in S) = 1$ . Let

$$U = r_1(X, Y) \quad \text{and} \quad V = r_2(X, Y),$$

where  $r_1$  and  $r_2$  are one-to-one differentiable transformations from  $S$  onto some subset  $T$  of  $\mathbb{R}^2$ . If

$$X = s_1(U, V) \quad \text{and} \quad Y = s_2(U, V)$$

describe the inverse transformation, where  $s_1$  and  $s_2$  are also one-to-one and differentiable. Then the joint pdf of  $U$  and  $V$  is

$$g_{UV}(u, v) = \begin{cases} f_{XY}(s_1(u, v), s_2(u, v)) |\det(J)|, & (u, v) \in T \\ 0, & \text{otherwise.} \end{cases}$$

where

$$J = \begin{pmatrix} \frac{\partial s_1}{\partial u} & \frac{\partial s_1}{\partial v} \\ \frac{\partial s_2}{\partial u} & \frac{\partial s_2}{\partial v} \end{pmatrix}$$

is the Jacobian.

# Change of Variables Example

## Example

Suppose  $f_{XY}(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ . Find the joint pdf of  $r$  and  $\theta$ , where  $x = r \cos \theta$  and  $y = r \sin \theta$ .

Sol

$$J = \begin{vmatrix} x & \theta \\ \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix}$$

$$\Rightarrow \det(J) = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \cos^2 \theta - (-r \sin^2 \theta) \\ = r \cos^2 \theta + r \sin^2 \theta \\ = r (\cos^2 \theta + \sin^2 \theta) \\ = r$$

So, the pdf of  $r$  and  $\theta$  is

$$f_{R\theta}(r, \theta) = f_{XY}(x, y) \det(J) = \frac{1}{2\pi} e^{-\frac{r^2(\cos^2 \theta + \sin^2 \theta)}{2}} \cdot r \\ = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \cdot r$$

$$f_{R\theta}(r, \theta) = \frac{r}{2\pi} e^{-r^2/2}$$

# Covariance

## Definition

Let  $X$  and  $Y$  be random variables having finite means. Suppose  $E[X] = \mu_X$  and  $E[Y] = \mu_Y$ . The **covariance of  $X$  and  $Y$**  is

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

A useful identity is

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y].$$

$$\begin{aligned}
 \text{Cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\
 &= \iint_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx dy \quad \text{cont. returns} \\
 &= \iint_{-\infty}^{\infty} (xy - x\mu_Y - y\mu_X + \mu_X\mu_Y) f(x, y) dx dy \\
 &= \iint_{-\infty}^{\infty} xy f(x, y) dx dy - \iint_{-\infty}^{\infty} x\mu_Y f(x, y) dx dy \\
 &\quad - \iint_{-\infty}^{\infty} y\mu_X f(x, y) dx dy \\
 &\quad + \iint_{-\infty}^{\infty} \mu_X\mu_Y f(x, y) dx dy \\
 &= \underbrace{\iint_{-\infty}^{\infty} xy f(x, y) dx dy}_{E[XY]} - \mu_Y \iint_{-\infty}^{\infty} x f(x, y) dx dy \\
 &\quad - \mu_X \iint_{-\infty}^{\infty} y f(x, y) dx dy \\
 &\quad + \mu_X\mu_Y \iint_{-\infty}^{\infty} f(x, y) dx dy \\
 &= \overline{E[XY]} - \mu_Y \overline{E[X]} - \mu_X \overline{E[Y]} + \mu_X\mu_Y
 \end{aligned}$$

$\boxed{\text{Cov}(X, Y) = E[XY] - E[X]E[Y]}$

# Covariance Example 1

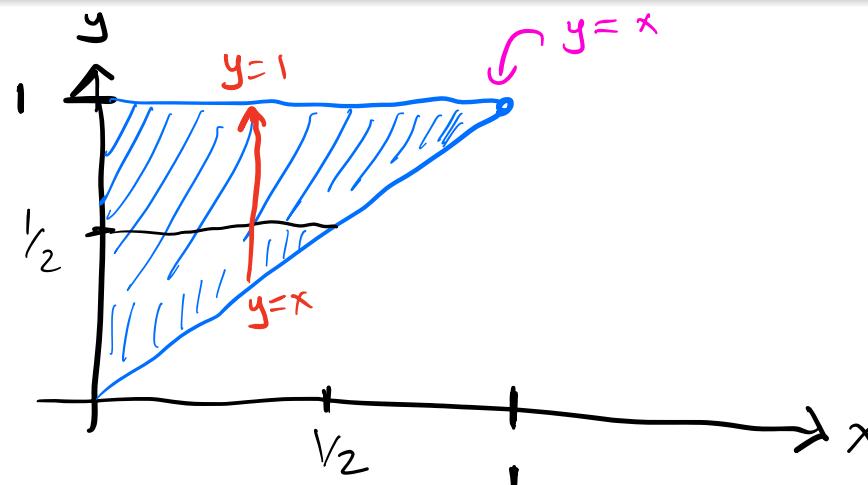
## Example

Consider  $X$  and  $Y$  with joint pdf

$$f_{XY}(x, y) = \begin{cases} 2, & 0 \leq x \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Calculate the covariance of  $X$  and  $Y$ . Hint:  $E[Y] = \frac{2}{3}$ , and  $E[X] = \frac{1}{3}$ .

Sol



$$\begin{aligned}
 \text{Cov}(x, y) &= E[X^4] - E[X]E[Y] \\
 &= \iint_{\mathbb{R}^2} xy f(x,y) dA - \frac{1}{3} \cdot \frac{2}{3} \quad \text{given} \\
 &= \int_0^1 \int_{y=x}^1 xy \cdot 2 dy dx - \frac{1}{3} \\
 &= 2 \int_0^1 \int_{y=x}^1 xy dy dx - \frac{1}{3} \\
 &= 2 \int_0^1 \frac{xy^2}{2} \Big|_{y=x}^1 dx - \frac{1}{3} \\
 &= 2 \int_0^1 \frac{x}{2} - \frac{x^3}{2} dx - \frac{1}{3} \\
 &= \int_0^1 x - x^3 dx - \frac{1}{3} \\
 &= \left[ \frac{x^2}{2} - \frac{x^4}{4} \right]_0^1 - \frac{1}{3} \\
 &= \frac{1}{2} - \frac{1}{4} - 0 - \frac{1}{3} \\
 &= \frac{2}{4} - \frac{1}{4} - \frac{1}{3} \\
 &= \frac{1}{4} - \frac{1}{3} \\
 &= \frac{3-4}{12} \\
 &= -\frac{1}{12}
 \end{aligned}$$

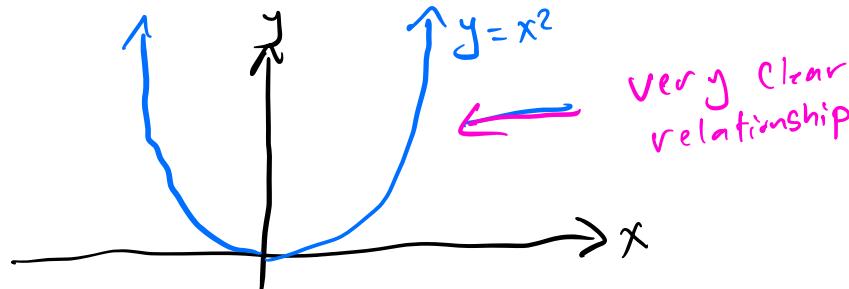
negative is fine

:-)

# Covariance Example 2

## Example

Suppose  $X \sim U(-1, 1)$ . Calculate the covariance between  $\underline{X}$  and  $\underline{\underline{X^2}}$ .



Sol  $\text{Cov}(X, X^2) = E[X \cdot X^2] - E[X] \cdot E[X^2]$

$$= E[X^3] - E[X] \cdot E[X^2]$$

$E[X^3] = \int_{-1}^1 x^3 \cdot \frac{1}{2} dx = \frac{1}{2} \int_{-1}^1 x^3 dx = 0$  *Original symmetry*

$f_X(x) = \frac{1}{1-(-1)} = \frac{1}{2}$

$E[X] = \int_{-1}^1 x \cdot \frac{1}{2} dx = \frac{1}{2} \int_{-1}^1 x dx = 0$

$E[X^2] = \int_{-1}^1 x^2 \cdot \frac{1}{2} dx = \frac{1}{2} \int_{-1}^1 x^2 dx = \frac{x^3}{3} \Big|_0^1 = \frac{1}{3}$

$$\begin{aligned}\text{Cov}(x, x^2) &= E[x^3] - E[x] \cdot E[x^2] \\ &= 0 - 0 \cdot \frac{1}{3} \\ &= 0\end{aligned}$$

The covariance between  $x$  and  $x^2$  is 0.

# Properties of Covariance

Suppose  $X_i$  and  $Y_j$  are random variables for all  $i$  and  $j$ , and  $\alpha$  is a real number.

