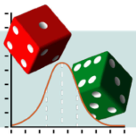


An Intuitive Introduction to Probability – *Decision-Making in an Uncertain World*

1- PROBABILITY THEORY

INTRODUCTION

WHY STUDY PROBABILITIES?

1. **Over the long term, the S&P500 outperforms corporate bonds, which in turn outperforms government bonds. This is due to uncertainty and risk.**
2. **Oswald Grübel (former CEO of Credit Suisse and UBS) said referring to banks: "The more risk you are willing to take, the more profitable you will be."**
3. **Patrick Leach (author of "Why Can't You Just Give Me The Number?") on the other side:**
   1. "Business adds value because of the existence of uncertainty."
   2. "... all value generated by business executives comes - directly or indirectly - from how they manage uncertainty. Without uncertainty, a share of a company's stock is effectively a bond, with guaranteed cash flows. Guaranteed bonds don't need management. But stocks (or rather, companies issuing stock) certainly do."
4. **Managers must take important decisions in uncertain business environments.**
5. **The most fundamental concept of dealing with risk and uncertainty? Probabilities!**

SOME TERMINOLOGY

Random Experiment
A process leading to an uncertain outcome.

State Space
The state space is the collection of all possible outcomes of a random experi**ment, usually denoted by** S**.**

Basic Outcome
A possible outcome of a random experiment**.**

Event

An event is a subset of basic outcomes. Any event which consists of a single outcome in the state space is called a simple event.

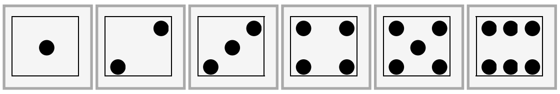
Probability
A probability is a measure for how likely a**n event of a random experiment is.**

***Notation:* The probability on an event A is usually denoted by $P(A)$.**

EXAMPLES

Before we can talk about probabilities, we need a good understanding of random experiments, state spaces, basic outcomes and events. Let us look at some simple examples.

EXAMPLE 1: THE SIX-SIDED DIE



BASIC OUTCOME

A basic outcome of rolling a six-sided die is for example a ‘6’.

RANDOM EXPERIMENT

Rolling a six-sided die is a random experiment, since the outcome is uncertain. However, once the die is rolled, the outcome can be precisely assessed.

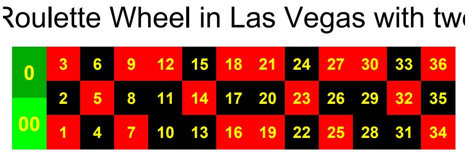
STATE SPACE

The state space S for rolling a six-sided die is equal to $\{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$.

EVENT

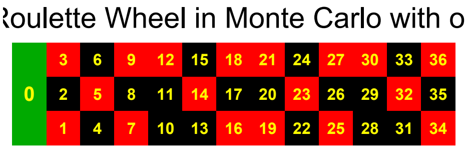
The event $A\ =\ '\mathit{score\ is\ smaller\ than\ 4}'$ is $\{1,\ 2,\ 3\}$. The event $B\ =\ '\mathit{score\ is\ 8}'$ is equal to $\emptyset$ (the empty set), because this is impossible with a six-sided die.

EXAMPLE 2: ROULETTE WHEEL IN MONTE CARLO



Roulette Wheel in Las Vegas with tw

0	3	6	9	12	15	18	21	24	27	30	33	36
	2	5	8	11	14	17	20	23	26	29	32	35
00	1	4	7	10	13	16	19	22	25	28	31	34



Roulette Wheel in Monte Carlo with o

	3	6	9	12	15	18	21	24	27	30	33	36
0	2	5	8	11	14	17	20	23	26	29	32	35
	1	4	7	10	13	16	19	22	25	28	31	34

While a roulette wheel in Las Vegas has two zeros, a roulette wheel in Monte Carlo only has a single zero. Besides this difference, both roulette wheels are the same.

Every outcome on a roulette wheel is equally likely, because each basket at each number has exactly the same size. Therefore it is just as likely that the ball lands on the single zero as on any other number on the roulette wheel.

RANDOM EXPERIMENT

Rolling a ball in a roulette wheel is a random experiment, since the outcome is uncertain. However, once the ball lands in a basket, the outcome can be precisely assessed.

STATE SPACE

The state space S for a roulette wheel in Monte Carlo is equal to $\{1,\ 2,\ \ldots,\ 35,\ 36,\ 0\}$.

BASIC OUTCOME

A basic outcome of a game of roulette could be for example ‘18’, ‘0’, ‘black’ or ‘odd and red’. Each outcome represents exactly one number, has a color and is even or odd.

EVENT

The event $A\ =$ ‘the ball lands on a number smaller than 4’ is $\{1,\ 2,\ 3\}$. The event $B\ =\ '\mathit{the\ ball\ lands\ on\ red\ and\ black}\,'$ is $\emptyset$ (the empty set), because this is impossible. This can be seen on the map of the roulette wheel above. There are no numbers that are red and black at the same time. Now we can turn to the definition of a probability. In fact, there are three possible definitions that are applied in practice.

PROBABILITY DEFINITIONS

CLASSICAL PROBABILITY

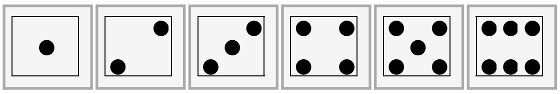
Classical Probability

$$P(A)\ =\ \frac{\mathit{Number\ of\ outcomes\ that\ satisfy\ the\ event}}{\mathit{Total\ number\ of\ outcomes\ in\ the\ state\ space}}$$

Remark: **This method assumes all outcomes in the sample space to be equally likely to occur.**

EXAMPLES

EXAMPLE 1 CONT'D: THE SIX-SIDED DIE



In this example, we throw a six-sided die. Recall that the state space S is equal to $\{1, 2, 3, 4, 5, 6\}$.

QUESTION

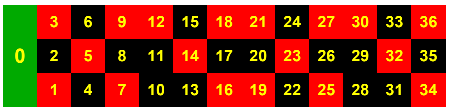
What is the probability of getting a '6'?

Solution.

All outcomes for throwing a six-sided die are equally likely (one can say in this case that the die is 'fair'). Therefore we can use classical probability to determine this particular probability. In this example, S, i.e. the total number of outcomes, is equal to six. There is exactly one outcome that satisfies the event of getting a '6' and this is rolling a '6'. Hence, the probability of getting a '6' is exactly equal to $\frac{1}{6}$ (≈ 0.167).

EXAMPLE 2 CONT'D: ROULETTE WHEEL IN MONTE CARLO

ulette Wheel in Monte Carlo with a si



0	3	6	9	12	15	18	21	24	27	30	33	36
	2	5	8	11	14	17	20	23	26	29	32	35
	1	4	7	10	13	16	19	22	25	28	31	34

Recall that the state space S for a roulette wheel in Monte Carlo is equal to $\{1, 2, \ldots, 35, 36, 0\}$.

QUESTION

What is the probability that the ball lands on a red number?

Solution.

In this question, S represents the state space that contains all possible outcomes of this random experiment. Every basket on the roulette wheel has the same size and hence all outcomes on the roulette wheel are equally likely. Therefore we can use classical probability to assess this particular probability. In this example, S consists of 37 elements in total. There are 18 outcomes that satisfy the event of getting a red number. Hence, the probability that the ball lands on a red number is equal to $\frac{18}{37}$ (≈ 0.486).

RELATIVE FREQUENCY PROBABILITY

Relative Frequency Probability

$$P(A) = \left(\textit{Number of times the event } A \textit{ occurs in repeated trials}\right) / \left(\textit{Total number of trials in a random experiment}\right)$$

Remark: $P(A)$ **onverges to the true probability in the limit.**

EXAMPLES

EXAMPLE 3: COMBUNOX

Combunox is a drug that contains a combination of oxycodone and ibuprofen. Oxycodone is an opioid pain medication. Ibuprofen is a non steriodal anti-inflammatory drug (NSAID). Combunox works by reducing substances in the body that cause pain, fever, and imflammation. During an experiment, it was reported that 49 out of 923 people vomited after taking Combunox.

QUESTION

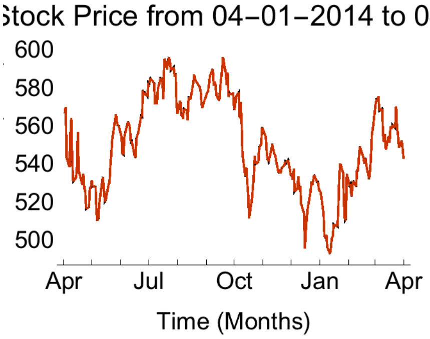
What is the probability of vomiting after taking Combunox?

Solution.

We want to determine the relative frequency that people vomited after taking combunox and therefore we can use relative frequency probability to determine this particular probability. The total number of trials in this random experiment is equal to 923. The number of times that people vomited is equal to 49. The probability of vomiting after taking Combunox is equal to $\frac{49}{923}$ (≈ 5.3 %).

EXAMPLE 4: GOOGLE STOCK PRICE

The end-of-day stock prices of Google are given in the figure below from 04-01-2014 to 04-01-2015 (which is equivalent to roughly 250 trading days).



QUESTION

Assume it is 04-01-2015. Based on the figure above, what is the probability that there will be a stock price increase on the next trading day? *Hint:* first determine the state space.

Solution.

In order to determine the state space initially, we need to determine all possible prices that the Google stock can attain.

- **The minimum price of the stock is zero.**

- **The maximum price cannot be derived from historical stock price data with certainty.**

We can define the following three events for the Google stock price change:

1. **'Up' (denoted by u), in case the stock price increases.**
2. **'Even' (denoted by m), in case the stock price remains unchanged.**
3. **'Down' (denoted by d), in case the stock price decreases.**

Therefore the state space S is equal to $\{u, m, d\}$. Even though we defined the state space, it is impossible to determine the probability for each of the three elements in the state space, let alone for each price that is theoretically possible.

SUBJECTIVE PROBABILITY

Subjective Probability
An individual opinion or belief about the probability of occurence.

EXAMPLES

EXAMPLE 5: STEVE JOBS

Steve Jobs had to use subjective probability in order to predict the probability of a success of the first iPad. There was no data available on which Steve Jobs could rely (except his experience and beliefs) and therefore he could not use relative frequency probability or classical probability. Considering the huge success of the iPad, Steve Jobs assessed this subjective probability quite well.

EXAMPLE 6: WHAT IS YOUR SUBJECTIVE PROBABILITY?

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

QUESTION

Which of the following eight options is more likely for Linda to be?
(1) Linda is a teacher in elementary school.
(2) Linda works in a bookstore and takes Yoga classes.
(3) Linda is active in the feminist movement.
(4) Linda is psychiatric social worker.
(5) Linda is a member of the *League of Women Voters*.
(6) Linda is a bank teller.
(7) Linda is an insurance salesperson.
(8) Linda is a bank teller and active in a feminist movement.

SET THEORY

INTRODUCTION

For simple probability calculations, it is useful to know just a little set theory. Therefore we cover in this section a few basic definitions. This part may be boring, but please stay with us. Below are a few definitions for arbitrary sets A and B.

SET THEORY: DEFINITIONS

Union

The union of two events A and B is the set that contains all the events that are either in A or in B or in both sets.

Notation: $A \bigcup B$**.**

Intersection

The intersection of two events A and B is the set that contains all the events that are both in A and in B.

Notation: $A \bigcap B$**.**

Complement

The complement of an event A with respect to an event B refers to all elements that are in B, but not in A. This is usually denoted by $B \bigcap A^c$ or $B \backslash A$.

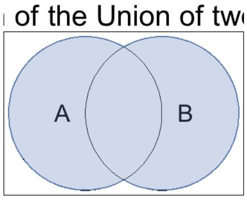
Mutually Exclusive

Two events A and B are said to be mutually exclusive if they have no basic outcomes in common. This is usually denoted by $B \bigcap A = \emptyset$.

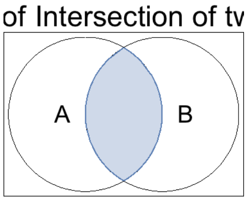
Collectively Exhaustive

Events $A_1,\ A_2,\ \ldots,\ A_n$ are said to be collectively exhaustive **events if $A_1 \bigcup\ A_2 \ldots \bigcup\ A_n = S$, i.e. the events in union completely cover the sample space S.**

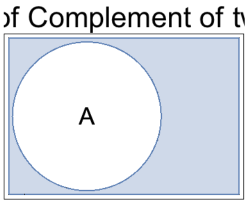
SET THEORY: ILLUSTRATIONS WITH VENN DIAGRAMS



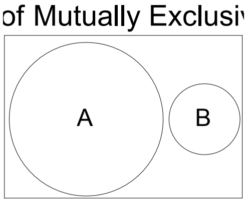
The shaded area is A ∪ B



The shaded area is A ∩ B



The shaded area is A^c



A ∩ B = Ø

EXAMPLES

EXAMPLE 1 CONT'D: THE SIX-SIDED DIE

In this example, we throw a six-sided die. Let A be the event that the outcome is either 1, 2 or 3. Let B be the event that the outcome is an even number, so either 2, 4 or 6.

QUESTION

Determine the following sets:

_ **The *union* of A and B, $A \cup B$.**

_ **The intersection of A and B, $A \cap B$.**

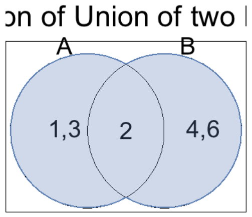
_ **The completement of A, A^c.**

- **The complement of B, B^c.**

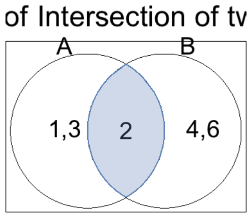
Solution.

First of all, recall that the state space S for throwing a six-sided die is equal to $\{1, 2, 3, 4, 5, 6\}$. Also, both sets A and B are a *subset* of S (that means, both A and B are ‘contained’ inside S). In this example, $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$.

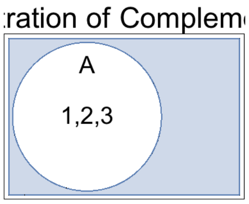
- $A \bigcup B = \{1, 2, 3, 4, 6\}$**, as these are the numbers that are either in** A **or in** B**.**
- $A \bigcap B = \{2\}$**, as this is the only number that is both in** A **and in** B **or both.**
- $A^c = \{4, 5, 6\}$**, as these are all the numbers in** S **that are not in** A**.**
- $B^c = \{1, 3, 5\}$**, as these are all the numbers in** S **that are not in** B**.**



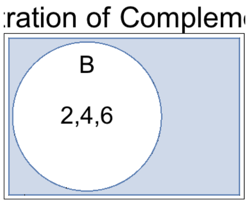
The shaded area is A ∪ B



The shaded area is A ∩ B



The shaded area is A^c



The shaded area is B^c

LAWS OF PROBABILITY

INTRODUCTION

Given a sample space S, the probabilities assigned to events must always satisfy the three laws of probability, which are listed below.

LAWS OF PROBABILITY

Law 1

$P(S) = 1$**, where** S **is the** state space.

Law 2

For any event A, $0 \leq P(A) \leq 1$. The probability of an event can never be negative or larger than 1.

Law 3

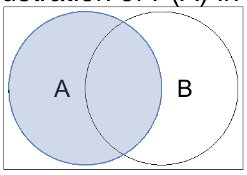
For two disjoint **events** A **and** B **(disjoint means that** $A \cap B = \emptyset$**),**

$P(A \cup B) = P(A) + P(B)$**.**

TRIVIAL IMPLICATION FOR LAWS OF PROBABILITY THEORY

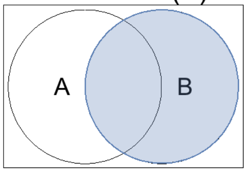
For two events A and B, it always holds that $P(A \cap B) \leq P(A)$ and also that $P(A \cap B) \leq P(B)$. The Venn diagrams below show this trivial implication.

ustration of P(A) in



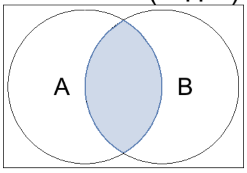
The shaded area is A

ustration of P(B) in



The shaded area is B

ɜtration of P(A ∩ B)



The shaded area is A ∩ B

EXAMPLES

EXAMPLE 6 CONT'D: WHAT IS YOUR SUBJECTIVE PROBABILITY?

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

QUESTION

Which of the following eight options is more likely for Linda to be?
(1) Linda is a teacher in elementary school.
(2) Linda works in a bookstore and takes Yoga classes.
(3) Linda is active in the feminist movement.
(4) Linda is a psychiatric social worker.
(5) Linda is a member of the *League of Women Voters*.
(6) Linda is a bank teller.
(7) Linda is an insurance salesperson.
(8) Linda is a bank teller and active in a feminist movement.

Solution.

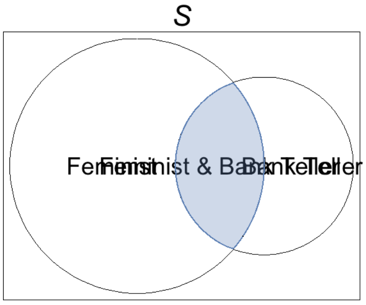
Out of the eight possible events that we have, it is possible to say something about the relative likelihood of the following three events:

(3) Linda is active in the feminist movement.
(6) Linda is a bank teller.
(8) Linda is a bank teller and active in a feminist movement.