(a)  $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$

(b)  $\text{Cov}(X_1, X_1) = \text{Var}(X_1)$

(c)  $\text{Cov}(\alpha X_1, X_2) = \alpha \text{Cov}(X_1, X_2)$

(d)  $\text{Cov}\left(\sum_{i=1}^m X_i, \sum_{j=1}^n Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n \text{Cov}(X_i, Y_j)$

$$\text{Cov}(x, y) = E[xy] - E[x]E[y]$$

$$x = y$$

$$\begin{aligned}\text{Cov}(x, x) &= E[x^2] - (E[x])^2 \\ &\approx \text{Var}(x)\end{aligned}$$

# Correlation

## Definition

Let  $X$  and  $Y$  be random variables with finite variances  $\sigma_X^2$  and  $\sigma_Y^2$ , respectively. Then the **correlation of  $X$  and  $Y$**  is

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

In Python the correlation between two variables can be calculated using

`np.corrcoef`. See the NumPy documentation for more details.

Correlation btw  $x$  and  $y$  is in  $E[xy]$

# Correlation Plot Python Code

```
# Import modules
import numpy as np, matplotlib.pyplot as plt
from scipy.stats import norm

# Use LaTeX
plt.rcParams['text.usetex'] = True

# Use Seaborn style
plt.style.use('seaborn')

# Generate standard normals
Z = norm.rvs(size = 100)

# List correlations
rho_vals = [-0.9, -0.75, -0.25, 0.25, 0.75,
             0.9]

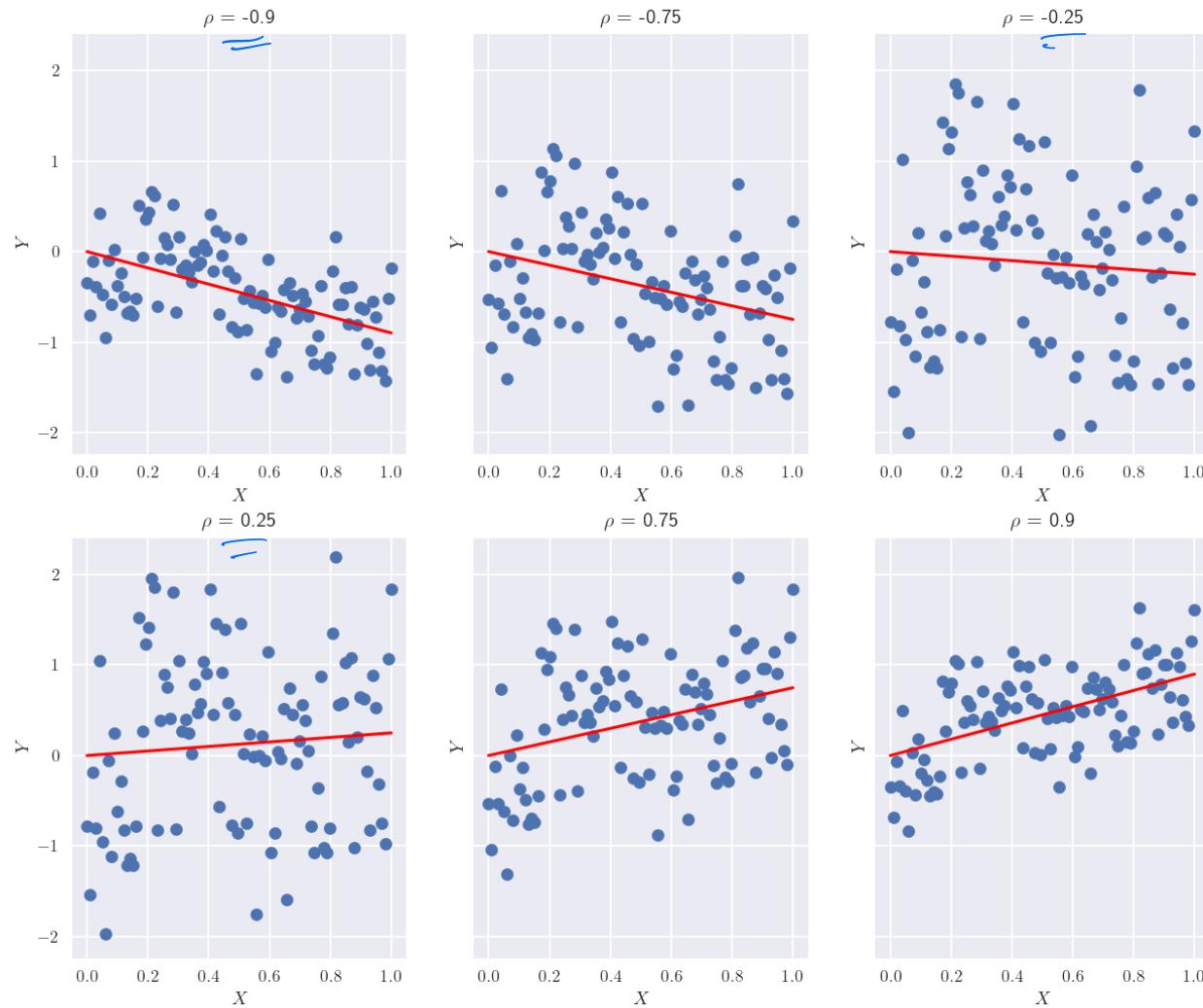
# Set up subplots
fig, ax = plt.subplots(2, 3, sharey = True,
                      figsize = (12, 10),
                      dpi = 125)

for i, rho in enumerate(rho_vals):
    # Get X and Y values
```

```
X = np.linspace(0, 1, len(Z))
Y = rho * X + np.sqrt(1 - rho**2) * Z
# Get row and column
row, col = i//3, i%3
# Draw scatter plot
ax[row, col].scatter(X, Y)
# Draw plot of underlying relationship
ax[row, col].plot(X, rho * X, color = 'red')
# Add title to subplot
ax[row, col].set_title(r'$\rho = $' +
                      str(rho))
# Add labels for x and y axes
ax[row, col].set_xlabel(r'$X$')
ax[row, col].set_ylabel(r'$Y$')
# Save the figure
plt.savefig(path + r'ex3-3.png')
plt.show()
```

$\text{corr}(x, t) = \rho$

# Correlation Plot



# Variance of Sum

## Theorem

If  $X$  and  $Y$  are random variables and  $\alpha$  and  $\beta$  are constants, then

$$\text{Var}(\underline{\alpha}X + \underline{\beta}Y) = \alpha^2 \text{Var}(X) + \beta^2 \text{Var}(Y) + 2\alpha\beta \text{Cov}(X, Y).$$

# Variance of Sum Example

## Example

Suppose  $X \sim U(-1, 1)$ . Calculate the variance of  $\underline{X - 2X^2}$ .

Sol  $\text{Var}(X - 2X^2) = 1^2 \text{var}(X) + (-2)^2 \text{var}(X^2) + 2 \cdot 1 \cdot (-2) \text{cov}(X, X^2)$

$= \text{var}(X) + 4 \text{var}(X^2) - 4 \text{cov}(X, X^2)$  ← From the previous problem

$= \text{var}(X) + 4 \text{var}(X^2)$   $\text{Var}(x) = E[X^2] - (E[x])^2$

$= E[X^2] - (E[X])^2 + 4(E[X^4] - (E[X^2])^2)$  ←

$$E[X] = \int_{-1}^1 x \cdot \frac{1}{2} dx = 0$$

$$E[X^2] = \int_{-1}^1 x^2 \cdot \frac{1}{2} dx = \int_0^1 x^2 dx = \frac{x^3}{3} \Big|_0^1 = \frac{1}{3}$$

$$E[X^4] = \int_{-1}^1 x^4 \cdot \frac{1}{2} dx = \frac{1}{2} \int_{-1}^1 x^4 dx = \frac{1}{2} \cdot 2 \int_0^1 x^4 dx = \frac{x^5}{5} \Big|_0^1 = \frac{1}{5}$$

$$\text{Var}(X - 2X^2) = \frac{1}{3} - 0^2 + 4\left(\frac{1}{5} - \left(\frac{1}{3}\right)^2\right)$$