First, let A be the event that Linda is a bank teller and B be the event that Linda is active in a feminst movement. By using trivial implication, we get:

- $P(A \cap B) \le P(A)$,
- $P(A \cap B) \le P(B)$.

Therefore the event that Lisa is active in a feminist movement and also a bank teller is less likely than that she is only a bank teller.



INDEPENDENCE

INTRODUCTION

QUESTION

Assume you throw a fair, six-sided die once. You write down the outcome and then you throw the die again.

- **What is the probability of getting a ‘1’ after the first roll?**
- **What is the probability of getting a ‘2’ after the second roll, given that you got a ‘1’ after the first roll?**

Solution.

- **The probability of getting a ‘1’ after the first roll is $\frac{1}{6}$.**
- **The probability of getting a ‘2’ after the second roll is completely unaffected by the outcome of the first roll. In other words, the probability of getting a ‘2’ after the second roll is independent from previous outcomes. Therefore the probability of getting a ‘2’ after the second roll is also equal to $\frac{1}{6}$.**

QUESTION

Assume you throw again a fair six-sided die twice. What is the probability that you first roll a ‘1’ and then a ‘2’?

Solution.

This question is slightly different from the previous question. In this question, we need to calculate $P('1' \cap '2')$. As determined in the previous question, both individual probabilities are equal to $\frac{1}{6}$. We can multiply the individual probabilities, because two subsequent rolls of a die are independent. Therefore we get

$$P('1' \cap '2') = P('1') \cdot P('2') = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \ (\approx 0.0278) .$$

INDEPENDENCE

Two events A and B can be independent in the sense that the outcome of A does not influence the outcome of B and vice versa.

Multiplication Rule

Two events A and B are (statistically) independent if and **only if the probability of both A and B occuring is the product of the probabilities of the two events,** $P(A \cap B) = P\left(A\ and\ B\right) = P(A) \cdot P(B)$**.**

EXAMPLES

EXAMPLE 7: GAME OF CRAPS

Craps is a dice game in which the players make wagers on the outcome of the roll, or a series of rolls, of a pair of dice. Let us assume that there are two dice and that each player can bet on the sum of the two dice. Note that the sum is at least equal to 2 (a ‘1’ and a ‘1’) and at most equal to 12 (a ‘6’ and a ‘6’). The state space S for the sum of the two dice is therefore equal to $\{2,\ 3,\ \ldots,\ 11,\ 12\}$.

QUESTION

- **What is the probability that the sum is equal to the the minimum value (e.g. 2)?**
- **What is the probability that the sum is equal to the maximum value (e.g. 12)?**

Solution.

The outcome of the first die is completely independent from the outcome of the second die. We know that the individual probabilities of getting for example a ‘1’ or a ‘6’ with either of the two dice is equal to $\frac{1}{6}$. Since the outcomes of the two dice are independent, we can use the multiplication rule. We get:

- $P\left(Sum\ equals\ 2\right) = P('1' \cap\ '1') = P('1')\ \cdot\ P('1') = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.
- $P\left(Sum\ equals\ 12\right) = P('6' \cap\ '6') = \frac{1}{36}$

QUESTION

- **What is the probability that the sum is equal to the the median value (e.g. 7)?**
- **What is the probability that the sum is equal to the second largest value (e.g. 11)?**

Solution.

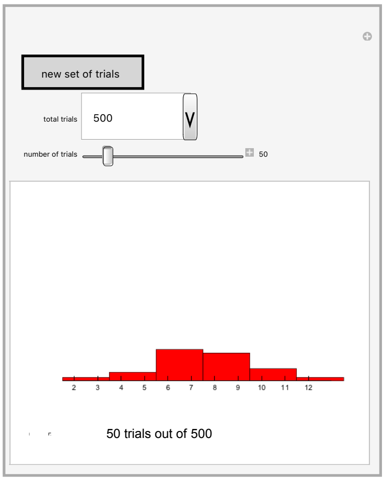
- **The outcome of the first die is** independent **from the outcome of the second die. The outcomes of the first respectively the second die can be: - ‘3 and 4’ *or* ‘4 and 3’ *or* ‘2 and 5’ *or* ‘5 and 2’ *or* ‘1 and 6’ *or* ‘6 and 1’.**

Hence, there are in total six possible combinations of numbers that satisfy the event 'sum equal to 7' and there are 36 possible outcomes in total. Therefore $P\left(Sum\ equals\ 7\right) = \frac{6}{36} = \frac{1}{6}$**.**

- **The outcome of the first die is again** independent **from the outcome of the second die. The outcomes of the first respectively the second die can be:**
  - '6 and 5' ***or*** **'5 and 6'*.***
  Hence, there are in total two possible combinations of numbers that satisfy the event 'sum equal to 11' and there are 36 possible outcomes in total. Therefore $P\left(Sum\ equals\ 11\right) = \frac{2}{36} = \frac{1}{18}$**.**

EXAMPLE 7 CONT'D: GAME OF CRAPS

The probability distribution for the sum of two dice is not uniform (i.e. not each sum is equally likely to occur). We can simulate the outcomes of the two dice multiple times and see how often we get each sum.



total trials 500

number of trials 50

50 trials out of 500

LAW OF LARGE NUMBERS

It is clear from the simulation above that the probability that the sum equals 7 is far greater than the probability that the sum equals 2 or 3. If it were possible to do the simulation infinitly often, all probabilities will converge to their real exact probabilities. If we perform a large number of simulations (for example 1.000 or 10.000), we can already expect it to converge to quite accurate estimations of the true probabilities. This essentially means that in the long run, the relative frequency probability converges to classical probability. This principle is called the law of large numbers.

EXAMPLE 8: APPLE STOCK

Suppose that during the last 250 trading days, the Apple stock went up (denoted by u) on 150 days and went down (denoted by d) on 100 days. There were no days that the stock price of Apple remained unchanged and therefore we omit this event in this example. We assume that the probability of a future up movement respectively a future down movement of the stock price is equal to:

- $P(u) = \frac{150}{250} = 0.6$,

- $P(d) = \frac{100}{250} = 0.4$.

In addition we assume that the performance of the Apple stock price on the current trading day is independent of the performance of the Apple stock price on previous trading days. This means that the probability of an up or down movement from one day to the next is not affected by the number of previous up and down movements, e.g. if the Apple stock went up three days in a row, the probability that it goes up the next day remains unchanged.

Do you think this assumption is reasonable?

QUESTION

In this question, we consider a week from Monday to Friday.

- **What is the probability that the stock price of Apple will increase on each consecutive day in the week?**
- **What is the probability that the Apple stock price will decrease on Monday, but subsequently will increase on Tuesday, Wednesday, Thursday and Friday?**
- **What is the probability that the Apple stock price will decrease on either Monday, Tuesday, Wednesday, Thursday or Friday and will increase on the other four days?**

Solution.

- **There are five trading days in a week, so we need to calculate the probability that there will be five subsequent up movements of the Apple stock price. It is given that each movement of the stock price is independent of the previous one. Therefore, we get by the multiplication rule:**

$$P(u \cap u \cap u \cap u \cap u) \overset{Independence}{=} P(u) \cdot P(u) \cdot P(u) \cdot P(u) \cdot P(u)$$
$$= 0.6 \cdot 0.6 \cdot 0.6 \cdot 0.6 \cdot 0.6 = 0.6^{5} \approx 0.078.$$

- **Similar reasoning as for the previous question, we get by the multiplication rule:**

$$P(d \cap u \cap u \cap u \cap u) \overset{Independence}{=} P(d) \cdot P(u) \cdot P(u) \cdot P(u) \cdot P(u)$$
$$= 0.6 \cdot 0.4 \cdot 0.6^{3} \approx 0.052.$$

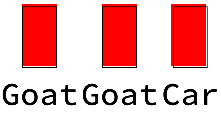
- **The decrease could happen either on Monday, Tuesday, Wednesday, Thursday or Friday. As calculated in the previous question, the probability that the down movement happens on for example Monday is equal to 0.052. This probability does** not **depend on which day the down movement happens! Therefore, by using the multiplication rule for each scenario and summing these up, we get:**

$$P(d \cap u \cap u \cap u \cap u) + P(u \cap d \cap u \cap u \cap u) + P(u \cap u \cap d \cap u \cap u) + P(u \cap u \cap u \cap d \cap u) + P(u \cap u \cap u \cap u \cap d) \approx 5 \cdot 0.052 = 0.26.$$

APPLICATIONS

MONTY HALL PROBLEM

The Monty Hall problem is based on the game shows ‘Let’s make a Deal’ and named after the host Monty Hall. There are three doors. Behind two doors is a goat and behind one door there is a car. The game contestant does not know behind which door the car is and he must find the car in order to win it. The game contestant can choose one door. The host, who knows behind which door the car is, then opens a door not chosen by the contestant and with a goat behind. The game contestant is then asked if he wants to stay with his initial door or switch to the other door that is still closed.



QUESTION

Should the game contestant switch doors or stick to his original choice?

Solution.

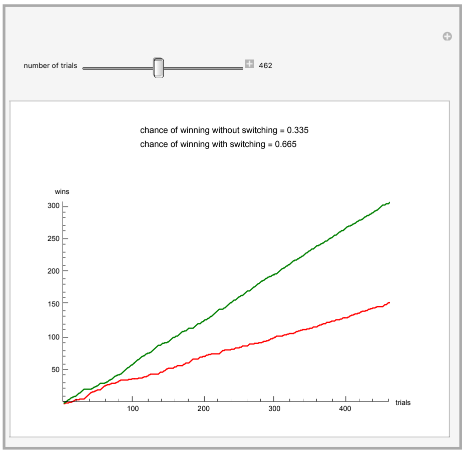
We only give the idea of the correct solution. The formal solution is beyond the scope of this lecture. We have:

- $P(\text{'}\mathit{car\ behind\ door\ number\ 1}\text{'}) = \frac{1}{3}$,
- $P(\text{'}\mathit{car\ behind\ door\ number\ 2}\text{'}\ \mathit{or}\ \text{'}\mathit{car\ behind\ door\ number\ 3}\text{'}) = \frac{2}{3}$.

Therefore switching doors means the host automatically ‘pushes’ the game contestant out of one wrong door. We get:

$P(\text{'}\mathit{winning}\text{'}\ \mathit{given}\ \text{'}\mathit{switching}\text{'}) = \frac{2}{3}$ and

$P(\text{'}\mathit{winning}\text{'}\ \mathit{given}\ \text{'}\mathit{not\ switching}\text{'}) = \frac{1}{3}$.



number of trials 462

chance of winning without switching = 0.335

chance of winning with switching = 0.665

THE WORLD’S MOST INTELLIGENT WOMAN

The Monty Hall Problem became famous when it was addressed as a question to Marilyn vos Savant’s “Ask Marilyn” column in Parade magazine in 1990. In 1991, she was listed in the Guinness Book of World Records Hall of Fame for “Highest I.Q.”. She answered a question similar to the Monty Hall problem and came up with the solution that switching doors is the only rational choice the game contestant can make. Her arguments were heavely critized by Professors and PhD’s in Mathematics. Some quotes:

- **- Mary Jane Still, Professor at Palm Beach Junior College:**
  “*Our math department had a good, self-righteous laugh at your expense*”
- **- Robert Sachs, Professor of Mathematics at George Mason University.**
  “*You blew it!*”
- **- E. Ray Bobo, Professor of Mathematics at Georgetown University.**
  “*You are utterly incorrect*”

TAKE AWAY

RANDOM EXPERIMENT

A process leading to an uncertain outcome.

STATE SPACE

The state space is the collection of all possible outcomes of a random experiment, denoted by S.

BASIC OUTCOME

This is the outcome of a random experiment.

EVENT

An event is a subset of basic outcomes.

PROBABILITY

A probability is a measure for how likely an outcome of a random experiment is. For an event A, this is usually denoted by $P(A)$.

LAWS OF PROBABILITIES

Law 1. For every state space S, $P(S) = 1$.

Law 2. For any event A, $0 \leq P(A) \leq 1$.

Law 3. For disjoint events A and B, $P(A \cup B) = P(A) + P(B)$.

INDEPENDENCE

Two events A and B are (statistically) independent if and only if the probability of both A and B occuring is the product of the probabilities of the two events, $P(A \cap B) = P\left(A \; and \; B\right) = P(A) \cdot P(B)$.

BEST PRACTICES

1. Make sure that your sample space includes all of the possibilities.
2. Check that the probabilities assigned to all of the possible outcomes add up to 1.

PITFALLS

1. Do not multiply probabilities of dependent events.
2. Avoid assigning the same probability to every outcome (unless, of course, you are convinced that all outcomes are equally likely).
3. Do not confuse independent events with disjoint events.

2 - CONDITIONAL PROBABILITY

INTRODUCTION

CONTINGENCY TABLE

Contingency Table
A contingency table shows counts of cases on one **categorical** variable contingent on the value of another (for every combination of both variables). Contingency tables are useful for calculating conditional probabilities, as they contain all the ingredients necessary for the computation.

EXAMPLES

EXAMPLE 1: AMAZON.COM

We want to investigate which host sends more buyers to the internet shopping website Amazon.com. In order to answer this question, we must gather data on two (categorical) variables:

1. **The host that identifies the originating site, which is either MSN, RecipeSource, or Yahoo.**
2. **A binary variable that indicates whether the visit results in a purchase.**

The contingency table below shows data for web shoppers at Amazon.com. It contains the following pieces of information:

- **The website through which a customer ended up at Amazon.com**
- **Whether the customer made a purchase or not**

	MSN	Recipe Source	Yahoo	Total
No Purchase	6.973	4.282	5.848	17.103
Purchase	285	1	230	516
Total	7.258	4.283	6.078	17.619

The contingency table above shows counts (number of customers) and not probabilities. We can transform the counts into probabilities by dividing each number by the total (17.619 in this case). The table below shows the probabilities corresponding to the counts in the contingency table above.

	MSN	Recipe Source	Yahoo	Total
No Purchase	0.396	0.243	0.332	0.971
Purchase	0.016	0.000	0.013	0.029
Total	0.412	0.243	0.345	1.0

PROBABILITY TYPES

MARGINAL, JOINT AND CONDITIONAL PROBABILITIES

Marginal Probability
The marginal probability is the unconditional probability on an event. This probability is not conditioned on any other event.

Notation: **The marginal probabilities of two arbitrary events A and B is denoted by respectively $P(A)$ and $P(B)$.**

Joint Probability
The joint probability is the probability of simultaneous events. This is the probability of the intersection of two or more events.

Notation: **The joint probability of two arbitrary events A and B is denoted by $P(A \cap B)$.**

Conditional Probability
The conditional probability is the probability of an event *given* that some other event has occured. The information from the event that has occured can influence the probability of the original event.

Notation: **For two events A and B, the conditional probability of A given B is denoted by $P(A \mid B)$ and the conditional probability of B given A is denoted by $P(B \mid A)$.**

General Rule

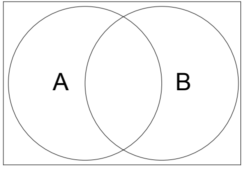
In general, for two arbitrary events A and B, $P(A \mid B) \neq P(B \mid A)$.

The first figure below shows the ‘old’ sample space S for two events A and B. Note that there is an overlap between A and B.

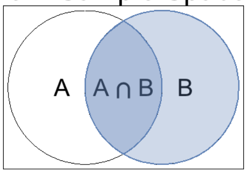
The second figure below shows the ‘new’ sample space S for two events A and B, because it shows the relation with respect to conditional probabilities.
The conditional probability can be calculated by dividing the probability of the dark blue area ($P(A \cap B)$) by the probability of the light blue area ($P(B)$). Note that the area of $A \cap B$ can be at most as large as the area of B (which is the case if B is fully contained in A). Therefore, if we calculate the conditional probability $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$, we immediately have that $P(A \mid B)$ is smaller than or

equal to 1.

Old' Sample Space.



lew' Sample Space



Relationship: P(A|B) = P(A ∩ B) / P(B)

PROOF OF $P(A \mid B) \neq P(B \mid A)$

Assume that:

- $P(A) \neq P(B), P(A) \neq 0$ **and** $P(B) \neq 0$**.**

We get:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \overset{[P(A \cap B) = P(B \cap A)]}{=} \frac{P(B \cap A)}{P(B)} \neq \frac{P(B \cap A)}{P(A)} = P(B \mid A),$$

where we used in the second last step that $P(A) \neq P(B)$.

EXAMPLES

EXAMPLE 1 CONT’D: AMAZON.COM

Consider the internet shopping example we saw earlier. Recall that the contingency table showed the following pieces of information:

- **The website through which a customer ended up at Amazon.com**
- **Whether the customer made a purchase or not**

The probability types that are given in a table corresponding to the contingency table are always

- **The marginal probabilities**
- **The joint probabilities**

The *conditional probabilities* can be calculated based upon these marginal and joint probabilities.

In the table below, the marginal probabilities are visible in the column ‘Total’ and the row ‘Total’. The joint probabilities are visible in the columns ‘MSN’, ‘Recipe Source’ and ‘Yahoo’ and the rows ‘NP’ and ‘P’.

	MSN	Recipe Source	Yahoo	Total
P	$P(MSN \cap P)$	$P(Recipe\ Source \cap P)$	$P(Yahoo \cap P)$	$P(P)$
NP	$P(MSN \cap NP)$	$P(Recipe\ Source \cap NP)$	$P(Yahoo \cap NP)$	$P(NP)$

Total	$P(MSN)$	$P(Recipe\ Source)$	$P(Yahoo)$	1

QUESTION

Consider again the internet shopping example that we saw earlier. Customers from which website are most likely to make a purchase on Amazon.com? *Hint:* treat that customers make a purchase as a given.

Solution.

The table below shows again the data (in terms of probabilities) for web shopping at Amazon.com. The table shows the website through which a customer ended up at Amazon.com and whether the customer made a purchase or not.

	MSN	Recipe Source	Yahoo	Total
No Purchase	0.396	0.243	0.332	0.971
Purchase	0.016	0.000	0.013	0.029
Total	0.412	0.243	0.345	1

In the question it is given that the customer made a purchase. Therefore we need to calculate the following probabilities:

- $P\left(MSN \mid Purchase\right)$
- $P\left(Recipe\ Source \mid Purchase\right)$
- $P\left(Yahoo \mid Purchase\right)$

We get:

- $P\left(MSN \mid Purchase\right) = \frac{P(MSN \cap Purchase)}{P(Purchase)} = \frac{0.016}{0.029} \approx 0.552.$
- $P\left(Recipe\ Source \mid Purchase\right) = \frac{P(Recipe\ Source \cap Purchase)}{P(Purchase)} = \frac{0.000}{0.029} \approx 0.$
- $P\left(Yahoo \mid Purchase\right) = \frac{P(Yahoo \cap Purchase)}{P(Purchase)} = \frac{0.013}{0.029} \approx 0.448.$

Therefore, based on the contingency table above, it is most likely that visitors from MSN make a purchase at Amazon.com

PROBABILITY TREES

PROBABILITY TREE

Probability Tree

A probability tree is a graphical depiction of conditional probabilities. It shows a sequence of events as a path, like branches of a tree.

EXAMPLES

EXAMPLE 2: SUCCESS OF TV ADVERTISING

Assume there are three programs that can be viewed on a Sunday evening. Viewers can either watch '60 Minutes', 'Desperate Housewives' or a football match. We want to investigate how successfull TV advertisement is given the TV program that can be watched. For each program, we collect data on the percentage of viewers that:

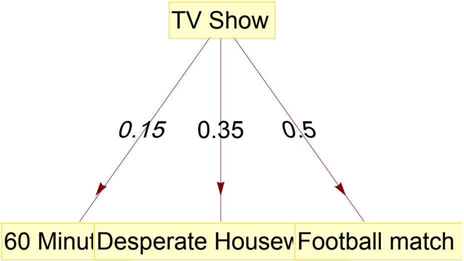
- **Watch the ads,**
- **Skip the ads.**

MARGINAL PROBABILITIES

It is given that the viewers watch the three respective shows with the following probabilities:

- $P\left(60\ Minutes\right) = 0.15,$
- $P\left(Desperate\ Housewives\right) = 0.35,$
- $P\left(Football\ match\right) = 0.5$**.**

The three probabilities above represent marginal probabilities. These probabilities are illustrated in the probability tree below.

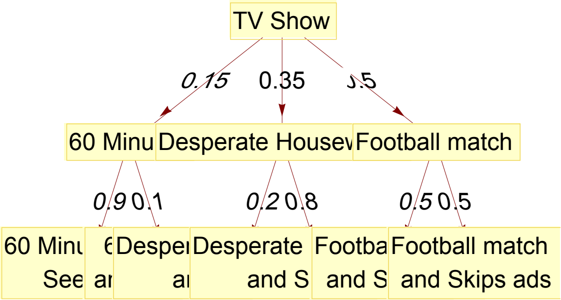


CONDITIONAL PROBABILITIES

There are six different conditional probabilities on whether a viewer watches ads or not, e.g.:

- $P\left(Sees\ ads \mid Football\ Match\right) = 0.5,$
- $P\left(Skips\ ads \mid Football\ Match\right) = 0.5,$
- ...
- $P\left(Skips\ ads \mid 60\ Minutes\right) = 0.1$**.**

These conditional probabilities are illustrated in the probability tree below.



WORKING WITH PROBABILITY TREES

Working with Probability Trees

Let $A_1,\ A_2,\ \dots,\ A_n$ **be** n **mutually exclusive and collectively exhaustive events. In addition, let** $B_1,\ B_2,\ \dots,\ B_k$ **be** k **mutually exclusive and collectively exhaustive events. Then** $P(A_1 \cap B_2)$ **is the joint probability on the events** A_1 **and** B_2**. Computing the marginal probability of** A_1 **can be done in the following way:**

$P(A_1) = P(A_1 \cap B_1) + P(A_1 \cap B_2) + \dots + P(A_1 \cap B_k)$**.**

EXAMPLES

QUESTION

Using the probability tree above, what is the joint probability for ‘Football match’ and ‘Sees Ads’ (e.g. $P(Sees\ Ads \cap Football\ match)$) and what is the marginal probability for ‘Sees Ads’ (e.g. $P\left(Sees\ Ads\right)$)?

Solution.

- $P(Sees\ Ads\ \cap\ Football\ match)$**.**

  We know from conditional probability that

  $P(Sees\ Ads \cap Football\ match)\ =$ **. It is given that**

  $$P(Sees\ Ads\,|\,Football\ match) \cdot P(Football\ match)$$

  $P(Football\ match) = 0.5$ **and** $P(Sees\ Ads\,|\,Football\ match) = 0.5$**.**

  Therefore $P(Sees\ Ads \cap Football\ match) = 0.5 \cdot 0.5\ = 0.25$**.**

- $P(Sees\ Ads)$**.**

  We know from ‘working with probability trees’ that we can calculate $P(Sees\ Ads)$ **the following way:**

$$P(Sees\ Ads) = P(60\ Min \cap Sees\ Ads) + P(D.H. \cap Sees\ Ads) + P(Football\ match \cap Sees\ Ads)$$

. We get:

$$P(Sees\ Ads) = 0.15 \cdot 0.9 + 0.35 \cdot 0.2 + 0.5 \cdot 0.5 = 0.455.$$

QUESTION

Consider the previous question. Fill in the contingency table below.

	60 Minutes	Desperate Housewives	Football match	Total
Sees Ads				
Skips Ads				
Total				1.0

Solution.

From the previous question, we know that $P(Sees\ Ads) = 0.455$ and $P(Sees\ Ads \cap Football\ match) = 0.25$. Therefore we can calculate $P(Skips\ Ads)$ and $P(Skips\ Ads \cap Football\ match)$. We get:

- $P(Skips\ Ads) = 1 - 0.455 = 0.545$,
- $P(Skips\ Ads \cap Football\ match) = P(Football\ match) - P(Football\ match \cap Sees\ Ads) = 0.5 - 0.25 = 0.25$.

From the probability tree, we know that $P(60\ Minutes) = 0.15$, $P(Desperate\ Housewives) = 0.35$ and $P(Football\ match) = 0.50$. Similar reasoning as in the previous question for the other joint probabilities yields the contingency table below.

	60 Minutes	Desperate Housewives	Football match	Total
Sees Ads	0.135	0.07	0.25	0.455
Skips Ads	0.015	0.28	0.25	0.545
Total	0.15	0.35	0.50	1.0

EXAMPLE 3: SPAM FILTER

Assume workers of a company want to filter out junk mail from important mail messages. They base their method on past data. For example, $20\ \%$ of all emails that were considered junk mail contained the word combination "Nigerian general". Past data indicates the following probabilities:

- $P(Nigerian\ general\ appears \mid Junk\ mail) = 0.20$,
- $P(Nigerian\ general\ appears \mid Not\ junk\ mail) = 0.001$,
- $P(Junk\ mail) = 0.50$.

QUESTION

Fill in the contingency table below.

	Junk mail	Not junk mail	Total

Nigerian general appears			
Nigerian general does not appear			
Total			1.0

Solution.

- **It is given that** $P(\mathit{Junk\ mail}) = 0.5$**. Therefore** $P(\mathit{Not\ junk\ mail}) = 1 - 0.5 = 0.5$.
- $P(\mathit{Nigerian\ general\ appears} \cap \mathit{Junk\ mail}) = P(\mathit{Nigerian\ general\ appears} \mid \mathit{Junk\ mail}) \cdot P(\mathit{Junk\ mail}) = 0.20 \cdot 0.50 = 0.10.$
- $P\,(\mathit{Nigerian\ general\ appears} \cap\ \mathit{Not\ junk\ mail}) = P\,(\mathit{Nigerian\ general\ appears} \mid\ \mathit{Not\ junk\ mail}) \cdot P\,(\mathit{Not\ junk\ mail}) = 0.001 \cdot 0.50 = 0.0005.$
- **Similar reasoning for** $P(\mathit{Nigerian\ general\ does\ not\ appear} \cap \mathit{Junk\ Mail})$ **and** $P(\mathit{Nigerian\ general\ does\ not\ appear}\ \cap\ \mathit{Not\ junk\ mail})$ **leads to the contingency table below.**

	Junk mail	Not junk mail	Total
Nigerian general appears	0.1	0.0005	0.1005
Nigerian general does not appear	0.4	0.4995	0.8995
Total	0.5	0.5	1.0

QUESTION

Using the contingency table above, calculate the probability that an email should be considered junk mail given that the phrase “Nigerian general” appears.

Solution.

$$P(\mathit{Junk\ mail} \mid \mathit{Nigerian\ general\ appears}) = \quad .$$

$$\frac{P(\mathit{Junk\ mail} \cap \mathit{Nigerian\ general\ appears})}{P(\mathit{Nigerian\ general\ appears})} = \frac{0.1}{0.1005} = 0.995$$

We can conclude that email messages to this employee with the phrase “Nigerian general” have a high probability (more than $99\ \%$) of being spam. The spam filter should move emails containing this phrase straight to the junk folder.

APPLICATIONS

BIRTHDAY PROBLEM

QUESTION

Assume there are 30 people in one room. How much would you bet that there is nobody in the room who shares the same birthday, assuming that a year has 365 days?

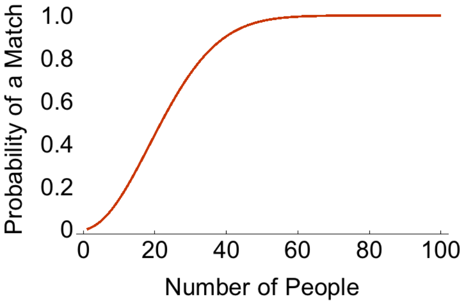
Solution.

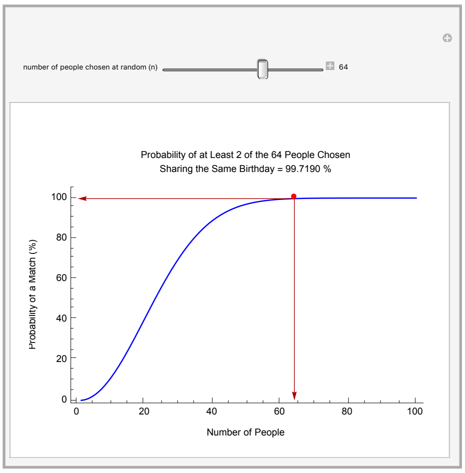
Let n be the number of people in the room. We denote A the event that *at least* two people share the same birthday. The probability on A is then denoted by $P(A)$. Calculating $P(A)$ results in the following table:

A	5	10	20	30	40	50	60
$P(A)$	0.0271	0.1169	0.4114	0.7063	0.8912	0.9794	0.9941

Note that for only 30 people, there is almost a $71\ \%$ chance that at least two people share the same birthday. This is surprisingly likely for a group that seems so small compared to the number of days in a year!

The graph below shows on the x-axis the number of people in one room and on the y-axis the probability that at least two people share the same birthday.





number of people chosen at random (n) 64

Probability of at Least 2 of the 64 People Chosen
Sharing the Same Birthday = 99.7190 %

CALCULATION

Let us make the following two assumptions:

- **Every day of the year is equally likely to be a birthday**
- **There are 365 days in a year**

The probability that at least two people share the same birthday (denoted by

$P(A)$) is equal to 1 minus the probability that nobody shares the same birthday (denoted by $P(A^c)$).

For n people in the room, there are in total 365^n possible outcomes and the event A can happen in $365 \cdot 364 \cdot \ldots \cdot (365 - n + 1)$ ways. We get:

$$P(A) = 1 - P(A^c) = 1 - \frac{365 \cdot 364 \cdot 363 \cdot \ldots \cdot (365 - n + 1)}{365^n}.$$

Putting $n = 30$ people into the formula above yields indeed an probability of 0.7063 that at least two people share the same birthday.

NAIVE BAYES CLASSIFICATION

> “***Naive Bayes is one of the most efficient and effective inductive learning algorithms for machine learning and data mining. Its competitive performance in classification is surprising, because the conditional independence assumption on which it is based, is rarely true in real-world applications.***” **(Zhang, 2004)**

- **Naive Bayes is useful for classification of emails whether they belong to the spam folder or are legitimate.**
- **Naive Bayes is useful for machine learning, (algorithms that can learn from data).**
- **Naive Bayes is useful for clustering techniques, where all kinds of objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).**

BAYES’ RULE

Naive Bayes Classification

Naive Bayes classification depends on Bayes’ Rule. For two events A and B, Bayes’ rule can be expressed by the following relationship:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \overset{P(B \cap A) = P(A \cap B)}{=} \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)},$$ **provided that** $P(A) \neq 0$**.**

Here, $P\,(B)$ **is called the** a priori **probability of** B **and** $P(B \mid A)$ **is called the** a posteriori **probability of** B**. Note that we expressed the probability of** B **given** A **in terms of** A **given** B**.**

Conditional Independence Assumption

Naive Bayes Classification relies heavily on two assumptions:

1. All events are equally important (they all have equal weights).
2. All events are mutually independent.

EXAMPLE 4: A NEW DAY

Let us assume the weather is as follows:

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
			Cold	3	1								
Rainy	3	2											

Converting the counts into probabilities yields the table below. Note that the table above is not a contingency table!

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
Sunny	$\frac{2}{9}$	$\frac{3}{5}$	Hot	$\frac{2}{9}$	$\frac{2}{5}$	High	$\frac{3}{9}$	$\frac{4}{5}$	False	$\frac{6}{9}$	$\frac{2}{5}$	$\frac{9}{14}$	$\frac{5}{14}$
Overcast	$\frac{4}{9}$	0	Mild	$\frac{4}{9}$	$\frac{2}{5}$	Normal	$\frac{6}{9}$	$\frac{1}{5}$	True	$\frac{3}{9}$	$\frac{3}{5}$		
			Cold	$\frac{3}{9}$	$\frac{1}{5}$								
Rainy	$\frac{3}{9}$	$\frac{2}{5}$											

QUESTION

What is the probability that the outlook is sunny, the temperature is considered cold, the humidity is high and that it is windy and play is either ‘yes’ or ‘no’?

Solution.

We get for a sunny outlook, a cool temperature, high humidity, windy weather and play equal to yes:

- $\frac{2}{9}\cdot\frac{3}{9}\cdot\frac{3}{9}\cdot\frac{3}{9}\cdot\frac{9}{14}\approx 0.0053$.

We get for no sunny outlook, no cool temperature, no high humidity, no windy weather and play equal to no:

- $\frac{3}{5}\cdot\frac{1}{5}\cdot\frac{4}{5}\cdot\frac{3}{5}\cdot\frac{5}{14}\approx 0.0206$.

So, transforming this to probabilities by normalizing the above values, we get

- $P('\,yes\,') = \frac{0.0053}{0.0053+0.0206} = 0.205 = 20.5\ \%$.
- $P('\,no\,') = \frac{0.0206}{0.0053+0.0206} = 0.795 = 79.5\ \%$.

We can immediately see that these probabilities add up to 1.

TAKE AWAY

ORDER OF UNIONS AND INTERSECTIONS

For events A and B,

1. $P(A\cap B) = P(B\cap A)$.
2. $P(A\cup B) = P(B\cup A)$.

CONTINGENCY TABLE

A contingency table shows counts of cases on one categorical variable contingent on the value of another (for every combination of both variables).

GENERAL RULE

In general, for two events A and B, $P(A \mid B) \neq P(B \mid A)$.

MARGINAL PROBABILITY

The unconditional probability on an event.

JOINT PROBABILITY

The probability on simultaneous events.

CONDITIONAL PROBABILITY

The probability of an event given some other event. For two events A and B, it holds that:

$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$, where $P(B) \neq 0$.

$P(B \mid A) = \frac{P(B \cap A)}{P(A)}$, where $P(A) \neq 0$.

INDEPENDENCE

Two events A and B are independent if $P(A \cap B) = P(B \cap A) = P(A) \cdot P(B)$

For independent events A and B, combining the rules for conditional probability and independence yields:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A).$$

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B).$$

PROBABILITY TREE

A probability tree is a graphical depiction of conditional probabilities.

BAYES' RULE

$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)}$, provided that $P(A) \neq 0$.

BEST PRACTICES

1. Presume events are dependent and use the General Multiplication Rule.
2. Use tables to organize probabilities.
3. Check that you have included all events.

PITFALLS

1. Do not confuse $P(A \mid B)$ for $P(B \mid A)$.
2. Do not confuse counts with probabilities in contingency tables.

3. Understand that conditional probabilities are not shown in contingency tables, but can be calculated directly from them.

4 - RANDOM VARIABLES

INTRODUCTION

RANDOM VARIABLES

Random Variable
A random variable is a variable whose value is subject to variations due to chance.
***Notation:* A r**andom variable is usually denoted with a capital letter, for example X or Y. The individual values ar**e denoted by respectively** x **and** y**.**

Probability

The likelihood that a random variable X **is** equal to an individual value x***.***

Notation: $P(X = x) = P(x) = p(x)$**.**

DISCRETE DISTRIBUTIONS

DISCRETE RANDOM VARIABLES

There are two types of random variables, discrete random variables and continuous random variables. In this section, we discuss discrete random variables.