$$\rightarrow = \frac{1}{3} - 0 + 4 \left( \frac{1}{5} - \frac{1}{9} \right)$$

$$= \frac{1}{3} + \frac{4}{5} - \frac{4}{9}$$

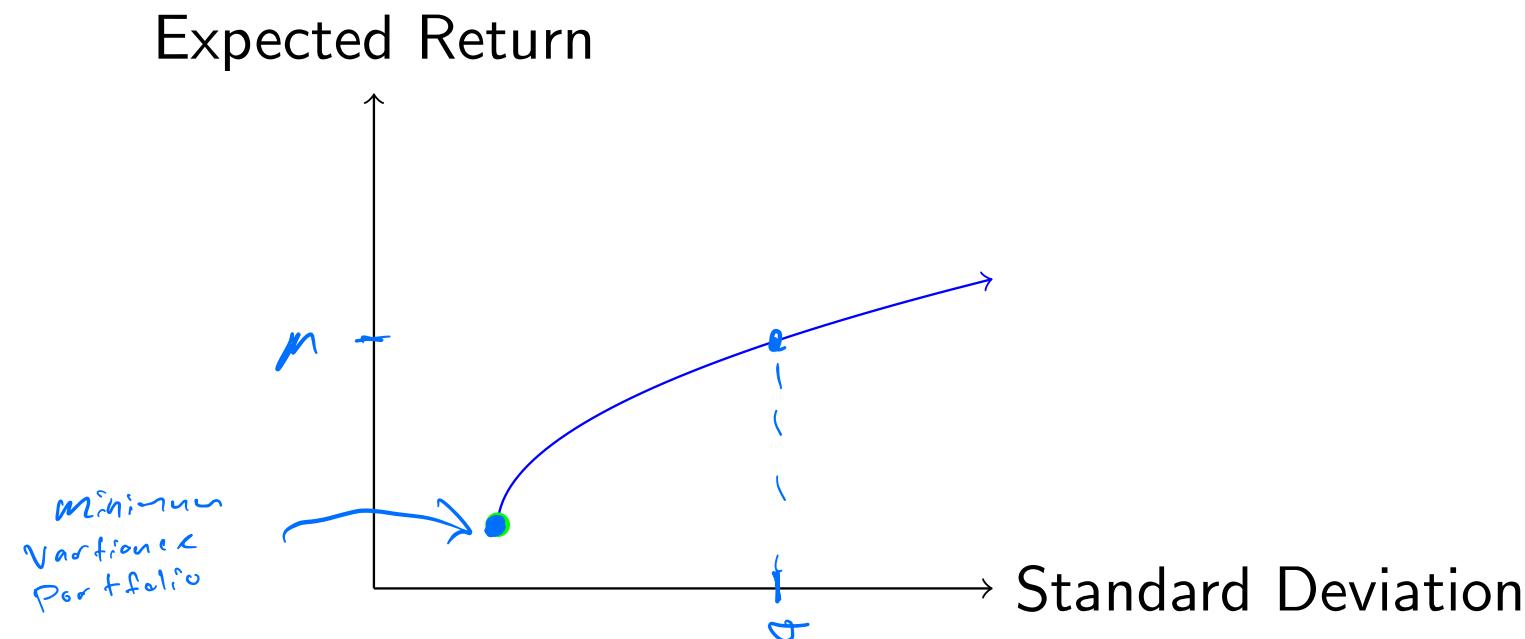
$$= \frac{15}{45} + \frac{36}{45} - \frac{20}{45}$$

$$\boxed{\text{Var}(x - 2x^2) = \frac{31}{45}}$$

# Efficient Frontier

## Definition

The **efficient frontier** is the set of portfolios which satisfy the condition that no other portfolio in the investment universe exists with a higher expected return and the same standard deviation of return.



# Portfolio Construction

## Example

Consider assets 1 and 2 with monthly expected returns of 1% and 0.5% and standard deviations of 10% and 6%, respectively. Suppose the correlation between returns is 25%.

- (a) Find the minimum variance portfolio.
- (b) Graph the efficient frontier.

Assume that you cannot take short positions in this investment universe.

(a) The return of your portfolio is

$$R_p = \omega R_1 + (1-\omega) R_2.$$

↑ weight in first security      ↑ weight in second security

← weights sum to 1 and  $0 \leq \omega \leq 1$ , because no short positions

$$\Rightarrow \text{Var}(R_p) = \text{Var}(\omega R_1 + (1-\omega) R_2)$$

$$= \omega^2 \text{Var}(R_1) + (1-\omega)^2 \text{Var}(R_2) + 2\omega(1-\omega) \text{Cov}(R_1, R_2)$$

$$= \sigma_1^2 \omega^2 + \sigma_2^2 (1-\omega)^2 + 2\omega(1-\omega) \rho \sigma_1 \sigma_2$$

$$\rho = 0.25$$

$$\text{SP}(R_1) = 10\%$$

$$\text{SD}(R_2) = 6\%$$

$$= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$\Rightarrow \frac{d}{d\omega} (\text{Var}(R)) = 2\sigma_1^2 - 2\sigma_2^2(1-\omega) + 2(1-\omega)\rho\sigma_1\sigma_2 + 2\omega(-1)\rho\sigma_1\sigma_2$$

$$\Leftrightarrow \text{Cov}(X, Y) = \rho \sigma_X \sigma_Y$$

$$= (2\sigma_1^2 + 2\sigma_2^2 - 2\rho\sigma_1\sigma_2 - 2\rho\sigma_1\sigma_2) \omega$$

$$- 2\sigma_2^2 + 2\rho\sigma_1\sigma_2$$

$$= (2\sigma_1^2 + 2\sigma_2^2 - 4\rho\sigma_1\sigma_2) \omega - 2\sigma_2^2 + 2\rho\sigma_1\sigma_2$$

$$\stackrel{\text{let}}{=} 0$$

$$\Rightarrow \omega = \frac{2\sigma_2^2 - 2\rho\sigma_1\sigma_2}{2\sigma_1^2 + 2\sigma_2^2 - 4\rho\sigma_1\sigma_2}$$

$$\approx 19.18\%$$

So, 19.18% in first security and 80.19% in the second security.

Slightly lower than 66%, so this is the min variance portfolio

$$\text{Var}(R_p) \approx 0.002795399 \Rightarrow \text{SP}(R_p) \approx 5.29\%$$

Sorry, I messed up on this when I did it in class!!

# Portfolio Construction Python Code

```
# Import modules
import numpy as np, pandas as pd, matplotlib.pyplot as plt
from scipy.optimize import minimize_scalar

# Use LaTeX
plt.rcParams['text.usetex'] = True

# Use Seaborn style
plt.style.use('seaborn')

# Define expected returns
mu1, mu2 = 0.01, 0.005

# Define standard deviations
sigma1, sigma2 = 0.10, 0.06

# Define correlation
rho = 0.25

# Create function for portfolio expected return
def port_mu(w):

    return mu1 * w + mu2 * (1 - w)

# Create function for portfolio standard deviations
def port_sigma(w):

    # Calculate variance...
    # ... uncorrelated variance
    var = w**2 * sigma1**2 + (1 - w)**2 * sigma2**2
    # ... variance due to correlation
    var += 2 * w * (1 - w) * rho * sigma1 * sigma2
    # Take square root to obtain standard deviation
    return np.sqrt(var)
```

# Portfolio Construction Python Code

```
# Get weights for minimum variance portfolio; no short positions
w_min_var = minimize_scalar(port_sigma, bounds = (0, 1), method = 'bounded').x

# Define number of samples
samples = 100

# Create data frame to save results
port_results = pd.DataFrame(index = range(samples), columns = [ 'mu' , 'sigma' ])

# Get weights to loop over
wt_vals = np.linspace(w_min_var, 1.0, samples)

# Calculate efficient frontier
for i, w in enumerate(wt_vals):

    # Calculate expected return
    mu = port_mu(w)

    # Calculate standard deviation
    sigma = port_sigma(w)

    # Save results
    port_results.loc[i, [ 'mu' , 'sigma' ]] = mu, sigma
```

# Portfolio Construction Python Code

```
# Plot efficient frontier
plt.plot(port_results['sigma'], port_results['mu'], label = 'Efficient Frontier')

# Add dot for minimum variance portfolio
plt.scatter(port_results.loc[0, 'sigma'], port_results.loc[0, 'mu'], label = 'Minimum Variance', color = 'green')

# Add dot for maximum expected return
plt.scatter(port_results.loc[samples - 1, 'sigma'], port_results.loc[samples - 1, 'mu'], label = 'Maximum Expected Return', color = 'red')

# Add legend
plt.legend()

# Add x-label
plt.xlabel(r'$\sigma$', fontsize = 15)

# Add y-label
plt.ylabel(r'$\mu$', fontsize = 15)

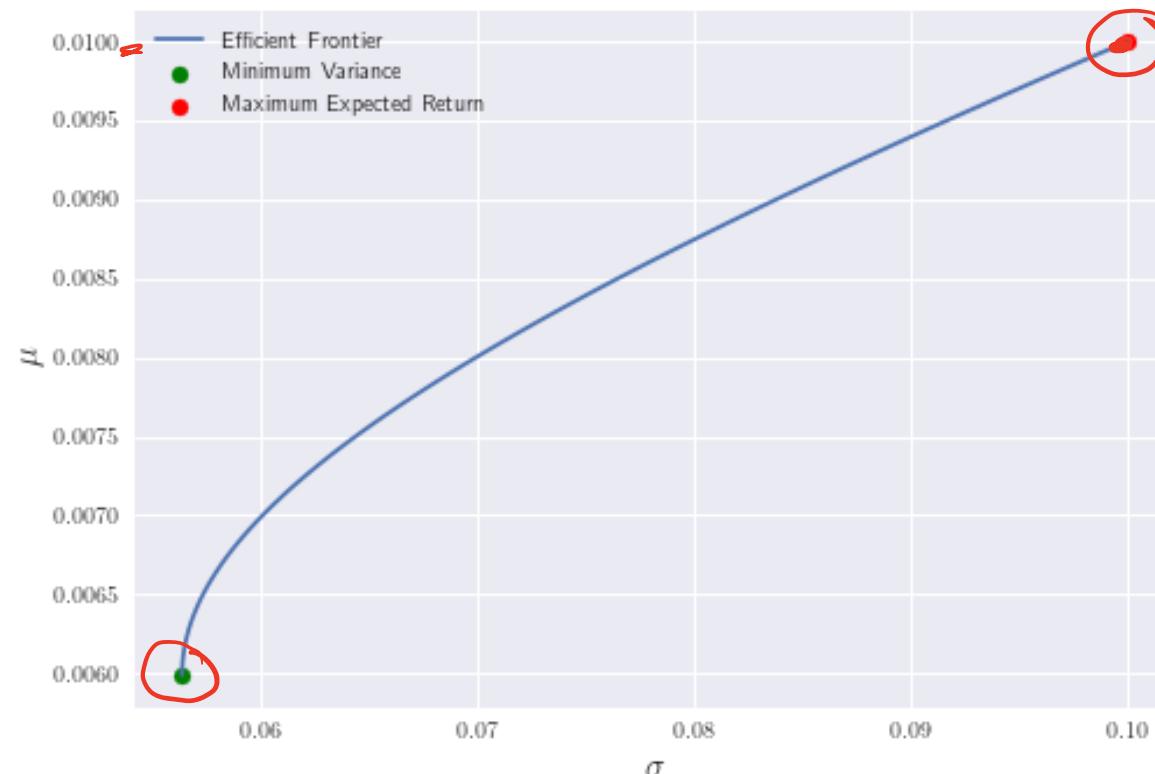
# Save the figure
plt.savefig(path + r'ex3-4.png')

plt.show()

print(r'The minimum variance portfolio has respective weights in',
      r'assets 1 and 2 of',
      f'{100 * w_min_var:.1f}% and {100 - 100 * w_min_var:.1f}%.',
      f'The standard deviation of the minimum variance portfolio is {100 * port_sigma(w_min_var):.1f}%.')
```

# Portfolio Construction Result

The minimum variance portfolio has respective weights in assets 1 and 2 of 19.8% and 80.2%. The standard deviation of the minimum variance portfolio is 5.6%.