Discrete Random Variable

A random variable X **is said to be discrete, if** X **can take on only a** finite **number of values (or at most a countably infinite number).**

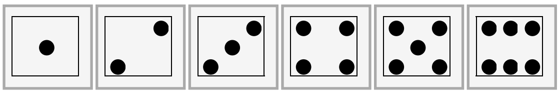
Probability Mass Function
A probability mass function (abbreviated PMF) completely describes the probability properties of the random variable. It shows the probability that a random variable X is exactly equal to some deterministic value x.

Cumulative Distribution Function
The cumulative distribution function (abbreviated CDF) shows the probability that a random variable X takes a value less than or equal to a deterministic value x.
Notation: $F(x) = P(X \leq x)$, where $-\infty < x < \infty$.

Discrete Uniform Distribution
The discrete uniform distribution is a symmetric probability distribution whereby a finite number of values are equally likely to be observed; every one of n values has equal probability $\frac{1}{n}$.

EXAMPLES

EXAMPLE 1: FAIR DIE - THE DISCRETE UNIFORM DISTRIBUTION



The number of possible outcomes for throwing a fair die is obviously finite. Let X be the random variable representing the possible outcomes of throwing a fair die.

- **Reason that X is a discrete random variable.**
- **Reason that X is uniformly distributed (the discrete case).**

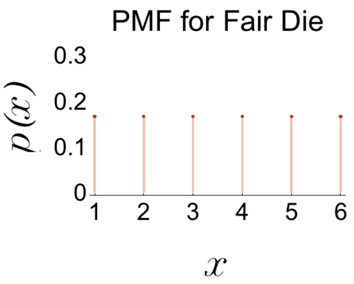
Solution.

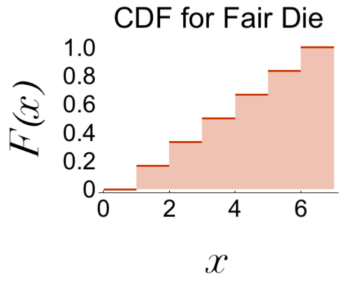
- **The number of possible outcomes for X is clearly finite. Therefore X is a discrete random variable.**
- **All outcomes are equally likely. Therefore X is uniformly distributed.**

  Since X is a discrete random variable, X follows the discrete uniform distribution.

PMF AND CDF

In the figures below, the probability mass function and the cumulative distribution function of X are shown.





EXAMPLE 2: SUM OF TWO DICE

In this example, two fair dice are thrown simultaneously. The sum is taken of the outcomes of each individual die. Each sum has a different probability. Consider the case where the sum is equal to 7 and where the sum is equal to 2. From the lecture on independence, recall that there are more combinations possible where the sum is equal to 7 than where the sum is equal to 2 and therefore the former is more likely to occur than the latter.

QUESTION

Let us denote by X the sum of the outcomes of the two dice. The state space S in this example is equal to $\{2,\ 3,\ \ldots,\ 11,\ 12\}$.

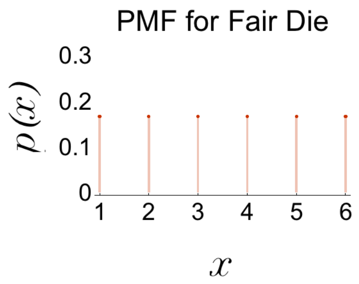
- **Reason that X is a discrete random variable.**
- **Reason that X is not uniformly distributed.**

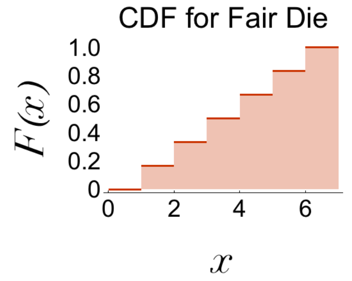
Solution.

- **The number of possible outcomes for X is clearly finite. Therefore X is a discrete random variable.**
- **All possible outcomes for the sum of two fair dice are not equally likely. Therefore X is not uniformly distributed.**

PMF AND CDF

In the figures below, the probability mass function and the cumulative distribution function of X are shown.





EXAMPLE 3: NUMBER OF DEFECTIVE PIECES

Assume there are 10 lamps in one box and they could either be new and working (with probability 0.7) or they could be old and defect (with probability 0.3).

QUESTION

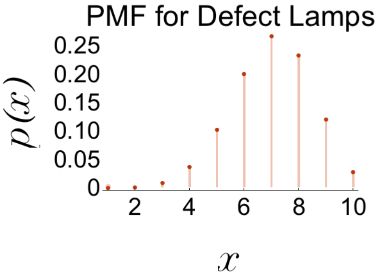
- **Reason that X is a discrete random variable.**
- **Reason that X is not uniformly distributed.**

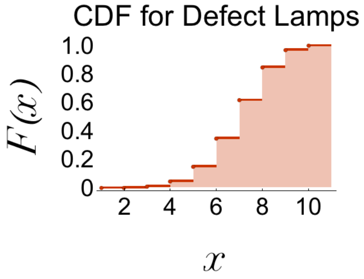
Solution.

- **The number of possible outcomes for X is clearly finite. Therefore X is a discrete random variable.**
- **All possible outcomes are not equally likely. This can be seen for example in the probability mass function of X.**

PMF AND CDF

In the figures below, the probability mass function and the cumulative distribution function of X are shown. In this particular case, X is said to be binomially distributed. The binomial distribution will be treated in further detail in lecture 5 of this course.





EXAMPLE 4: WAITING TIMES AT CALL CENTERS

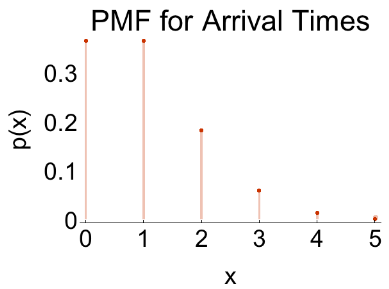
The priority of every call center should be handling calls from customers efficiently and minimizing waiting times. This is not a straightforward task, as call arrivals and service times of calls are random. The difficult part is to staff each call center with a sufficient amount of people. Too many people is costly and inefficient, but usefull at times when all of a sudden many calls arrive. Too few people is cost efficient, but leads to long waiting when many calls arrive at the same time.

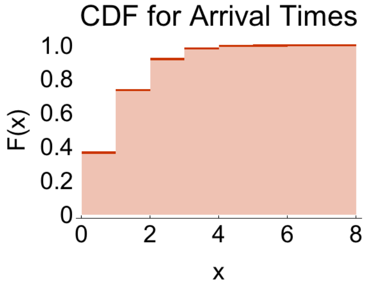
CALL ARRIVALS

A call center can be naturally viewed as a queueing system. Here we look at the likelihood of a certain number of calls arriving per minute. We denote by X the number of calls that is arriving per minute. Note that X is a discrete random variable. We omit technical details here, but this type of process can be modelled with the Poisson distribution.

PMF AND CDF

In the figures below, the probability mass function and the cumulative distribution function of X are shown.





CONTINUOUS DISTRIBUTIONS

CONTINUOUS RANDOM VARIABLES

Continuous Random Variable

A random variable X is said to be continuous, if the number of possible outcomes for X is (uncountably) infinite.

Probability Density Function

For a continuous random variable, the probability density function, usually denoted by $f\,(x)$, has the following properties:

1. $f(x) \geq 0$

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$**, where $\int$ refers to the integral.**

If X is a continuous random variable with a density function f, then for any $a < b$, the probability that X falls in the interval $(a,\ b)$ is the area under the density function between a and b:

$$P(a < X < b) = \int_a^b f(x)\,dx.$$

Cumulative Distribution Function

The cumulative distribution function (abbreviated CDF) shows the probability that a random variable X takes a value less than or equal to a deterministic

value x. It shows the area under the probability density function from $-\infty$ to x.
Notation: $F(x) = P(X \leq x)$, where $-\infty < x < \infty$.

EXAMPLES

Returns
Returns show the relative percentage of increase or decrease of the price of a stock at time t in comparison to a price at time $t-1$. The return is calculated based on the price at time t, P_t, and the price at time $t-1$, P_{t-1}. A time step can be a second, a day, a month, etc.
Notation: $R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$**.**
Remark: **For finance experts, we consider returns without dividens.**

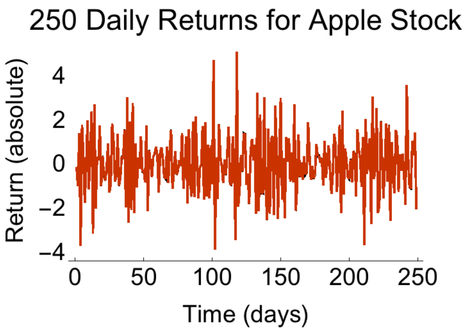
Histogram
A histogram **is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous random variable. It shows the counts of the number of times the stock price had a certain increase or decrease.**

RETURNS

Assume that the price of an Apple stock on 01-01-2015 is $100. We want to make a prediction for the Apple stock price in one year from now by predicting the daily returns. We make the following two assumptions:

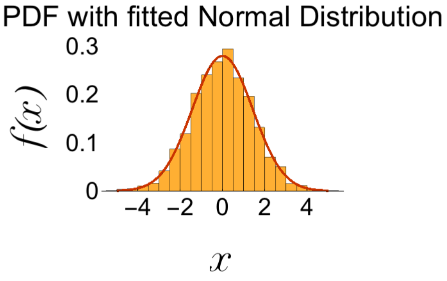
- **The returns are normally distributed**
- **There are 250 trading days in one year**

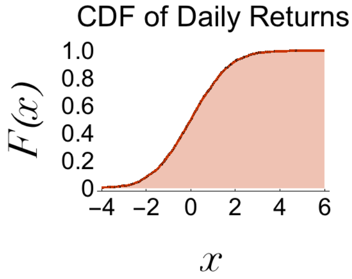
The 250-day forecast for the returns is visible in the figure below. We can clearly see that the returns ‘hover’ around zero. This makes intuitively sense, because (a series of) negative returns are usually followed by (a series of) positive returns.



HISTOGRAM AND CDF

We can visualize the count of daily stock return increases and decreases in a histogram. The histogram is shown in the top figure below together with a fitted normal distribution. The CDF for the daily return forecasts is shown in the bottom figure below.

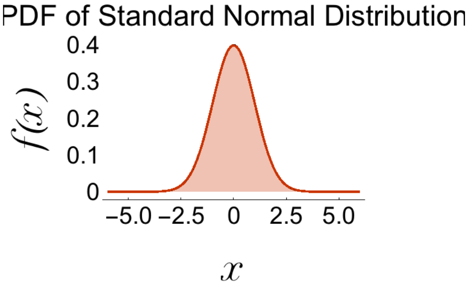




SUMMARY MEASURES

INTRODUCTION

Below the probability density function from the standard normal distribution is shown (i.e. the normal distribution with mean 0 and variance 1. This will be explained in chapter 6 of this lecture).



Showing the PDF can be informative (as in the standard normal case), but usually probability distributions are difficult to grasp. In order to communicate the largest amount of information as simple as possible, we use summary measures.

MEAN AND EXPECTED VALUE

Measure of Location

The arithmetic mean is normally used as a measure of location/central tendency.

Notation: the arithmetic mean is usually denoted by μ.

Expected Value

The expected value $E[X]$ of a discrete random variable X is the probability-weighted sum of all possible values.

Notation: **$E[X] = x_1\, p(x_1) + x_2\, p(x_2) + \ldots + x_k\, p(x_k)$, for k possible values of X.**

EXAMPLES

EXAMPLE 7: EXPECTED VALUE OF A FAIR DIE

Recall that the state space S of the fair die is equal to $\{1, 2, 3, 4, 5, 6\}$ and that every possible outcome has an equal probability of $\frac{1}{6}$. We denote by X the random variable that represents the outcome of rolling the fair die. The table below shows $P(X = x)$ for each possible outcome.

x	1	2	3	4	5	6
P(X=x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

QUESTION

What is the expected value of the fair die?

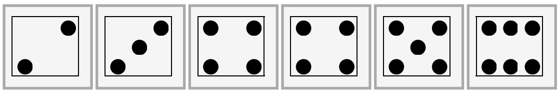
Solution.

Every outcome is equally likely and therefore we assign equal weight to each outcome. We simply multiply each outcome by its individual probability. We get:

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5.$$

Interpretation: The expected value of an extremely large number of dice rolls will very likely almost be equal to 3.5. Note that the average of rolling the die is not an element of the state space.

EXAMPLE 8: EXPECTED VALUE OF A LOADED DIE



A loaded die has instead of the ‘1’ an extra ‘4’. Therefore a ‘4’ is twice as likely as each other individual outcome. We denote by X the random variable that represents the outcome of rolling the loaded die. The table below shows the $P(X = x)$ for each possible outcome.

x	2	3	4	5	6
P(X=x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

QUESTION

What is the expected value of the loaded die?

Solution.

Multiplying each outcome by its individual probability yields the expected value of X. We get:

$$E[X] = 2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{2}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=4.$$

Interpretation: The expected value of an extremely large number of dice rolls with the loaded die will very likely (almost) be equal to 4.

RANDOM VARIABLES AS MODELS

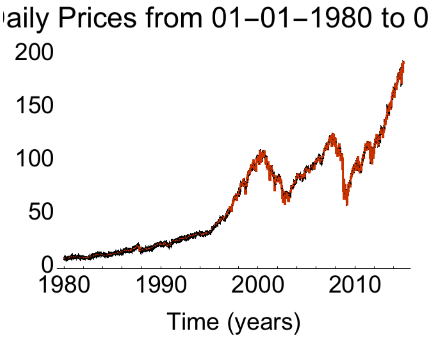
INTRODUCTION

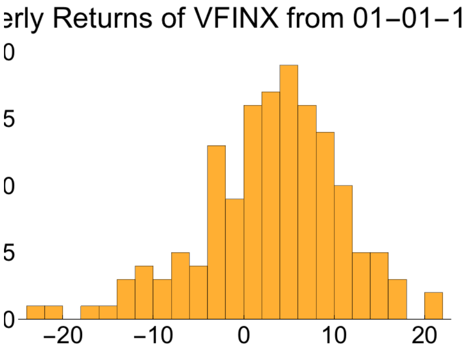
A random variable is a type of statistical model. In business applications, a statistical model usually represents a simplified or idealized view of reality.

EXAMPLES

EXAMPLE 9: VANGUARD 500 INDEX FUND

You are considering an investment in the oldest S&P 500 index fund. The “Vanguard 500 Index Fund” (ticker symbol: VFINX) was the first index fund for individual investors. This mutual fund invests in 500 of the largest U.S. companies, which span many different industries and account for about $75\ \%$ of the U.S. stock market’s value. In the top figure below, the daily price development is visible from 01-01-1980 to 01-01-2015. The time period 01-01-1977 to 12-31-1980 is not visible in the figure below. We are interested in the quarterly returns, which we can visualize in a histogram. The histogram of the quarterly returns from 01-01-1977 to 01-01-2015 is displayed in the bottom figure below.





QUESTION

Denote the first quarterly return in 1977 by R_1, the second quarterly return by R_2. Using similar notation for all subsequent returns up to the fourth quarterly return in 2014, R_{152}. Express the historical average of the returns in terms of $R_1, \ \ldots, \ R_{152}$.

Solution.

Using the notation introduced in the question, we can express the average quarterly return in terms of $R_1, \ \ldots, \ R_{152}$ in the following way: $\frac{R_1+R_2+\ldots+R_{152}}{152}$. Calculating the average quarterly return for our data sample yields $\frac{R_1+R_2+\ldots+R_{152}}{152} = 2.97\ \% \approx 3\ \%$. Based on a statistical model that assumes that all past outcomes are equally likely to occur, the forecast for the first quarter in 2015 would be $3\ \%$.

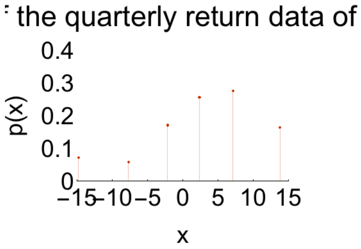
EXAMPLE 9 CONT’D: VANGUARD 500 INDEX FUND

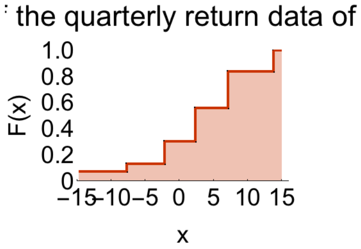
Based on the histogram of the quarterly return data of VFINX, we can construct the probability distribution by deriving a table for the returns with their corresponding probabilities. This table is shown below.

Return	−14.7491	−7.6566	−2.14	2.3877	7.1442	13.8112
p	0.0723	0.0592	0.171	0.2565	0.2763	0.1644

PMF AND CDF

In the figure below, the probability mass function and the cumulative distribution function of X are visible.





EXAMPLE 9 CONT'D: VANGUARD 500 INDEX FUND

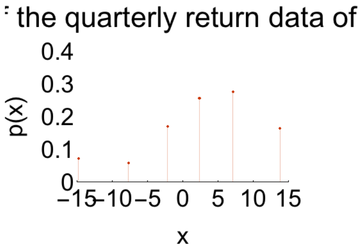
Now, the economic forecast for the next quarter is very good. Therefore, we build a model by subjectively changing the probabilities (and rounding the conditional expectations). There are six scenarios with six probabilities. These probabilities with their corresponding returns are displayed in the table below.

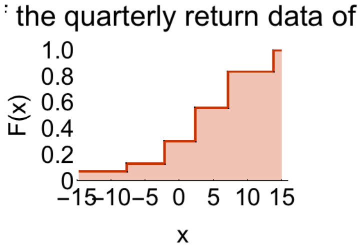
Return	−15.0	−7.5	−2.5	2.5	8.0	14.0
p	0.02	0.05	0.13	0.30	0.30	0.20

Our new expected value of the next quarterly return of VFINX is equal to 4.95% based on the model where we subjectively changed the probabilities.

PMF AND CDF

In the figure below, the probability mass function and the cumulative distribution function of X are visible.





SHORTCOMINGS OF EXPECTED VALUE AS A MEASURE

We have a new expected value of 4.95% in our new model. This single number does not indicate anything about the risk of investing in VFINX.

VARIATION

INTRODUCTION

We need a measure which captures the ‘deviation’ from the mean, i.e. the variation of VFINX. If all historical data points are very close to the mean, then the realized return of VFINX will likely not be far away from the expected value and therefore there is not much risk. If, on the other hand, many historical data points are ‘far away’ from the mean, one could argue that it is more likely that VFINX either greatly outperforms the expectation or that a huge negative return is realized. This could indicate that investing in VFINX is very risky. How could we measure this variation?

SIMPLE DEVIATION

Simple Deviation

The simple deviation **for a set of data values** x_i**,** $i = 1, \ldots, n$**, is equal to the** sum of the differences **of the data values and the mean** ***μ*****.**

Notation: $\sum_{i=1}^{n} (x_i - \mu)$**.**

EXAMPLES

EXAMPLE 10: SIMPLE DEVIATION OF A FAIR DIE

Consider once more the fair die with state space $S = \{1, 2, 3, 4, 5, 6\}$.

QUESTION

Calculate the simple deviation for the fair die.

Solution.

The simple deviation in this example with $n = 6$ and $\mu = 3.5$ is equal to $\sum_{i=1}^{6} (x_i - 3.5)$. The results for x_i with $i = 1, \ldots, 6$ are visible in the table below.

x_i	1	2	3	4	5	6
$x\_i{-}\mu$	-2.5	-1.5	-0.5	0.5	1.5	2.5

So, taking the sum $\sum_{i=1}^{6} (x_i - 3.5) = -2.5 - 1.5 - 0.5 + 0.5 + 1.5 + 2.5 = 0$. Therefore the simple deviation is equal to zero in this example.

SIMPLE DEVIATION AS A MEASURE OF VARIATION

In example 10, the simple deviation was equal to zero. In fact, it will turn out that the simple deviation will always be equal to zero for any random variable X, as will be proved later.

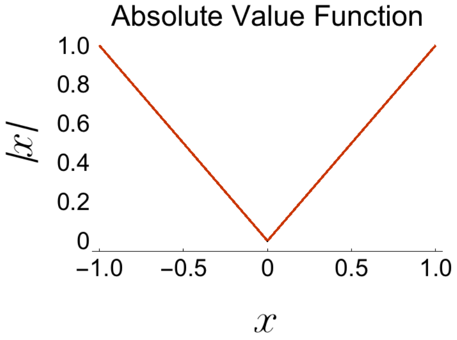
PROOF THAT SIMPLE DEVIATION IS EQUAL TO ZERO

Let us assume that there are n observations for a random variable X. Using the definition of simple deviation, we get:

$$\sum_{i=1}^{n} (x_i - \mu) =$$

$$\sum_{i=1}^{n} x_i - n \cdot \mu = \sum_{i=1}^{n} x_i - n \cdot \frac{1}{n} \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0.$$

ABSOLUTE VALUE FUNCTION



The kink at the origin in the absolute value function is undesirable, because therefore the function $|x|$ is not differentiable at $x = 0$.

VARIANCE

The next easiest measure for variation is taking the squares of the differences of the data values and the mean and sum those values up. This leads to a measure of variation called variance.

Variance: General Case

The variance **of a random variable** X **is** the expected value of squared deviations **from** μ**.**

Notation: $\sigma^2 = Var[X] = E\left[(X - (E[X]))^2\right] = E\left[X^2\right] - (E[X])^2$**,**

where $E[X]$ **is the expected value of** X **and** $E\left[X^2\right]$ **is the second moment* of** X**.**

Variance: Discrete Case

Only the notation changes for the variance.

Notation: $Var[X] = \sum_{i=1}^{n} (x_i - E[X])^2 \, p(x_i)$,

where $p(x_i)$is the probability of X being equal to x_i. When all n values are equally likely, the expression simplifies to

$Var[X] = \frac{1}{n}\sum_{i=1}^{n}(x_i - E[X])^2$.

***Remark:* the units of the variance are the** squared units of the random variable. Therefore it is convenient to take the square root of the variance. This leads us to the standard deviation**.**

Standard Deviation

The standard deviation of a random var**iable** X **is the** square-root of the variance **of** X**.**

Notation: $SD[X] = \sigma = \sqrt{Var[X]}$ **.**

REMARK ON SQUARE ROOTS OF SQUARED UNITS

In general, the square root cancels out the square of a squared number. For example, $\sqrt{9} = \sqrt{3^2} = 3$. However, the square root of the variance does not cancel out the square.
Illustration:

$$\sqrt{3^2 + 4^2} = \sqrt{5^2} = 5 \overset{!}{\neq} 7 = 3 + 4.$$

$$\sqrt{VAR[X]} \neq SDEV[X]$$

To show that the square-root of the variance is not equal to the simple deviation of X:

$$Var[X] =$$

$$\sqrt{\sum_{i=1}^{n}(x_i - \mu)^2 \cdot p(x_i)} = \sqrt{\sum_{i=1}^{n}\left[x_i^2 + \mu^2 - 2 \cdot x_i\, \mu\right]] \cdot p(x_i)}$$

$$\overset{!}{\neq} \sum_{i=1}^{n}(x_i - \mu) \cdot p(x_i) = SDev[X]$$

EXAMPLES

EXAMPLE 7 CONT'D: VARIANCE OF A FAIR DIE

Consider once more the fair die with state space $S = \{1, 2, 3, 4, 5, 6\}$.

QUESTION

Calculate the variance and standard deviation for the fair die.

Solution.

Recall from the definition that for a discrete random variable X with $n = 6$,

$$Var[X] = \sum_{i=1}^{6}(x_i - \mu)^2\, p(x_i) =$$
$$(x_1 - \mu)^2\, p(x_1) + (x_2 - \mu)^2\, p(x_2) + (x_3 - \mu)^2\, p(x_3) +$$

$(x_4 - \mu)^2\, p(x_4) + (x_5 - \mu)^2\, p(x_5) + (x_6 - \mu)^2\, p(x_6)$. With mean $\mu = 3.5$, we get:

x_i	p(x_i)	(x_i−μ)	(x_i−μ)^2	p(x_i)·(x_i−μ)^2
1	$\frac{1}{6}$	−2.5	6.25	1.04167
2	$\frac{1}{6}$	−1.5	2.25	0.375
3	$\frac{1}{6}$	−0.5	0.25	0.04167
4	$\frac{1}{6}$	0.5	0.25	0.04167
5	$\frac{1}{6}$	1.5	2.25	0.375
6	$\frac{1}{6}$	2.5	6.25	1.04167

$$Var[X] = \qquad .$$
$$1.04167 + 0.375 + 0.04167 + 0.04167 + 0.375 + 1.04167 \approx 2.917$$

$SD[X] = \sqrt{Var[X]} = \sqrt{2.917} \approx 1.71$. Hence, the variance is equal to approximately 2.917 and the standard deviation is equal to approximately 1.71.

EXAMPLE 8 CONT'D: VARIANCE OF A LOADED DIE

Consider once more the example of the loaded die, which has instead of the '1' an extra '4'. Recall that it is therefore twice as likely that a '4' comes up. We denote by X the random variable that represents the outcome of rolling the loaded die.

QUESTION

Calculate the variance and standard deviation for the loaded die. The probability on each outcome of the loaded die is given in the table below.

x	2	3	4	5	6
P(X=x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

Solution.

Recall from the definition that for a discrete random variable X with $n = 5$ we have,

$$Var[X] = \sum_{i=1}^{5}(x_i - \mu)^2\, p(x_i) = (x_1 - \mu)^2\, p(x_1) + (x_2 - \mu)^2\, p(x_2)$$
$$+(x_3 - \mu)^2\, p(x_3) + (x_4 - \mu)^2\, p(x_4) + (x_5 - \mu)^2\, p(x_5).$$

The mean μ is equal to 4. We get:

x_i	$p(x\_i)$	$(x\_i-E[X])$	$(x\_i-E[X])$^2	$p\ (x\_i)\cdot(x\_i-E[X])$^2
2	1/6	−2	4	2/3

3	1/6	−1	1	1/6
4	2/6	0	0	0
5	1/6	1	1	1/6
6	1/6	2	4	2/3

$Var[X] = \sum_{i=2}^{6} (x_i - \mu)^2 \, p(x_i) = \frac{5}{3}$.

$SD[X] = \sqrt{Var[X]} = \sqrt{\frac{5}{3}} \approx 1.291$. Hence, the variance is equal to approximately 1.667 and the standard deviation is equal to approximately 1.291.

EXAMPLE 9 CONT'D: VFINX

Recall the model we made that we used to predict the next quarterly returns of VFINX. The model consists of six different returns with six subjectively assigned probabilities, which are displayed in the table below. Calculate the variance of this statistical model.

p	0.02	0.05	0.13	0.3	0.3	0.2
Return	−15	−7.5	−2.5	2.5	8	14

Solution.

For a discrete random variable X, the variance is given by:

$Var[X] = \left(x_1 - \mu\right)^2 p(x_1) + (x_2 - \mu)^2 p(x_2) + \ldots + (x_n - \mu)^2 p(x_n)$. In our case, $n = 6$, so we get:

$$Var[X] = (x_1 - \mu)^2 p(x_1) + (x_2 - \mu)^2 p(x_2) + (x_3 - \mu)^2 p(x_3) + \ldots +. \\ (x_6 - \mu)^2 p(x_6)$$

With $\mu = 4.95\ \%$, this yields to the table below.

x_i	$p(x\_i)$	$(x\_i - \mu)$	$(x\_i - \mu)\^2$	$p(x\_i) \cdot (x\_i - \mu)\^2$
−15	0.02	−19.95	398.0025	7.9600
−7.5	0.05	−12.45	155.0025	7.7501
−2.5	0.13	−7.45	55.0025	7.2153
2.5	0.3	−2.45	6.0025	1.8008
8	0.3	3.05	9.3025	2.7907
14	0.2	9.05	81.9025	16.3805

$$Var[X] = \\ 7.9600 + 7.7501 + 7.2153 + 1.8008 + 2.7907 + 16.3805 = 43.8975.$$

$SD[X] = \sqrt{Var[X]} = \sqrt{43.8975} = 6.6255$. Hence, the variance is equal to approximately 43.90 and the standard deviation is equal to approximately 6.63.

CALCULATION RULES

INTRODUCTION

We can add constants to random variables or multiple random variables by constants. This has an impact on the expected value, the variance and the standard deviation. The rules for these operations on random variables are given below.

OPERATIONS ON RANDOM VARIABLES

Adding a constant

1. $E[X \pm a] = E[X] \pm a$
2. $Var[X \pm a] = Var[X]$
3. $SD[X \pm a] = SD[X]$

Multiplying by a constant

1. $E[c \cdot X] = c \cdot E[X]$
2. $Var[c \cdot X] = c^2 \cdot Var[X]$
3. $SD[c \cdot X] = |c| \cdot SD[X]$**, where** $|c|$ **is the** absolute value **of** c

Linear Functions of *X* (follows from adding and multiplying a constant)

1. $E[a + c \cdot X] = a + c \cdot E[X]$
2. $Var[a + c \cdot X] = c^2 \cdot Var[X]$
3. $SD[a + c \cdot X] = |c| \cdot SD[X]$

EXAMPLES

QUESTION

Assume you throw a fair die once. Calculate:

- **The expected value of** X **times 5,**
- **The variance of** X **times 5.**

Solution.

Denote by X the random variable representing the outcome of throwing a fair die.

- **The expected value of throwing a fair die once is equal to 3.5, i.e.** $E[X] = 3.5$**. We now need to calculate** $E[5 \cdot X]$**. We can use the rule for multiplying by a constant, i.e.** $E[c \cdot X] = c \cdot E[X]$**. By using this rule, we**

get $E[5 \cdot X] = 5 \cdot E[X] = 5 \cdot 3.5 = 17.5$.

- **The variance of throwing a fair die once is equal to 2.92. We can use the rule for multiplying by a constant, i.e $Var[c \cdot X] = c^2 \cdot Var[X]$. By using this rule, we get $Var[5 \cdot X] = 25 \cdot 2.92 = 72.92$.**

APPLICATIONS

SHARPE RATIO

One of the most popular measures to compare investments with different means and standard deviations is the Sharpe Ratio.

Sharpe Ratio

The Sharpe ratio is **the average return μ earned in excess of the risk-free rate r_f per unit of volatility (equal to SD) σ or total risk. By subtracting the risk-free rate r_f from the mean return μ, the performance associated with risk-taking activities can be isolated. The higher the Sharpe Ratio, the better the performance of the investment**

Notation: $S(X) = \frac{\mu_X - r_f}{\sigma_X}$**, where X is the investment opportunity.**

Risk-Free Interest Rate

The risk-free interest rate r_f is the theoretical rate of return of an investment with no risk of financial loss.

Assume you can invest in either Disney or Mc Donald's stocks. The mean and standard deviation for both stock returns are given in the table below.

Company	Random Variable	Mean	SD
Disney	D	0.61%	8.3%
Mc Donald's	M	0.53%	7.6%

QUESTION

Suppose the risk-free rate is 0.4%. In which company should you invest when you compare their Sharpe Ratios?

Solution.

Let us denote by μ_D and σ_D the mean and standard deviation of Disney's stock returnand and by μ_M and σ_M the mean and standard deviation of Mc Donald's stock return. Calculating the Sharpe Ratios for respectively Disney and Mc Donald's yields:

- $S(D) = \frac{\mu_D - r_f}{\sigma_D} = \frac{0.61-0.40}{8.3} \approx 0.0253$.
- $S(M) = \frac{\mu_M - r_f}{\sigma_M} = \frac{0.53-0.40}{7.6} \approx 0.0171$.

According to the Sharpe Ratio, Disney should be preferred to McDonald’s.

FINANCIAL DISTRESS AND AGENCY COST

Assume company XYZ has a loan of $10 million that is due at the end of the year. Company XYZ is in financial distress, because the market value of its assets will at the end of the year only be $9 million. In that case, company XYZ has to default on its debt. Company XYZ considers a new strategy with no upfront investment.

- **The probability of success of the new strategy is only 20%. Hence the probability of failure of the new strategy is 80%.**
- **If the new strategy is successfull, the value of the firm’s assets will increase to $15 million.**
- **If the new strategy is not successfull, the value of the firm’s assets will fall to $5 million.**

	Old Strategy [mln]	Success [mln]	Failure [mln]
Value of Assets	9	15	5
Debt	9	10	5
Equity	0	5	0

QUESTION

Consider the table above. Calculate the expected value of the XYZ’s assets under the new strategy. Is it beneficial for the equity and/or bond holders if company XYZ executes this strategy?

Solution.

We denote by the random variable X the possible values of XYZ’s assets. We get:

- $E[X] = 0.2 \cdot \$15mln. + 0.8 \cdot \$5mln. = \$7mln.$ **Hence, under the new strategy, the expected value of XYZ’s assets wi**ll decrease **from** $\$9mln.$ **to** $\$7mln.$
- **If company XYZ does nothing, it will ultimately default and equity holders will get nothing with certainty. If the new strategy is implemented and succeeds, equity holders will get $\$5$ million in total. The expected payoff under the new strategy is equal to** $0.8 \cdot \$0mln. + 0.2 \cdot \$5mln. = \$1mln.$ **Therefore equity holders have** nothing to lose **from this strategy.**
- **If company XYZ does nothing, it will ultimately default and debt holders will get $9 million with certainty. If the new strategy is implemented and**

succeeds, debt holders will get $10 million in total. If the new strategy is implemented and fails, debt holders will get $5 million in total. The expected payoff under the new strategy is equal to $0.8 \cdot \$5mln. + 0.2 \cdot \$10mln. = \$6mln$**. Therefore debt holders have** a lot to lose **from this strategy.**

The results are summarized in the table below.

	Old Strategy [mln]	Success [mln]	Failure [mln]	Expected [mln]
Value of Assets	9	15	5	7
Debt	9	10	5	6
Equity	0	5	0	1

Effectively, the equity holders are gambling with the debt holder's money. Shareholders have an incentive to invest in negative-NPV projects (where NPV stands for Net Present Value) that are risky, even though a negative-NPV project destroys value for the firm overall.

TAKE AWAY

RANDOM VARIABLE

A random variable (usually denoted with a capital letter) is a variable whose value is subject to variations due to chance.

INDIVIDUAL VALUE

A value that is not subject to variations due to chance, denoted by lowercase letters x, y, z, etc.

PROBABILITY

The likelihood that a random variable is equal to an individual value.

PROBABILITY MASS FUNCTION

A probability mass function (abbreviated PMF) completely describes the probability properties of a discrete random variable. It shows the probability that a discrete random variable X is exactly equal to some value.

PROBABILITY DENSITY FUNCTION

A probability density function (abbreviated PDF) describes the relative likelihood for a continuous random variable X on a given value.