# Markov Inequity

## Theorem (Markov Inequity)

Suppose  $X$  is a random variable such that  $P(X \geq 0) = 1$ . Then for each real number  $t > 0$ ,

$$P(X \geq t) \leq \frac{E[X]}{t}.$$

$$P(X \geq t) \leq \frac{E[X]}{t}$$

Proof  $P(X \geq 0) = 1$

$$\begin{aligned}
 E[X] &= \int_0^{\infty} x f(x) dx = \int_0^t x f(x) dx + \int_t^{\infty} x f(x) dx \\
 &\geq \int_t^{\infty} x f(x) dx \quad x \geq t \text{ for } x \in [t, \infty) \\
 &\geq \int_t^{\infty} t f(x) dx \\
 &= t \int_t^{\infty} f(x) dx \\
 &= t P(X \geq t)
 \end{aligned}$$

$$\Rightarrow E[X] \geq t P(X \geq t)$$

$$\Rightarrow P(X \geq t) \leq \frac{E[X]}{t}$$

# Chebyshev Inequality

## Theorem (Chebyshev Inequality)

Let  $X$  be a random variable for which  $\text{Var}(X)$  exists. Then for every number  $t > 0$ ,

$$P(|X - E[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$

Consider  $\gamma = |X - E[X]|^2$

$$P(\gamma \geq t^2) \leq \frac{E[\gamma]}{t^2} \Rightarrow P(|X - E[X]|^2 \geq t^2) \leq \frac{E[(X - E[X])^2]}{t^2}$$

$$\Rightarrow P(|X - E[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

# Mean and Variance of Sample Mean

## Theorem

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution with mean  $\mu$  and variance  $\sigma^2$ . Let  $\bar{X}_n$  be the sample mean. Then  $E[\bar{X}_n] = \mu$  and  $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$ .

$$\begin{aligned} E[\bar{X}_n] &= E\left[\frac{1}{n} \sum_{k=1}^n X_k\right] = \frac{1}{n} \sum_{k=1}^n E[X_k] = \frac{1}{n} \sum_{k=1}^n \mu = \mu \\ \text{Var}\left(\frac{1}{n} \sum_{k=1}^n X_k\right) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \quad \leftarrow \begin{array}{l} \text{independent} \\ \text{samples} \end{array} \quad \text{Cov}(X_i, X_j) = 0 \text{ for } i \neq j \\ &= \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) \\ &= \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{n \sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

# Mean and Variance of Sample Mean

## Example

Suppose  $E[X] = 0$  and  $\text{Var}(X) = 1$ . Use Chebyshev Inequality to find a minimum value of  $n$ , so that  $P(|\bar{X}_n| \geq 0.5) \leq 0.01$ ?

$$P(|Y - E[Y]| > t) \leq \frac{\text{Var}(Y)}{t^2}$$

In our case,

$$\begin{aligned} P(|\bar{X} - 0| \geq 0.5) &= P(|\bar{X} - E[\bar{X}]| \geq 0.5) \\ &\leq \frac{\text{Var}(\bar{X})}{0.5^2} \\ &= \frac{1/n}{0.25} \\ &= \frac{4}{n} \end{aligned}$$

$$E[X] = 0$$

$$E[\bar{X}] = E[X] = 0$$

$$\begin{aligned} \text{Var}(\bar{X}) &= \frac{\text{Var}(X)}{n} \\ &= \frac{1}{n} \end{aligned}$$

$$\Rightarrow P(|\bar{x}| \geq 0.5) \leq \frac{4}{n} \leq 0.01$$

$$\Rightarrow n \geq \frac{4}{0.01} = 400$$

Under these assumptions we must sample 400 observations.

# Convergence in Probability

## Definition

A sequence  $X_1, X_2, \dots$  of random variables **converges in probability** to  $b$  if for each  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|X_n - b| < \epsilon) = 1.$$

We often write  $X_n \xrightarrow{p} b$  to denote that the sequence converges in probability to  $b$ .

# Law of Large Numbers

## Theorem (Law of Large Numbers)

Suppose that  $X_1, X_2, \dots, X_n$  form a random sample from a distribution with mean  $\mu$  and finite variance. Let  $\bar{X}_n$  denote the sample mean. Then

$$\bar{X}_n \xrightarrow{p} \mu.$$

# Law of Large Numbers Example

## Example

Use an exponential distribution with  $\lambda = 1$  to illustrate the Law of Large Numbers. Generate 100,000 samples each of size  $n$ , for  $n = 100, 500, 1000$ , and 10,000.

# Law of Large Numbers Example Python Code

```
# Import modules
import numpy as np, matplotlib.pyplot as plt, time
from scipy.stats import expon

# Start the clock!
start_time = time.perf_counter()

# Set random seed
np.random.seed(0)

# Use LaTeX and Seaborn style
plt.rcParams['text.usetex'] = True
plt.style.use('seaborn')

# Choose n-values and set number of trials
n_vals, trials = [100, 500, 1000, 10_000], 100_000
# Set up subplots
fig, ax = plt.subplots(1, 4, sharex=True, sharey=True,
                      figsize=(12, 4), dpi=125)
for i, n in enumerate(n_vals):
    # Generate numbers of dimension trials x n
    # Scale is 1/lambda; default is 1 so unnecessary
    X = expon.rvs(scale = 1/1, size = (trials, n))
    # Take mean of each row and make histogram
    ax[i].hist(X.mean(axis=1), bins = int(np.log(n)), density = True)
    ax[i].vlines(x = 1, ymin = 0, ymax = 35, linestyle = 'dashed',
                 color = 'red', label = 'True Mean')
    # Give each histogram a title
    ax[i].title.set_text(f'$n = {n}$')
    # Add legend
    ax[i].legend()
    # Add annotations
    if n == 100:
        ax[i].text(1.5, 35, '100', color='blue', rotation=45, fontstyle='italic')
        ax[i].text(1.5, 30, 'Sample', color='blue', rotation=45, fontstyle='italic')
        ax[i].text(1.5, 25, 'means', color='blue', rotation=45, fontstyle='italic')
```

# Law of Large Numbers Example Python Code

```
# Clear up a little RAM
del X

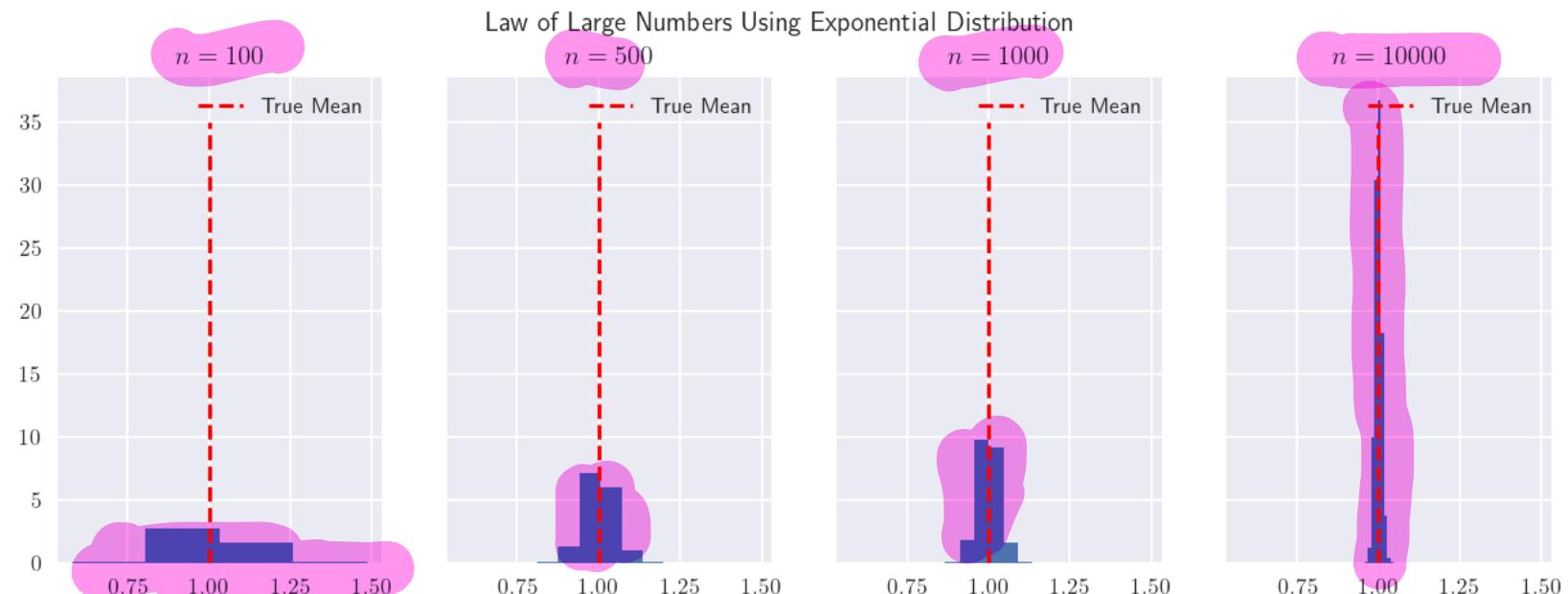
# Give the entire figure a title
fig.suptitle('Law of Large Numbers Using Exponential Distribution')

# Save the figure
plt.savefig(path + r'ex3-5.png')

plt.plot()

# When my programs run slowly, I like to monitor the time it takes
print(f'This program took {time.perf_counter() - start_time:.3f} seconds.')
```

# Law of Large Numbers Result



# Convergence in Distribution

## Definition

A sequence  $X_1, X_2, \dots$  of random variables with cdfs  $F_1, F_2, \dots$  **converge in distribution** to the random variable  $X$  with cdf  $F$  if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

for every number  $x$  at which  $F$  is continuous.

We often write  $X_n \xrightarrow{d} X$  to denote converges in distribution.

# Empirical CDFs

## Definition

Suppose we have observed data  $x_1, x_2, \dots, x_n$  which are sampled from the same distribution. The empirical cdf given these data is

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_k(x),$$

where

$$\mathbf{1}_k(x) = \begin{cases} 1, & x_k \leq x \\ 0, & \text{otherwise.} \end{cases}$$

In Python, it's very easy to code the empirical cdf:

```
ecdf = lambda x, data: np.mean(data <= x)
```

# Empirical CDFs Python Example

## Example

Graph the empirical cdf of a  $\chi^2(1)$  distribution with a sample of size 5, 50, and 1000 as well as the true cdf of the  $\chi^2(1)$ .

```
# Import modules
import numpy as np, matplotlib.pyplot as plt
from scipy.stats import chi2

# Use LaTeX
plt.rcParams['text.usetex'] = True

# Use Seaborn style
plt.style.use('seaborn')

# Set random seed
np.random.seed(0)

# Create empirical cdf
ecdf = lambda x, data: np.mean(data <= x)

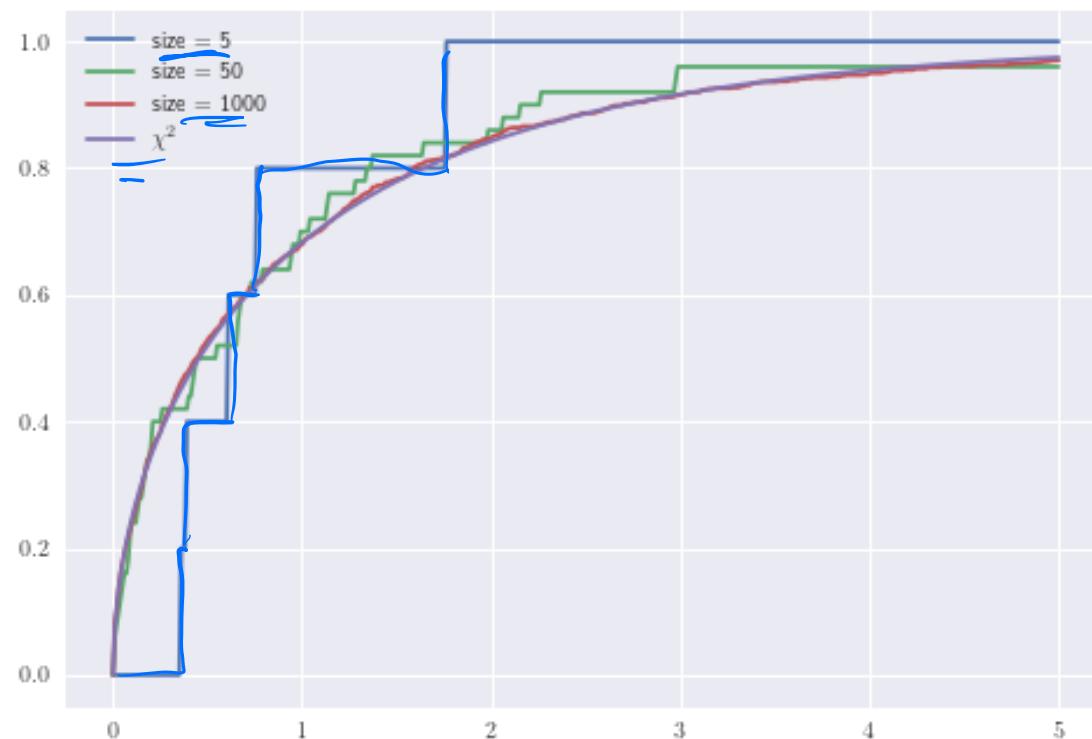
# List of sample sizes
sample_sizes = [5, 50, 1000]

# Create x-values for plot
x_vals = np.linspace(0, 5, 500)
```

# Empirical CDFs Python Example

```
# Loop over the sample sizes
for sample_size in sample_sizes:
    # Generate data
    data = chi2.rvs(df = 1, size = sample_size)
    # Get the y-values
    y_vals = [ecdf(x, data) for x in x_vals]
    # Plot values
    plt.plot(x_vals, y_vals, label = f'size = {sample_size}')
# Get y-values; function vectorized
y_vals = chi2.cdf(x_vals, df = 1)
# Plot standard normal cdf
plt.plot(x_vals, y_vals, label = f'$\chi^2$')
# Clear up RAM
del data, x_vals, y_vals
# Show legend
plt.legend()
# Save the figure
plt.savefig(path + r'ex3-6.png')
plt.show()
```

# Empirical CDFs Python Result



# Central Limit Theorem

## Theorem (Central Limit Theorem)

If the random variables  $X_1, X_2, \dots, X_n$  are a random independent sample of size  $n$  from a given distribution with mean  $\mu$  and finite variance  $\sigma^2$ , then

$$\frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, 1).$$

$$SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

$$E[\bar{X}] = E[X_k] = \mu$$

In other words, the sample mean should approximately follow the the distribution  $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$  when  $n$  is large.

# Central Limit Theorem Python Example

## Example

Illustrate the Central Limit Theorem with an exponential distribution,  
using the empirical cdf and means of size 2, 5, and 100.

# Central Limit Theorem Python Code

```
# Import modules
import numpy as np, matplotlib.pyplot as plt
from scipy.stats import expon, norm

# Use LaTeX
plt.rcParams['text.usetex'] = True

# Use Seaborn style
plt.style.use('seaborn')

# Set random seed
np.random.seed(0)

# Create empirical cdf
ecdf = lambda x, data: np.mean(data <= x)

# Create x-values for plot
x_vals = np.linspace(-5, 5, 500)

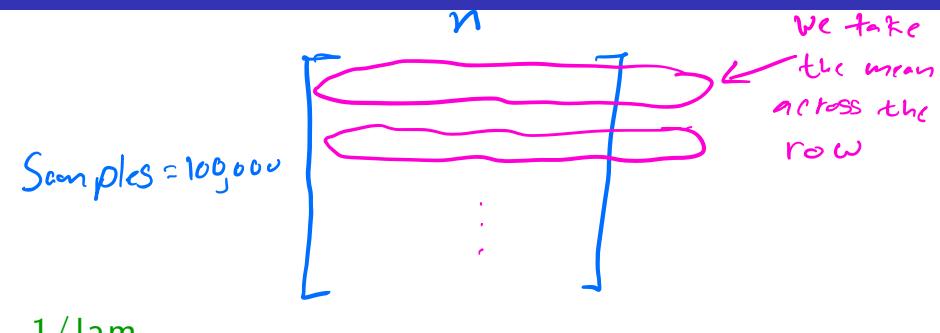
# Create n-values
n_vals = [2, 5, 100]

# Let's do 100,000 samples
samples = 100_000
```

$$F(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{x_1 \leq x_k}(x)$$

# Central Limit Theorem Python Code

```
# Scale is 1/lambda
lam = 1
# Loop over the n-values
for n in n_vals:
    # Generate the data; mean is 1/lam and std is 1/lam
    data = expon.rvs(scale = 1/lam, size = (samples, n)).mean(axis = 1)
    # Change data so it has mean 0 and sd 1; current mean 1/lam and sd 1/(lam * sqrt(n))
    data = np.sqrt(n)/(1/lam) * (data - 1/lam)
    # Get y-values
    y_vals = [ecdf(x, data) for x in x_vals]
    # Plot values
    plt.plot(x_vals, y_vals, label = f'$n = $ {n}')
# Get y-values
y_vals = norm.cdf(x_vals)
# Plot standard normal cdf
plt.plot(x_vals, y_vals, label = 'Normal')
# Show legend
plt.legend()
# Save the figure
plt.savefig(path + r'ex3-7.png')
plt.show()
```