PROBABILITY DISTRIBUTION

A probability distribution is a statistical function that shows the possible values and likelihoods that a discrete or continuous random variable can take within a given range. There exist many different probability distributions.

CUMULATIVE DISTRIBUTION FUNCTION

The cumulative distribution function (abbreviated CDF), shows the probability that a random variable X takes a value less than or equal to x.

Notation: $F(x) = P(X \leq x)$, where $-\infty < x < \infty$.

HISTOGRAM

A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous random variable. It shows for example the counts of the number of times the stock price had a certain increase or decrease.

MEASURE OF LOCATION

The arithmetic mean is normally used as a measure of location/central tendency.

MEASURE OF STATISTICAL DISPERSION

The standard deviation is normally used as a measure of statistical dispersion.

EXPECTED VALUE

The expected value of a discrete random variable X is the probability-weighted sum of all possible values.

Notation: $E[X] = x_1\, p(x_1) + x_2\, p(x_2) + \ldots + x_k\, p(x_k)$, for k possible values, $x_{1,}\ x_2,\ \ldots\ x_k$ for X.

VARIANCE: GENERAL CASE

The variance of a random variable X is the expected value of its squared deviations from $E[X]$.

Notation: $\sigma^2 = Var[X] = E\left[(X - E[X])^2\right] = E\left[X^2\right] - (E[X])^2$, where $E[X]$ is the expected value of X and $E\left[X^2\right]$ is the second moment of X. Note that $\mu = E[X]$.

VARIANCE: DISCRETE CASE

Only the notation changes for the variance.

$Var[X] = \sum_{i=1}^{n} (x_i - \mu)^2\, p(x_i)$,

where $p(x_i)$ is the probability of X being equal to x_i. When all n values are equally likely to occur, the expression simplifies to $Var[X] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$, where μ is the mean of X.

STANDARD DEVIATION

The standard deviation of a random variable X is the square-root of the variance of $X.$

Notation: $SD[X] = \sigma = \sqrt{VAR[X]}$.

ADDING A CONSTANT

1. $E[X \pm a] = E[X] \pm a.$
2. $Var[X \pm a] = Var[X]$.
3. $SD[X \pm a] = SD[X]$.

MULTIPLYING BY A CONSTANT

1. $E[c \cdot X] = c \cdot E[X]$.
2. $Var[c \cdot X] = c^2 \cdot Var[X]$.
3. $SD[c \cdot X] = |c| \cdot SD[X]$, where $|c|$ is the absolute value of c.

LINEAR FUNCTIONS OF *X*

This follows from adding and multiplying a constant.

1. $E[a + c \cdot X] = a + c \cdot E[X]$.
2. $Var[a + c \cdot X] = c^2 \cdot Var[X]$.
3. $SD[a + c \cdot X] = |c| \cdot SD[X]$.

SHARPE RATIO

The Sharpe ratio is the average return μ earned in excess of the risk-free rate r_f per unit of volatility (equal to SD) σ or total risk. By subtracting the risk-free rate r_f from the mean return μ, the performance associated with risk-taking activities can be isolated. The higher the Sharpe Ratio, the better the performance of the investment

Notation: $S(X) = \frac{\mu_X - r_f}{\sigma_X}$, where X is the investment opportunity.

BEST PRACTICES

1. Use random variables to represent uncertain outcomes.
2. Calculate the standard deviation by calculating the square-root of the variance.
3. Apply the formula for means and variances carefully, i.e. multiply each probability by the appropriate weight.

PITFALLS

1. Do not mix up X with $x.$
2. Apply the square after you deducted the mean in the variance formula.
3. Understand that the mean or expectation of a discrete random variable could possibly never be attained.

BINOMIAL DISTRIBUTION

INTRODUCTION

FLORIDA ELEVATED ROADWAYS

Florida Elevated Roadways INC. (“FERI”), is a construction company in Southern Florida that is specialized in building elevated roadways. FERI’s most recent construction project included building 218 pillars for an elevated expressway near Miami. The Florida Department of Transportation (FDOT) carefully evaluates each and every pillar whether it satisfies the proper FDOT specifications.

Toshi D., a Florida construction engineer and proud Kellogg alumnus, tells us that FDOT classifies such pillars as a “failure” if they do not satisfy its pre-specified technical requirements. In Toshi’s experience, the probability that a pillar is classified as a failure in a project like the one in Miami is 0.5%. Toshi also informs us that we can safely assume that all pillars are constructed independently of each other.

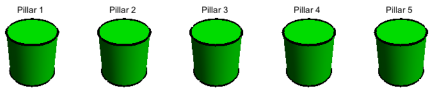
We are interested in the probability that for example ‘none’, ‘at most one’ or ‘at least two’ of the 218 pillars will fail. We are also interested in how many pillars we can expect to fail out of the total number of pillars and the standard deviation of the total number of failed pillars in this construction project.

EXAMPLES

EXAMPLE 1: FIVE PILLARS

Before we analyze the case with 218 pillars, let us first assume that there are 5 pillars in total. The probability that a pillar is classified as a failure is still $0.5\ \%$ and we can still safely assume that all pillars are constructed independently from each other.

QUESTION



What is the probability that none of the five pillars is classified as a failure?

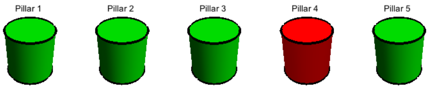
Solution.

We denote by n_i the event that pillar i (with $i = 1,\ \ldots,\ 5$), is not classified as a failure. We need to calculate the following probability: $P(n_1 \cap n_2 \cap n_3 \cap n_4 \cap n_5)$. The probability that pillar i is classified as a failure is independent from the probability that pillar j (with $j = 1, \ldots, 5$ and $j \neq i$) is classified as a failure. Using this independence, we can rewrite this probability as a product of all the individual probabilities. We get:

$P(n_1 \cap n_2 \cap n_3 \cap n_4 \cap n_5) \overset{Independence}{=} P(n_1) \cdot P(n_2) \cdot P(n_3) \cdot P(n_4) \cdot P(n_5)$. It is given that $P(n_i) = 0.995$ (for $i = 1, \ldots, 5$). Therefore we get $P(n_1 \cap n_2 \cap n_3 \cap n_4 \cap n_5) = 0.995^5 \approx 0.975$. Therefore, the probability that none of the five pillars is classified as a failure is approximately $97.5\ \%$.

QUESTION

What is the probability that exactly the fourth pillar is classified as a failure?



Solution.

We denote by n_i the event that pillar i (with $i = 1, 2, 3, 5$), is not classified as a failure. We denote by f_4 the event that pillar 4 is classified as a failure. We need to calculate the following probability: $P(n_1 \cap n_2 \cap n_3 \cap f_4 \cap n_5)$. Using the fact that the probability that pillar i is classified as a failure is independent from the probability that pillar j (with $j \neq i$) is classified as a failure, we can rewrite this probability as a product of all the individual probabilities. We get:

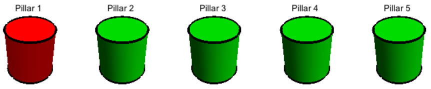
$P(n_1 \cap n_2 \cap n_3 \cap f_4 \cap n_5) \overset{Independence}{=} P(n_1) \cdot P(n_2) \cdot P(n_3) \cdot P(f_4) \cdot P(n_5)$. Since $P(n_i) = 0.995$ for $i = 1, 2, 3, 5$ and $P(f_4) = 0.005$, we get:

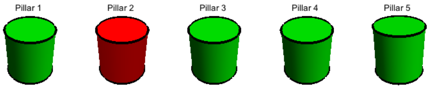
$P(n_1 \cap n_2 \cap n_3 \cap f_4 \cap n_5) = 0.995^3 \cdot 0.005 \cdot 0.995 \approx 0.0049.$

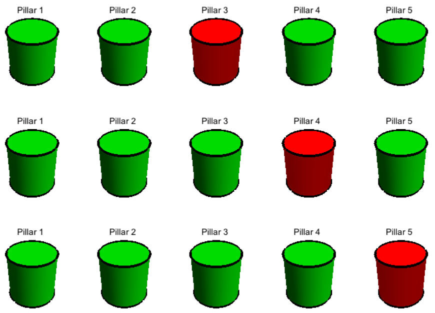
Therefore, the probability that exactly the fourth pillar is classified as a failure is approximately $0.49\ \%$.

QUESTION

What is the probability that exactly one of the five pillars is classified as a failure?







Solution.

Note that we have now five different possibilities where a pillar can be classified as a failure. We denote by n_i the event that pillar i (ranging from 1 to 5) is not classified as a failure and by f_j the event that pillar j is classified as a failure (also ranging from 1 to 5 and $j \neq i$). We need to calculate the following probability:

$$P(f_1 \cap n_2 \cap n_3 \cap n_4 \cap n_5) + P(n_1 \cap f_2 \cap n_3 \cap n_4 \cap n_5) + P(n_1 \cap n_2 \cap f_3 \cap n_4 \cap n_5) + P(n_1 \cap n_2 \cap n_3 \cap f_4 \cap n_5) + P(n_1 \cap n_2 \cap n_3 \cap n_4 \cap f_5).$$

We can rewrite each probability of the intersection of five events as the product of the individual probabilities of the five events. We get:

$$[P(f_1) \cdot P(n_2) \cdot P(n_3) \cdot P(n_4) \cdot P(n_5)] + [P(n_1) \cdot P(f_2) \cdot P(n_3) \cdot P(n_4) \cdot P(n_5)] + [P(n_1) \cdot P(n_2) \cdot P(f_3) \cdot P(n_4) \cdot P(n_5)] + [P(n_1) \cdot P(n_2) \cdot P(n_3) \cdot P(f_4) \cdot P(n_5)] + [P(n_1) \cdot P(n_2) \cdot P(n_3) \cdot P(n_4) \cdot P(f_5)] =$$

$$[0.005 \cdot 0.995^4] + [0.995 \cdot 0.005 \cdot 0.995^3] + [0.995^2 \cdot 0.005 \cdot 0.995^2] + [0.995^3 \cdot 0.005 \cdot 0.995] + [0.995^4 \cdot 0.005] = 5 \cdot 0.995^4 \cdot 0.005$$

$$\approx 0.0245$$

Therefore, the probability that exactly one of the five pillars is classified as a failure is approximately equal to $2.45\ \%$.

QUESTION

What is the probability that exactly 2 of the five pillars are classified as a failure?

Solution.

Note that in this case, we can have 10 different possibilities. They are listed below.

Using similar notation as in previous questions, we need to calculate the following probability:

$$P(f_1 \cap f_2 \cap n_3 \cap n_4 \cap n_5) +$$
$$P(f_1 \cap n_2 \cap f_3 \cap n_4 \cap n_5) + P(f_1 \cap n_2 \cap n_3 \cap f_4 \cap n_5) +$$
$$P(f_1 \cap n_2 \cap n_3 \cap n_4 \cap f_5) +$$
$$P(n_1 \cap f_2 \cap f_3 \cap n_4 \cap n_5) + P(n_1 \cap f_2 \cap n_3 \cap f_4 \cap n_5) +$$
$$P(n_1 \cap f_2 \cap n_3 \cap n_4 \cap f_5) +$$
$$P(n_1 \cap n_2 \cap f_3 \cap f_4 \cap n_5) + P(n_1 \cap n_2 \cap f_3 \cap n_4 \cap f_5) +$$
$$P(n_1 \cap n_2 \cap n_3 \cap f_4 \cap f_5)$$

This becomes quite cumbersome and messy. We will continue this example later on in this lecture.

POSSIBLE COMBINATIONS

Below are all the combinations listed for the case that two of the five pillars are classified as a failure.

- **Pillar 1 & 2, Pillar 1 & 3, Pillar 1 & 4, Pillar 1 & 5,**
- **Pillar 2 & 3, Pillar 2 & 4, Pillar 2 & 5,**
- **Pillar 3 & 4, Pillar 3 & 5,**
- **Pillar 4 & 5.**

FROM BERNOULLI TO BINOMIAL

INTRODUCTION

As we could see in the previous example, calculating the probability that exactly two of the five pillars are classified as a failure becomes quite a cumbersome calculation. Imagine how cumbersome this calculation would become if we want to calculate the probability that exactly two out of 218 pillars are classified as a failure. We create a new framework in order to do these kind of computations faster and more structured.

BERNOULLI RANDOM VARIABLES - DEFINITIONS

Bernoulli Random Variable

A Bernoulli random variable X considers only two outcomes: "success" (with probability p) or "failure" (with probability $1 - p$).

Notation: **$X \sim Ber(p)$.**

Bernoulli Trials

Bernoulli Trials are random experiments with three characteristics:

1. There are two possible outcomes (success or failure). Put differently, success and failure are mutually exclusive and collectively exhaustive.

2. Constant probability of success (p) and failure ($1 - p$).

3. Individual trials are independent. Put differently, the outcome of one trial does not affect the outcome of another trial.

Expected Value

The expected value of a Bernoulli random variable X is equal to p, where p is the probability of success.

Notation: **$E[X] = p$.**

Variance

The variance of a Bernoulli random variable X is equal to $p \cdot (1 - p)$, where p is the probability of success and $(1 - p)$ is the probability of failure.

Notation: **$Var[X] = p \cdot (1 - p)$.**

Standard Deviation

The standard deviaton of a Bernoulli random variable X is equal to $\sqrt{p \cdot (1 - p)}$, the square-root of the variance.

Notation: **$SD[X] = \sqrt{p \cdot (1 - p)}$.**

BERNOULLI RANDOM VARIABLES - PMF AND CDF

Let X be a Bernoulli random variable. Also, let $k = 1$ denote success and $k = 0$ denote failure. The probability mass function (in formula form) of X is then given by:

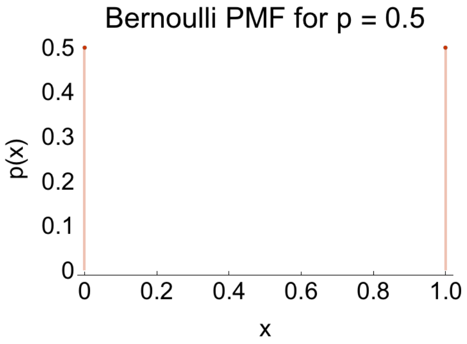
$$\begin{cases} 1-p & k == 0 \\ p & k == 1 \\ 0 & \text{True} \end{cases}$$

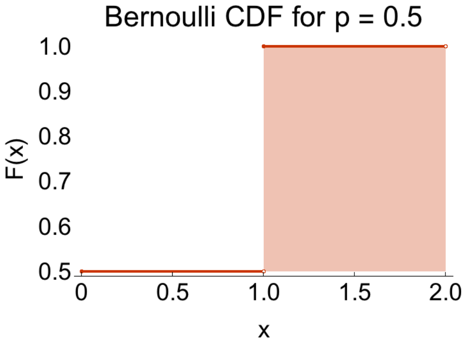
This means that there is a probability equal to p that $k = 1$ (i.e. success) and a probability equal to $1 - p$ that $k = 0$ (i.e. failure).

The corresponding cumulative distribution function (in formula form) of X is given by:

$$\begin{cases} 0 & k < 0 \\ 1-p & 0 \leq k < 1 \\ 1 & \text{True} \end{cases}$$

The probability mass function of X for $p = 0.5$ is shown in the top figure and the corresponding cumulative distribution function of X is shown in the bottom figure below.





EXAMPLES OF BERNOULLI TRIALS

- **A manufacturing plant labels items as either defective or acceptable.**
- **A firm bidding for contracts will either get the contract or not.**
- **A marketing research firm receives survey responses of “yes I will buy” or “no I will not”.**
- **A successful job applicant either accepts the offer or rejects it.**

BINOMIAL RANDOM VARIABLES

Binomial Random Variable

A binomial random variable X is the sum of n independent and identically distributed (IID) Bernoulli random variables. The random variable X is defined by two parameters, n (the total number of trials) and p (the probability of success). It is useful for calculating the probability of getting exactly k successes in n trials. This is exactly given by the probability mass function.

Notation: $X \sim Bin\left(n,\ p\right)$.

Expected Value

The expected value of a Binomial random variable X is equal to $n \cdot p$, the product of the number of trials and the probability of success.

Notation: $E[X] = n \cdot p$.

Variance

The variance of a Binomial random variable X is equal to $n \cdot p \cdot \left(1 - p\right)$, the number of trials times the probability of success times the probability of failure.

Notation: $Var[X] = n \cdot p \cdot \left(1 - p\right)$.

Standard Deviation

The standard deviaton of a Binomial random variable X is equal to the square-root of the variance.

Notation: $SD[X] = \sqrt{n \cdot p \cdot \left(1 - p\right)}$

DERIVATION OF EXPECTED VALUE, VARIANCE AND STANDARD DEVIATION FOR BINOMIAL RANDOM VARIABLES

Expected Value

Let $Y \sim Bin\left(n,\ p\right)$. By definition, Y can be written as the sum of n IID Bernoulli random variables, $X_1,\ X_2,\ \ldots,\ X_n$. Recall that the expected value of a Bernoulli random variable is equal to $p.$

We get:

$$E[Y] = E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] =$$

$$E[X_1] + E[X_2] + \ldots + E[X_n] = p + p + \ldots + p = n \cdot p.$$

Variance

Let $Y \sim Bin\left(n,\ p\right)$. By definition, Y can be written as the sum of n IID Bernoulli random variables, $X_1,\ \ldots,\ X_n$. Recall that the variance of a Bernoulli random variable is equal to $p \cdot \left(1 - p\right)$.