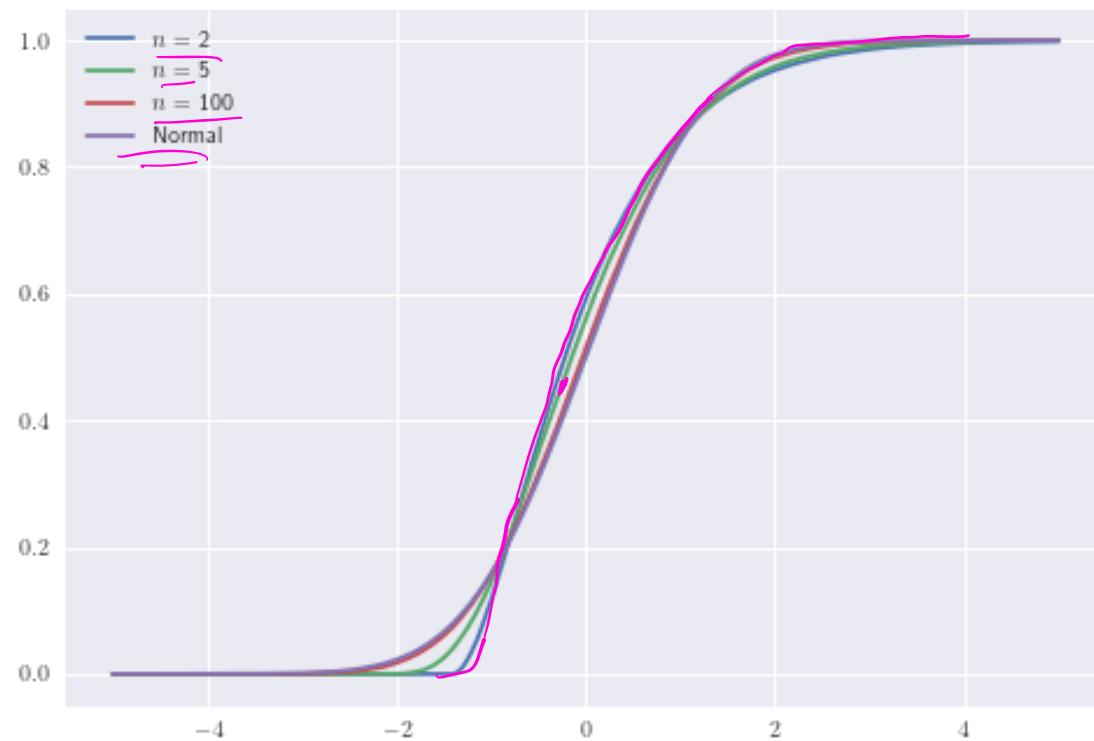
$$X \sim \mathcal{N}(\mu)$$

$$E[X] = \frac{1}{\lambda} \text{ and } \text{var}(X) = \frac{1}{\lambda^2}$$

$$\text{SD}(X) = \sqrt{\frac{1}{\lambda}}$$

$$\text{SE}(\bar{X}) = \frac{1/\lambda}{\sqrt{n}}$$

# Central Limit Theorem Python Result



# Central Limit Theorem Example

## Example

Assume that the weights of individuals are independent and normally distributed with a mean of 160 pounds and a standard deviation of 30 pounds. Suppose that twenty-five people squeeze into an elevator that is designed to hold 4300 pounds. What is the probability that the total weight exceeds the limit?

Sol Total weight  $\geq 4300$  lbs  $\Rightarrow$  Average weight  $\geq \frac{4300}{25} = 172$

$$E[w] = 160 \quad SD(w) = 30 \leftarrow \text{Just one person}$$

$$\Rightarrow E[\bar{w}] = 160 \quad SD(\bar{w}) = \frac{30}{\sqrt{25}} = \frac{30}{5} = 6$$

↑ Standard normal + use Z-Score  
↓ Convert to Standard Normal Distribution

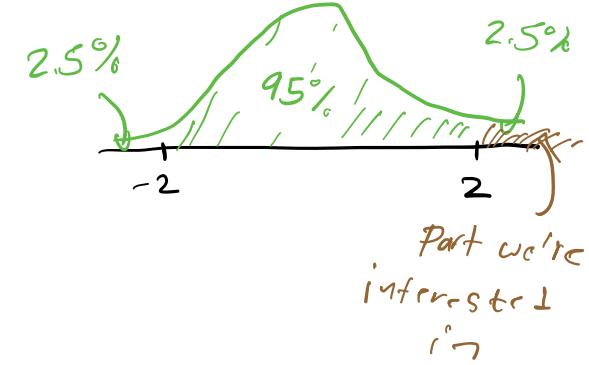
We want to find

$$P(\bar{w} \geq 172) = P(Z \geq \frac{172 - 160}{6})$$

$$= P(Z > \frac{12}{6})$$

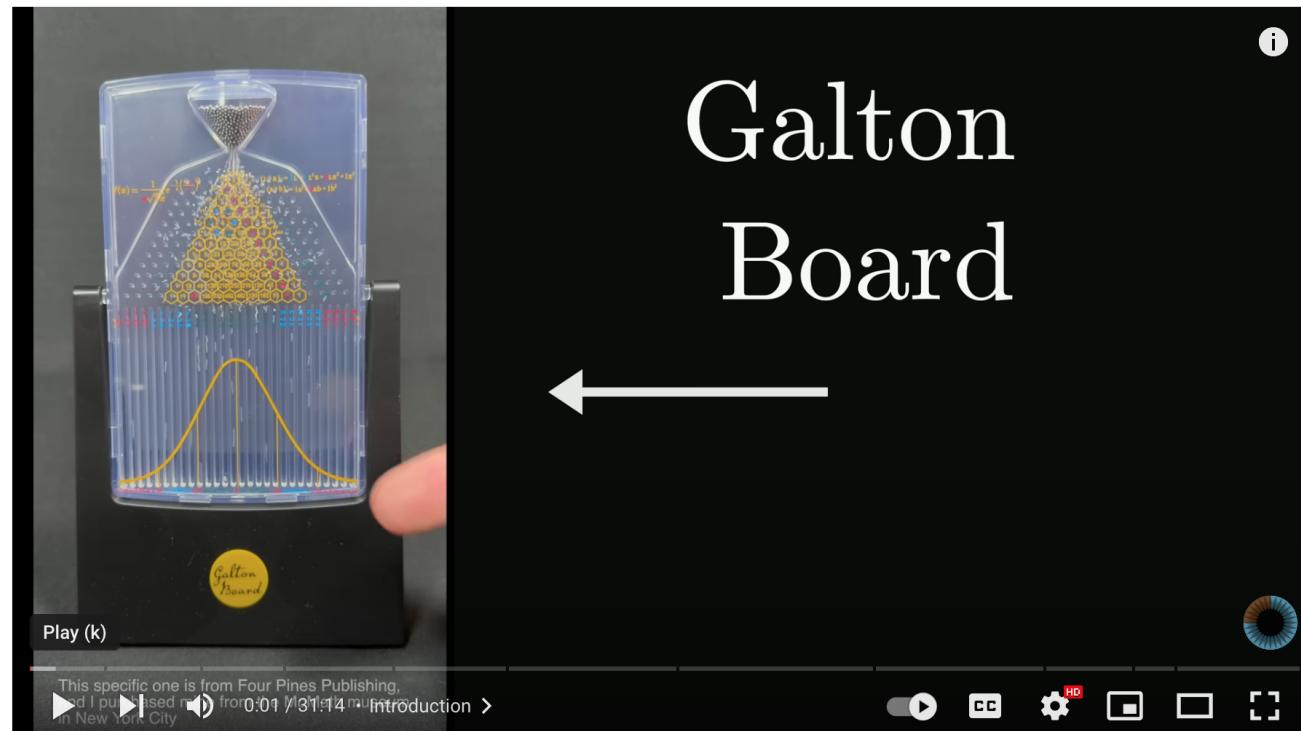
$$= P(Z > 2)$$

$$= 2.5\%$$



# Central Limit Theorem on YouTube

3Blue1Brown has a great video on the Central Limit Theorem  
(<https://youtu.be/zeJD6dqJ5lo>)



But what is the Central Limit Theorem?



Subscribed

48K



Share

Download

...

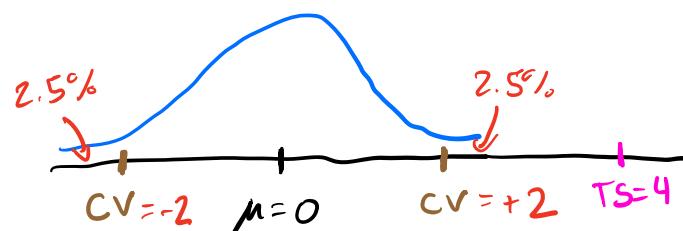
# Hypothesis Testing Example

## Example

You sample the numbers 8, 0.5, 0, and  $-0.5$  from a normal distribution with known variance of 1. Perform a hypothesis test to determine whether the mean of the distribution is statistically different from 0 at a significance level of 5%.

$$H_0: \mu = 0$$

$$H_1: \mu \neq 0$$



Let's calculate our test statistic:

$$\bar{x} = \frac{8 + 0.5 + 0 - 0.5}{4} = 2$$

$$SD(\bar{x}) = \sqrt{1} = 1$$

$$\Rightarrow SD(\bar{x}) = \frac{1}{\sqrt{4}} = \frac{1}{2}$$

So, our test statistic is

$$TS = \frac{2}{1/2} = 4$$

We reject the null hypothesis, and conclude that the mean is not 0 with 100% - 5% = 95% confidence.

# Hypothesis Testing Steps

- 1 State the hypotheses.
- 2 Set the criteria for a decision.
- 3 Compute the test statistic.
- 4 Make a decision.

# State the Hypotheses

## Definition

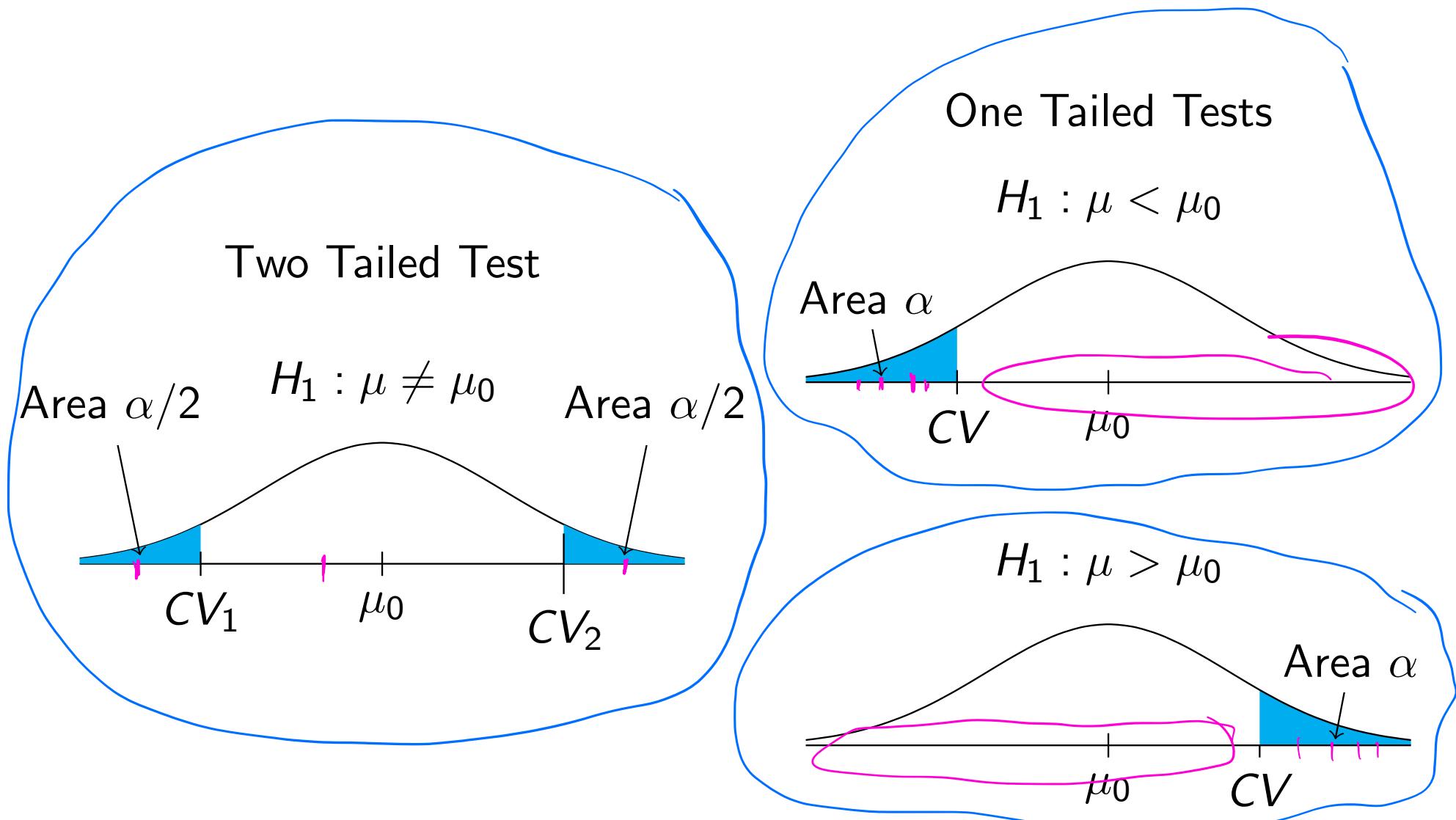
- The **null hypothesis** ( $H_0$ ) is a statement about a population parameter that is assumed to be true. We will test whether the value stated in the null hypothesis is likely to be true.
- An **alternative hypothesis** ( $H_1$ ) is a statement that directly contradicts the null hypothesis by stating that the actual value of a population parameter is less than, greater than, or not equal to the value stated in the null hypothesis.

# Set the Criteria for a Decision

## Definition

**Significance level ( $\alpha$ )** is a probability which denotes how strong the evidence must be before you reject the null hypothesis and conclude that the effect is statistically significant.

# Hypothesis test for $H_0 : \mu = \mu_0$



# Compute the Test Statistic

## Definition

A **test statistic** quantifies the observed data in a way that would distinguish the null from the alternative hypothesis.

# Make a Decision

## Definition

The **p-value** is the probability of an observation being at least as extreme as the test statistic assuming the null hypothesis is true.

$$p\text{-value} = P(|Z| > 4) \leftarrow \begin{array}{l} \text{This would be} \\ \text{the p-value} \\ \text{in the example} \\ \text{we did} \end{array}$$

$p\text{-value} < \alpha \Rightarrow$  Reject null hypothesis

$p\text{-value} > \alpha \Rightarrow$  Fail to reject null hypothesis

Watch out: Many modules state the p-value assuming a two tailed hypothesis test.

# Make a Decision Cont.

- If the  $p$ -value is greater than or equal to  $\alpha$ , we *fail to reject the null hypothesis*.
- If the  $p$ -value is less than  $\alpha$ , we *reject the null hypothesis*.

# Types of Error

		Decision	
		Fail to Reject Null	Reject Null
Ground Truth	Null True	Correct $1 - \alpha$	Type I Error $\alpha$
	Null False	Type II Error $\beta$	Correct $1 - \beta$

# Confidence and Power

## Definition

- The **confidence level** is  $1 - \underline{\alpha}$ .
- The **power** is  $1 - \underline{\beta}$ .

5% significance  $\Leftrightarrow$  95% confidence

# Types of Hypothesis Tests

Incomplete list of types of hypothesis tests.

Name	Test Statistic	Comments
One-sample $z$ -test	$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$	Normal population or large $n$ and known $\sigma$ . The value $\mu_0$ is the mean assuming the null hypothesis.
One-sample $t$ -test	$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$	Normal population or large $n$ and unknown $\sigma$ . The degrees of freedom are $df = n - 1$
One-proportion $z$ -test	$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	$p_0$ is the proportion under the null hypothesis and $\min\{n \cdot p_0, n \cdot (1 - p_0)\} > 10$

# Notation

	Population	Sample
Mean	$\mu = E[X]$	$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$
Variance	$\sigma^2 = E[(X - \mu)^2]$	$s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$
Standard Deviation	$\sigma = \sqrt{E[(X - \mu)^2]}$	$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$

# Sample Standard Deviation

$$s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$$

$\bar{x}$  is closer to observed data than  $\mu$ , so  
small than  $\sum_{k=1}^n (x_k - \mu)^2$

$$\sum_{k=1}^n (x_k - \bar{x})^2$$

# Sample Standard Deviation Example

## Example

Find the distribution of the biased and ~~unbiased~~ standard deviation when ten elements are sampled from  $\mathcal{N}(0, 1^2)$ .

```
# Import modules
import numpy as np, matplotlib.pyplot as plt
from scipy.stats import norm

# Use LaTeX
plt.rcParams['text.usetex'] = True

# Use Seaborn style
plt.style.use('seaborn')

# Set random seed
np.random.seed(0)

# Define number of trials
trials = 100_000

# Generate normal rvs with mean 0 and variance 1
Z = norm.rvs(loc = 0, scale = 1, size = (trials, 10))

# Calculate the sd without bias correction
sd0 = np.apply_along_axis(np.std, 1, Z, ddof = 0)

# Calculate the sd with bias correction
sd1 = np.apply_along_axis(np.std, 1, Z, ddof = 1)
```