We get:

$$Var[Y] = Var[X_1 + X_2 + \ldots + X_n] \overset{Independence}{=}$$

$$Var[X_1] + Var[X_2] + \ldots + Var[X_n]$$

$$= p \cdot (1 - p) + p \cdot (1 - p) + \ldots p \cdot (1 - p) =$$

$$n \cdot (p \cdot (1 - p)) = n \cdot p \cdot (1 - p).$$

Standard Deviation

Let $Y \sim Bin(n, p)$. By definition, the standard deviation of any random variable is the square root of the variance. We get: $SD[Y] = \sqrt{n \cdot p \cdot (1 - p)}$

PROBABILITY MASS FUNCTION & CUMULATIVE DISTRIBUTION FUNCTION

PROBABILITY MASS FUNCTION

Probability Mass Function
The binomial distribution has two input parameters:

- n: the number of times the experiment is performed.

- p: the probability of success.

For a Binomial random variable X, the probability of getting k successes in n trials is given by the probability mass function:

$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$**, where** $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$**. Necessarily, we have that** $k \le n$**.**

FACTORIAL AND BINOMIAL COEFFICIENT

- **‘$n\,!$’ is pronounced as ‘n-factorial’. It is the product of all positive integers less than or equal to n. For example,** $6\,! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 720$**. Therefore,** $n\,! = n \cdot (n - 1) \cdot (n - 2) \cdot \ldots \cdot 2 \cdot 1$**.**
- **‘$\binom{n}{k}$’ is called the binomial coefficient. It is equivalent to** $\frac{n!}{k!\,(n-k)!}$**, where necessarily** $k \le n$**. For**

example, for $n = 7$ and $k = 3$, we get:

$$\binom{7}{3} = \frac{7!}{3!\cdot(7-3)!} = \frac{7!}{3!\cdot(4)!} = \frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{(3\cdot2\cdot1)\cdot(4\cdot3\cdot2\cdot1)} \overset{\left(\frac{4\cdot3\cdot2\cdot1}{4\cdot3\cdot2\cdot1}=1\right)}{=} \frac{7\cdot6\cdot5}{3\cdot2\cdot1} = \frac{210}{6} = 35.$$

CUMULATIVE DISTRIBUTION FUNCTION

Cumulative Distribution Function

The CDF, $F(x)$, of a discrete random variable X, expresses the probability that X does not exceed a fixed value x.

Notation: **$F(x) = P(X \le x)$.**

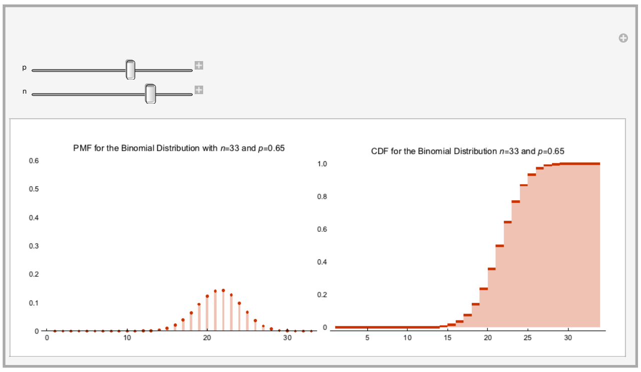
Probability as an Area

Let a and b be two possible values of X with $a < b$. The probability that X lies between a and b is $P(a < X < b) = F(b) - F(a)$, where F is the CDF of X.

EXAMPLES

PMF AND CDF: ILLUSTRATION

In the figure below, the probability mass function for a binomially distributed random variable X is visible for different n and p. The corresponding cumulative distribution function is visible in the figure below on the right.


p
n
PMF for the Binomial Distribution with n=33 and p=0.65
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
CDF for the Binomial Distribution n=33 and p=0.65
1.0
0.8
0.6
0.4
0.2
0
5
10
15
20
25
30

EXAMPLE 1 CONT’D: FIVE PILLARS

Consider again the FERI example from before.

QUESTION

Calculate the probability that exactly the fourth pillar is classified as a failure.

Solution.

We need to calculate the probability that the fourth pillar is classified as a failure. Therefore we cannot use the probability mass function of the binomial distribution. Since we would only be able to calculate the probability of exactly 1 failure in 5 trails (i.e. 4 successes in 5 trials) and not necessarily the probability that exactly the fourth pillar is classified as a failure.

Using the same approach as before, we need to calculate the following probability: $P(n_1 \cap n_2 \cap n_3 \cap f_4 \cap n_5)$. Recall that by n_i, we denote the event that pillar $i = 1, 2, 3, 5$ is not classified as a failure and by f_4 we denote the event that pillar 4 is classified as a failure. We get:

$$P(n_1 \cap n_2 \cap n_3 \cap f_4 \cap n_5) \overset{Independence}{=}$$

$$P(n_1) \cdot P(n_2) \cdot P(n_3) \cdot P(f_4) \cdot P(n_5)$$

$$= 0.995^3 \cdot 0.005 \cdot 0.995 \approx 0.0049.$$

Therefore, the probability that exactly the fourth pillar is classified as a failure is approximately $0.49\ \%$.

QUESTION

What is the probability that exactly one of the five pillars is classified as a failure?

Solution.

In this question, we are asked to calculate the probability of exactly 4 successes in 5 trials. Therefore we can use the probability mass function of the binomial distribution to answer this question. It is given that:

- **The number of trials n is equal to 5.**
- **The number of successes k is equal to 4.**
- **The probability of success p is equal to 0.995.**
- **The probability of failure $1 - p$ is equal to 0.005.**
- **The probability mass function for k-out-of-n successes is given by**

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

Let us denote by X the random variable that represents the number of pillars that do not fail. Clearly, $X \sim Bin\,(5,\ 0.995)$. We get:

$$P(X = 4) = \binom{5}{4} \cdot 0.995^4 \cdot 0.005^{5-4} = \binom{5}{4} \cdot 0.995^4 \cdot 0.005 \approx 0.0245.$$

Therefore there is a probability of approximately $2.45\ \%$ that exactly one of the five pillars is classified as a failure.

QUESTION

What is the probability that exactly two of the five pillars are classified as a failure?

In this question, we are asked to calculate the probability of exactly 3 successes in

5 trials. Therefore we can use the probability mass function of the binomial distribution to answer this question. It is given that:

- **The number of trials n is equal to 5.**
- **The number of successes k is equal to 3.**
- **The probability of success p is equal to 0.995.**
- **The probability of failure $1-p$ is equal to 0.005.**
- **The probability mass function for k-out-of-n successes is given by**

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

Let us denote by X the random variable that represents the number of pillars that do not fail. Clearly, $X\sim Bin\,(5,\ 0.995)$. We get:

$$P(X=3)=\binom{5}{3}\cdot 0.995^3\cdot 0.005^{5-3}=\binom{5}{3}\cdot 0.995^3\cdot 0.005^2\approx 0.000246.$$

There is a probability of approximately $0.025\ \%$ that two pillars are classified as a failure, hence a highly unlikely event.

APPLICATIONS

BEER BOTTLES

In 1981 during the halftime of the Super Bowl, the Joseph Schlitz Brewing Company broadcasted a live beer taste test for approximately 100 million viewers worldwide. Joseph Schlitz Brewing Company took 100 people who all preferred a Michelob beer. The 100 people conducted a blind taste test between their supposed favorite beer and Schlitz. Joseph Schlitz Brewing Company wanted to show that even beer drinkers who think they like another brand will prefer Schlitz in a blind taste test. In order to broadcast this, Joseph Schlitz Brewing Company paid $1.7 million.

How risky and naive is this strategy of the Joseph Schlitz Brewing Company?

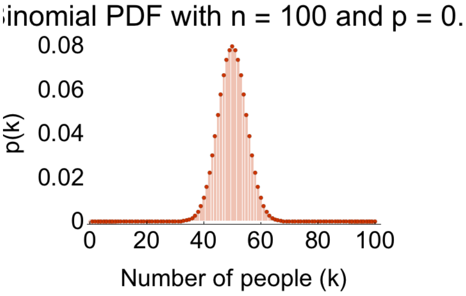
(A) Risky, but not naive
(B) Naive, but not risky
(C) Risky and naive
(D) Nor risky, nor naive

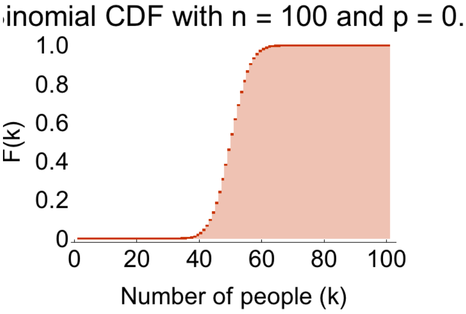
Solution.

In order to determine the correct answer, we need to know the underlying assumptions for this taste testing experiment.

- **The beer that was preferred by all the participants of the taste test, Michelob, actually had about the same taste as Schlitz beer. Therefore Schlitz beer did not have to taste particularly better in order to be chosen by (some) of the participants.**
- **Because of the similar taste, each participant would choose Schiltz over Michelob with a probability of about 50%. Schlitz would never conduct the test with people who preferred their own beer, because there is also a 50% chance that people would prefer the beer of the competitor, although they initially preferred Schlitz beer!**

By conducting the experiment only with people who liked the beer of the competitor better in the first place, Schlitz did something very smart: every single person who chose Schlitz was a win. This strategy is not naive. But is it risky? Let us first look at the probability distribution function and the cumulative distribution function.





We can clearly see in the top figure above that it is most likely that between 40 and 60 people will choose Schlitz over Michelob. We can also clearly see this in the CDF (the bottom figure above), since the CDF starts to increase around $k = 40$ people.

QUESTION

Marketing people say a taste test is a success if 40 people or more choose Schlitz. Calculate the probability that more than 40 people choose Schlitz.

Solution.

Let us denote by X the random variable that represents the number of people that choose Schlitz. Clearly, $X \sim Bin\ (100,\ 0.5)$. We need to calculate $P(X > 40)$, which is equal to $1 - P(X \leq 40)$. In the CDF in the previous question, we can see that the mass on the left hand side of 40 is quite small, so $P(X \leq 40)$ is quite small as well. As a result, $P(X > 40)$ will be large.

We need to calculate the following probability:

$$P(X > 40) =$$

$$1 - P(X \leq 40) = 1 - F(k; n,\ p) = 1 - \sum_{k=1}^{40} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} =$$

$$1 - \sum_{k=1}^{40} \binom{100}{k} \cdot 0.5^k \cdot 0.5^{100-k}.$$

This computation is very technical and hence we did it with *Mathematica*.

We get: $P(X > 40) = 1 - P(X \leq 40) = 1 - 0.0284 = 0.9716$. Therefore there is a chance of approximately $97\ \%$ that the taste test will be considered a success, i.e. that more than 40 people will choose Schlitz over Michelob.

TAKE AWAY

BERNOULLI RANDOM VARIABLE

A Bernoulli random variable X considers only two outcomes: “success” (with probability p) or “failure” (with probability $1 - p$).

Notation: $X \sim Ber(p)$

Expected value: $E[X] = p$

Variance: $Var[X] = p \cdot (1\ -\ p)$

Standard Deviation: $SD[X] = \sqrt{p \cdot (1 - p)}$

BINOMIAL RANDOM VARIABLE

A binomial random variable X is the sum of n independent and identically distributed Bernoulli random variables. The random variable X is defined by two parameters, n (the total number of trials) and p (the probability of success).

Notation: $X \sim Bin(n,\ \ p)$

Expected value: $E[X] = n \cdot p$

Variance: $Var[X] = np \cdot (1\ -\ p)$

Standard Deviation: $SD[X] = \sqrt{n \cdot p \cdot (1 - p)}$

PROBABILITY MASS FUNCTION

The binomial distribution has two input parameters:

- n: the number of times the experiment is performed.

- p: represents the probability of one specific outcome.

For a Binomial random variable X, the probability of getting k successes in n trials is given by the probability mass function:

$P(X = k) = \begin{pmatrix} n \\ k \end{pmatrix} \cdot p^k \cdot (1 - p)^{n-k}$, where $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k! \cdot (n-k)!}$ and $n! = n \cdot (n-1) \cdot \ldots \cdot 1$. Necessarily, we have that $k \leq n$.

CUMULATIVE DISTRIBUTION FUNCTION

The CDF of a discrete random variable X, denoted by $F(x)$, expresses the probability that X does not exceed a fixed value x.

Notation: $F(x) = P(X \leq x)$.

PROBABILITY AS AN AREA

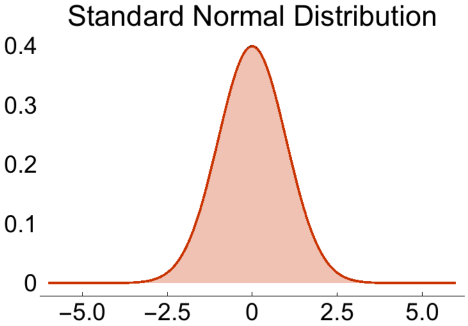
Let a and b be two possible values of X, with $a < b$. The probability that X lies between a and b is $P(a < X < b) = F(b) - F(a)$, where F is the CDF of X.

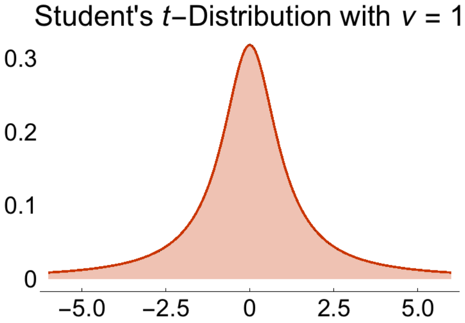
5 - NORMAL DISTRIBUTION AND STUDENT’S T-DISTRIBUTION

INTRODUCTION

NORMAL PDF VS. STUDENT’S T-DISTRIBUTION PDF

The normal distribution closely approximates the probability distributions of a wide range of random variables. The normal distribution is known for its iconic bell-shape. The student’s t-distribution is in an important way related to the normal distribution and therefore also treated in this lecture.


Standard Normal Distribution

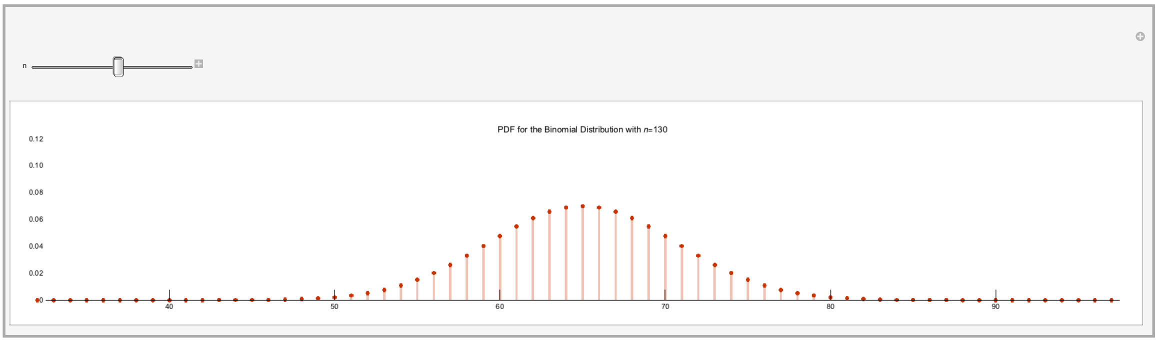

Student's t−Distribution with v = 1

The standard normal distribution (with $\mu = 0$ and $\sigma^2 = 1$) is visible in the top figure above. The student’s t-distribution (with $\nu = 1$, which means that it has 1 degree of freedom) is visibile in the bottom figure above. A notable difference is that the ‘tails’ of the student’s t-distribution are thicker. The tails are said to have more mass.

CONTINUOUS DISTRIBUTIONS

INTRODUCTION

The figure below shows the binomial distribution for a fixed probability p of success, but n varies from 50 to 200. From a graphical perspective, the larger n, the more the PDF gets bell-shaped.



PDF for the Binomial Distribution with n=130

NORMAL DISTRIBUTION

Normal Distribution
The normal distribution is a continuous probability distribution that is centered around the mean, bell-shaped, symmetric and is completely determined by two input paramters: the mean μ and the variance σ^2.

Notation: **A normally distributed random variable X with mean μ and variance σ^2 is denoted by $X \sim N(\mu, \sigma^2)$.**

Standard Normal Distribution

The standard normal distribution is a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. Any normally distributed random variable X with mean μ and variance σ^2 can be rewritten as a standard normal random variable Z in the following way:

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

Probability Mass Function
The normal distribution has two input parameters:

- μ: the mean of the distribution, located at the center of the probability density function.

- σ^2: the variance of the distribution.

The probability density function of the normal distribution with mean μ and variance σ^2 is given by: $f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

Cumulative Distribution Function

The CDF, $F(x)$, of a continuous random variable X expresses the probability that X does not exceed a fixed value x.

Notation: $F(x) = P(X \leq x)$**.**

Probability as an Area

Let a and b be two possible values of X, with $a < b$. The probability that X lies between a and b is $P(a < X < b) = F(b) - F(a)$, where F is the CDF of X.

EXAMPLES OF NORMALLY DISTRIBUTED RANDOM VARIABLES

- **The thickness of an item.**
- **The time required to complete a task.**
- **The height of people, in inches.**

EXAMPLES

EXAMPLE 1: FLORIDA ELEVATED ROADWAYS

Florida Elevated Roadways, Inc. ("FERI") is a construction company in Southern Florida that is specialized in building elevated roadways. FERI's most recent construction project included building 218 pillars for an elevated expressway near Miami. Such pillars are classified as a "failure" if they do not satisfy some pre-specified technical requirements. One of the technical requirements states that some parameter "x" describing the pillar must have a value of at most 1. If the value x exceeds 1, the pillar is classified as a failure.