$$\sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}$$

$$\sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$$

# Sample Standard Deviation Example Cont.

```
# Set up subplots
fig, ax = plt.subplots(1, 2, sharex = True, sharey = True, figsize = (15, 5), dpi = 125)
ax[0].hist(sd0, bins = int(np.sqrt(trials)), density = True)
ax[0].axvline(x = 1, color = 'red', label = 'True')
ax[0].axvline(x = np.mean(sd0), color = 'orange', label = 'Mean SD')
ax[0].legend()
ax[0].title.set_text(r'$\displaystyle \sqrt{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2}$')
ax[1].hist(sd1, bins = int(np.sqrt(trials)), density = True)
ax[1].axvline(x = 1, color = 'red', label = 'True')
ax[1].axvline(x = np.mean(sd1), color = 'orange', label = 'Mean SD')
ax[1].legend()
ax[1].title.set_text(r'$\displaystyle \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2}$')

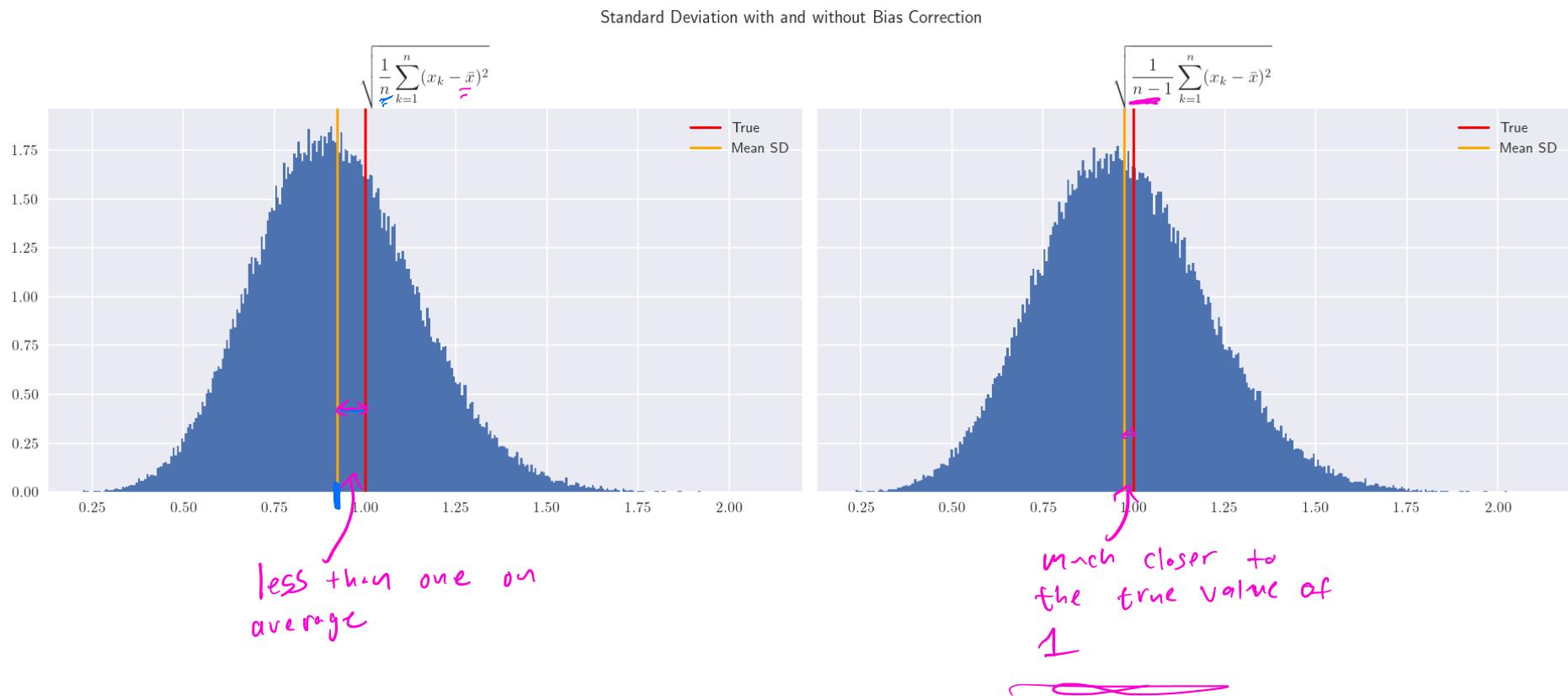
# Give entire figure title
fig.suptitle('Standard Deviation with and without Bias Correction')

# Add padding
fig.tight_layout(pad = 1)

# Save the figure
plt.savefig(path + r'ex3-8.png')

plt.show()
```

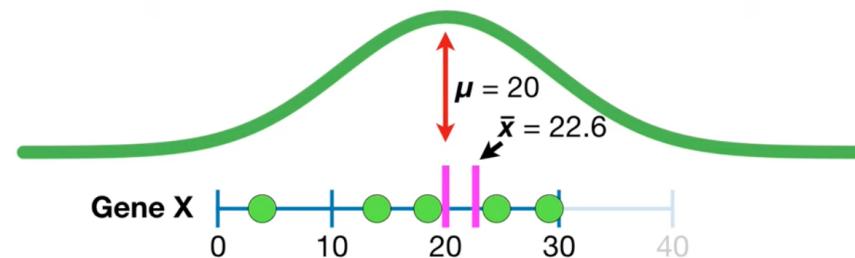
# Sample Standard Deviation Example Result



# Sample Variance on YouTube

There's a StatQuest about sample variance

(<https://www.youtube.com/watch?v=sHRBg6BhKjI>)!



$$\frac{\sum(x - \bar{x})^2}{n} = 105 < \frac{\sum(x - \mu)^2}{n} = 112$$

So far we have seen two simple examples where using the sample mean and dividing by  $n$  underestimated the variances we got with the population mean.

## Why Dividing By N Underestimates the Variance



StatQuest ...  
1.18M subscribers

Join

Subscribed

4.1K



Share

...

# Student's $t$ -Distribution

## Definition

Let  $x_1, x_2, \dots, x_n \sim \mathcal{N}(\mu, \sigma^2)$  be independent and identically distributed samples. Then

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

follows a  $t$ -distribution with  $n - 1$  degrees of freedom. We often write

$$T \sim t_{n-1}.$$

For  $\underline{n > 30}$  the  $t$ -distribution is approximately equal to a normal distribution. As a result, this distribution is only relevant for small samples.

# Hypothesis Testing with Unknown Variance

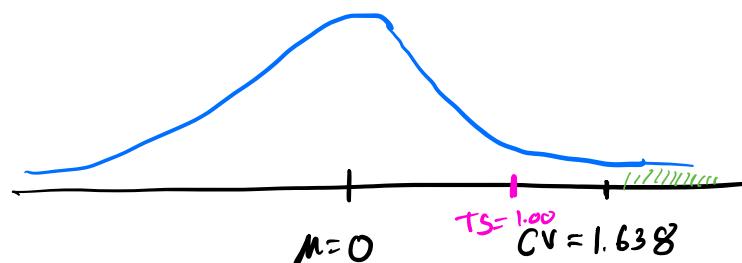
## Example

You sample the numbers 8, 0.5, 0, and  $-0.5$  from a normal distribution. Perform a hypothesis test to determine whether the mean of the distribution is larger than 0 at a significance level of 10%. Note: If  $T \sim t_3$ , then  $P(T > 1.638) \approx 10\%$ .

$$H_0: \mu = 0$$

$$H_1: \mu > 0$$

$$\alpha = 10\% =$$



Let's find our test statistic:

$$\bar{x} = \frac{8 + 0.5 + 0 - 0.5}{4} = 2$$

$$\begin{aligned} SD(x) &= \sqrt{\frac{1}{3} [(8-2)^2 + (0.5-2)^2 + (0-2)^2 + (-0.5-2)^2]} \\ &= \sqrt{\frac{1}{3} (6^2 + 1.5^2 + 2^2 + 2.5^2)} \\ &\approx 4.02 \end{aligned}$$

$$SD(\bar{x}) \approx \frac{4.02}{\sqrt{4}} = \frac{4.02}{2} = 2.01$$

$$\Rightarrow t = \frac{2 - 0}{2.01} \approx 1.00$$

$\Rightarrow$  Fail to reject the null hypothesis that  $\mu = 0$ .

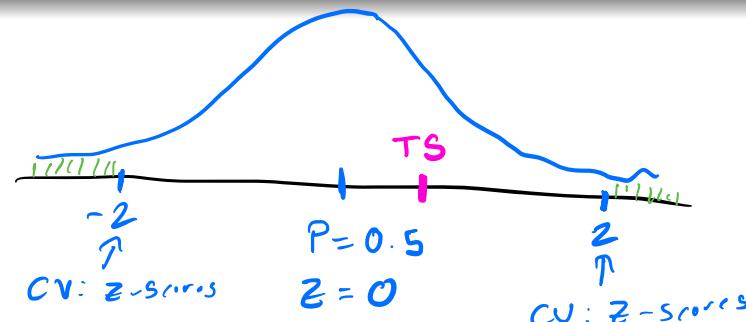
# Hypothesis Testing Proportions

# Example

You flip a coin 100 times and it lands heads 65 times. Perform a hypothesis test to determine whether the probability of landing heads is statistically different from 50% at a 95% confidence level.

$$H_0: P = 0.5$$

$H_1: P \neq 0.5$



$$\alpha = 100\% - 95\% \\ = 5\%$$

$$100 \cdot 0.5 = 50 > 10 \quad \checkmark$$

ан

$$100 \cdot (1 - 0.5) = 100 \cdot 0.5 = 50 > 10 \quad \checkmark$$

Let's calculate our test statistic

$$\hat{P} \approx \frac{65}{100} = 0.65$$

$$Z = \frac{\hat{P} - P_0}{\sqrt{n P_0 (1 - P_0)}}$$

$$TS: Z = \frac{0.65 - 0.5}{\sqrt{\frac{100 \cdot 0.5 \cdot 0.5}{100}}} \quad \checkmark$$

$X$  is binomial distributed

$$= \frac{0.15}{0.5}$$

$$= 0.3$$

$\Rightarrow$  We fail to reject the null hypothesis the  $p=0.5$  at a 95% confidence level.

# Confidence Interval

The population parameter lies in this interval with probability  $1 - \alpha$ .

- Known Variance:

$$\left( \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right).$$

- Unknown Variance:

$$\left( \bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right).$$

Degrees of freedom are  $n - 1$ .

- Proportion:

$$\left( \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right).$$