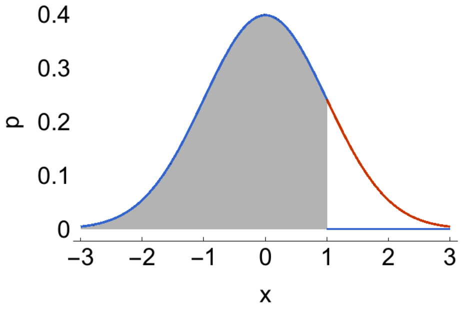
FERI's construction process for pillars achieves, on average, a value for x of 0.8. But FERI does not reach this exact value on every pillar. Its long-term experience shows that the value of x has a mean of 0.8 and a standard deviation of 0.1.

QUESTION

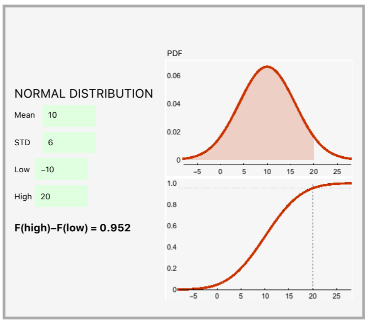
What is the probability that a pillar will fail?

Solution.

A pillar is classified as a failure if the value x is larger than 1.0. We use the normal distribution to model the parameter x of a pillar. Let us denote by X the random variable that represents x. Clearly, $X \sim N\left(0.8,\ 0.1^2\right)$. We need to calculate $P(X > 1) = 1 - P(X \leq 1)$. The shaded area in the figure below represents $P(X \leq 1)$.



We calculate $P(X \leq 1)$ below:



NORMAL DISTRIBUTION

Mean 10

STD 6

Low −10

High 20

F(high)−F(low) = 0.952

The probability that a pillar will fail therefore is equal to $1 - 0.97725 \approx 0.023$, i.e the chance that a particular pillar will fail is therefore about $2.3\ \%$. Since FERI builds many pillars, this rather high probability implies that FERI should be prepared to experience pillar failures in projects which involve the construction of many pillars.

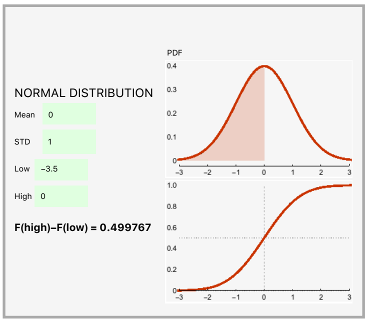
QUESTION

Calculate again the probability that a pillar will fail. Transform the random variable $X \sim N\left(0.8,\ (0.1)^2\right)$ to a standard normal variable Z and then perform the calculation.

Solution.

It is given that $X \sim N\left(0.8,\ (0.1)^2\right)$. Standardization of X leads to $Z = \frac{X - 0.8}{0.1}$. We get:

$P(X \leq 1) = P\left(\frac{X - \mu}{\sigma} \leq \frac{1 - \mu}{\sigma}\right) = P\left(Z \leq \frac{1 - 0.8}{0.1}\right) = P(Z \leq 2)$. Calculating this yields:



NORMAL DISTRIBUTION

Mean 0

STD 1

Low −3.5

High 0

F(high)–F(low) = 0.499767

The probability that a particular pillar will fail is therefore $1 - 0.97725 \approx 0.023$. Note that this approach leads to the exact same outcome as in the previous question.

INVERSE CALCULATIONS

INVERSE NORMAL DISTRIBUTION

Inverse Normal Distribution

With the inverse normal distribution, it is possible to calculate the value x such that $P(X \leq x)$ is equal to a probability p that is usually given for a normally distributed random variable.

EXAMPLES

EXAMPLE 1 CONT'D: FLORIDA ELEVATED ROADWAYS

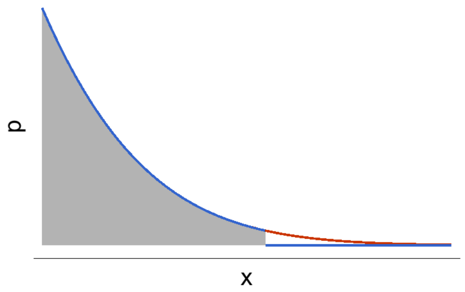
Recall that FERI's construction process for pillars achieves, on average, a value for x of 0.8. But FERI does not reach this exact value on every pillar. Its long-term experience shows that the value of x has a mean of 0.8 and a standard deviation of 0.1.

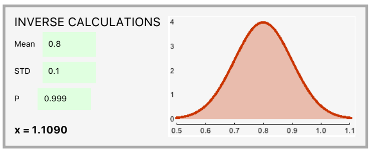
QUESTION

Calculate the value for x that is not exceeded with a probability of 99.9%.

Solution.

The mean $\mu = 0.8$, the standard deviation $\sigma = 0.1$ and the probability $p = 0.999$. Below is the right part of the normal distribution shown. The shaded area is 0.999 and ends precisely at the value for x that we need to find.




INVERSE CALCULATIONS

Mean 0.8

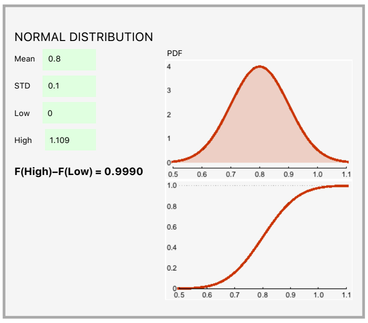
STD 0.1

P 0.999

x = 1.1090

So, with a probability of $99.9\ \%$ the value for x falls below the level of 1.109.

We can check this result by putting the value for x that we found in the normal CDF and see what probability comes out.


NORMAL DISTRIBUTION

Mean 0.8

STD 0.1

Low 0

High 1.109

F(High)−F(Low) = 0.9990

This is indeed approximately equal to $99.9\ \%$.

EXAMPLE 2: PROPERTY OF NORMAL DISTRIBUTIONS

Let X be a normally distributed random variable with mean μ and variance σ^2, so $X \sim N\left(\mu,\ \sigma^2\right)$.

QUESTION

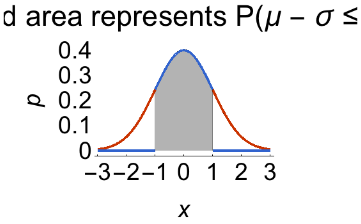
Calculate the following three probabilities:

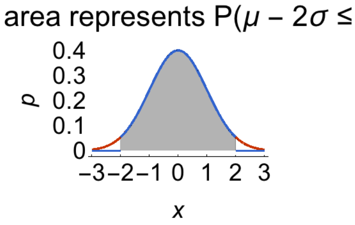
_ $P(\mu - \sigma \ \leq X \leq \mu + \sigma)$.

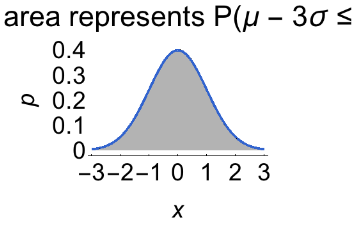
_ $P(\mu - 2\,\sigma \leq X \leq \mu + 2\,\sigma)$.

_ $P(\mu - 3\,\sigma \leq X \leq \mu + 3\,\sigma)$.

Solution.







- $P(\mu - \sigma \leq X \leq \mu + \sigma)$**:**
  $P(\mu - \sigma < X < \mu + \sigma) =$
  $$P\left(\frac{\mu-\sigma-\mu}{\sigma} < Z < \frac{\mu+\sigma-\mu}{\sigma}\right) = P\left(\frac{-\sigma}{\sigma} < Z < \frac{\sigma}{\sigma}\right) = F\left(\frac{\sigma}{\sigma}\right) - F\left(\frac{-\sigma}{\sigma}\right)$$
  $= F(1) - F(-1) = 0.8413 - 0.1587 = 0.6826$**. We can conclude that about $68\ \%$ of the total mass lies in the interval $[\mu - \sigma,\ \ \mu + \sigma]$. Hence, there is a probability of about $68\ \%$ that a normally distributed random variable X lies in the interval $[\mu - \sigma,\ \mu + \sigma]$. This is visualized in the top figure above.**
- $P(\mu - 2\sigma \leq X \leq \mu + 2\sigma)$**:**
  $P\ (\mu - 2\ \sigma < X < \mu + 2\ \sigma)\ = P\left(\frac{\mu-2\ \sigma-\mu}{\sigma} < Z < \frac{\mu+2\ \sigma-\mu}{\sigma}\right) =$
  $$P\left(\frac{-2\ \sigma}{\sigma} < Z < \frac{2\ \sigma}{\sigma}\right) = F\left(\frac{2\ \sigma}{\sigma}\right) - F\ \left(\frac{-2\ \sigma}{\sigma}\right)$$
  $= F(2) - F(-2) = 0.9772 - 0.0228 = 0.9544$**. Based on the question above, about $95\ \%$ of the total mass lies in the interval $[\mu - 2\ \sigma,\ \mu + 2\ \sigma]$. Hence, there is a probability of about 95% that a normally distributed random variable X lies in the interval $[\mu - 2\ \sigma,\ \mu + 2\ \sigma]$. This is visualized in the middle figure above.**
- $P(\mu - 3 \cdot \sigma \leq X \leq \mu + 3 \cdot \sigma)$**:**
  $P(\mu - 3\ \sigma < X < \mu + 3\ \sigma) = P\left(\frac{\mu-3\ \sigma-\mu}{\sigma} < Z < \frac{\mu+3\ \sigma-\mu}{\sigma}\right) =$
  $$P\left(\frac{-3\ \sigma}{\sigma} < Z < \frac{3\ \sigma}{\sigma}\right) = F\left(\frac{3\ \sigma}{\sigma}\right) - F\left(\frac{-3\ \sigma}{\sigma}\right)$$

$= F(3) - F(-3) = 0.9974$. Based on the question above, about $99.7\ \%$ of the total mass lies in the interval $[\mu - 3\,\sigma,\ \mu + 3\,\sigma]$. Hence, there is a probability of about $99.7\ \%$ that a normally distributed random variable X lies in the interval $[\mu - 3\,\sigma,\ \mu + 3\,\sigma]$. This is visualized in the bottom figure above.

QUANTILES

QUANTILES

Quantile

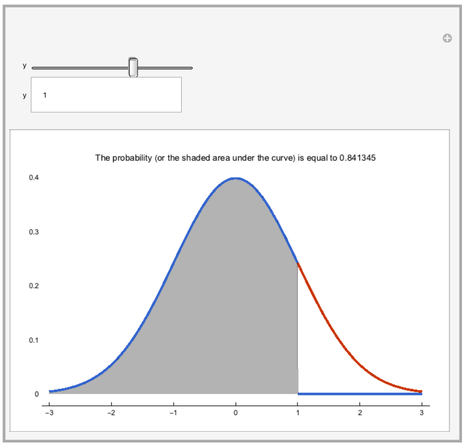
The p^{th} quantile (or percentile) of a probability distribution is the value x such that $P(X \le x) = p$. If X is normally distributed, $X \sim N\left(\mu,\ \sigma^2\right)$, then the x value such that $P(X \le x) = p$ is $x = InverseCDF\left[NormalDistribution[\mu,\ \sigma],\ p\right]$.

Value-at-Risk

Given a confidence level $\alpha \in (0,\ 1)$, Value-at-Risk (VaR) is defined as $VaR_\alpha = F_L{}^{-1}(\alpha) = inf\ \{x \in \mathbb{R} : F_L(x) \ge \alpha\}$. VaR is simply the α-quantile of a loss distribution F_L. Normally, we assume that F_L follows the normal distribution. For a given portfolio, time horizon, and probability p, the $100\ p\ \%$ VaR is defined as a threshold loss value, such that the probability that the loss on the portfolio exceeds this value over the given time horizon is p.

EXAMPLES

VISUALIZATION OF QUANTILES



y: 1

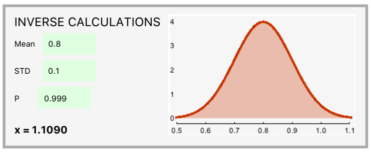
The probability (or the shaded area under the curve) is equal to 0.841345

QUESTION

FERI wants the probability of a failure to be at most $0.1\ \%$. Suppose it continues to build pillars with an average value for the parameter x of 0.8 and a standard deviation of 0.1. How large would the threshold for x (upper bound on x) have to be?

Solution.

It is given that $\mu = 0.8$, $\sigma = 0.1$ and $p = 1 - 0.001 = 0.999$. We need to calculate the threshold x. We get:



INVERSE CALCULATIONS

Mean: 0.8

STD: 0.1

P: 0.999

x = 1.1090

This function calculates the value x such that the probability of $p = F(X \leq x)$.

With a probability of $99.9\ \%$, the parameter x falls below the level 1.109.

EXAMPLE 3: VALUE-AT-RISK

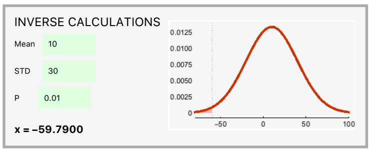
Suppose the $1 million portfolio of an investor is expected to grow on average 10 over the next year with a standard deviation of 30.

QUESTION

- **What is the $95\ \%$ VaR (Value-at-Risk)?**
- **What is the $99\ \%$ VaR (Value-at-Risk)?**

Solution.

Let the random variable X denote the percentage change next year in the portfolio. We model it by using the normal distribution with mean $\mu = 10$ and variance $\sigma^2 = 30^2$. We get that $X \sim N\left(10,\ 30^2\right)$. Calculating these two probabilities yields:



INVERSE CALCULATIONS

Mean 10

STD 30

P 0.01

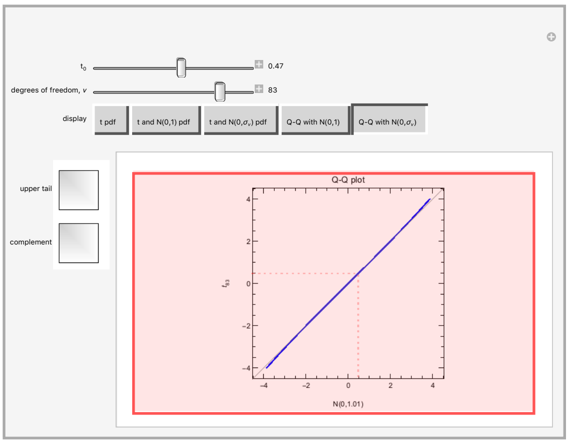
x = −59.7900

Therefore, the annual $95\ \%$ VaR for this portfolio is $\$393.456$. That is, there is a probability of $5\ \%$ that this portfolio loses more than $\$393,\ 456$. Put differently, with a probability of $95\ \%$, the portfolio will lose less than $\$393.456$.

The annual $99\ \%$ VaR for this portfolio is $\$597,\ 904$. That is, there is a probability of $1\ \%$ that this portfolio loses more than $\$597,\ 904$. Put differently, with a probability of $99\ \%$, the portfolio will lose less than $\$597.904$.

STUDENT’S T-DISTRIBUTION

INTRODUCTION



t_0 0.47

degrees of freedom, v 83

display: t pdf | t and N(0,1) pdf | t and N(0,σ_v) pdf | Q-Q with N(0,1) | [✓ Q-Q with N(0,σ_v)]

upper tail □

complement □

```
ParametricPlot:
{(-1 + 1/InverseBetaRegularized[2 p, 83/2, 1/2]) – 0, (-1 + 1/InverseBetaRegularized[2 (1 + Times[«2»]), 83/2, 1/2]) – 0, InverseBetaRegularized[2 p, 83/2, 1/2] – 0, InverseBetaRegularized[2 (1 – p), 83/2, 1/2] – 0, Max[- 1/2 + p, –p] – 0, Max[1/2 – p, –1 + p] – 0, 2 p – 2, 2 p – 0, p – 0, (- 1/2 + p) – 0, (1 – p) – 0, Im[
    1/InverseBetaRegularized[2 p, 83/2, 1/2]] – 0, Im[1/InverseBetaRegularized[2 (1 + Times[«2»]), 83/2, 1/2]] – 0} must be a list of equalities or real–valued functions.
```

STUDENT’S T-DISTRIBUTION

Student’s t-Distribution

In probability and statistics, the Student’s t-distribution is a symmetric and bell-shaped continuous probability distribution. The Student’s t-distribution is important when estimating the mean of a normally distributed population in situations where the sample size is small and the

population standard deviation is unknown. A Student's t-distribution is characterized by one parameter, which is called the degrees of freedom. This parameter is usually denoted by ν.

Normal Distribution vs. Student's t-Distribution

A normal distribution describes a full population with known parameters μ and σ. When only a sample of the full population is drawn, μ and σ have to be estimated. The Student's t-distribution describes samples drawn from a full population; accordingly, the Student's t-distribution for each sample size is different, and the larger the sample, the more the distribution resembles a normal distribution.

Degrees of Freedom

The degrees of freedom parameter ν in a Student's t-distribution is equal to $n-1$, where n is the number of observations in the sample. If $\nu \to \infty$ (infinity), the Student's t-distribution converges to the normal distribution with its true parameters μ and σ.

FORMAL DEFINITION OF THE STUDENT'S T-DISTRIBUTION

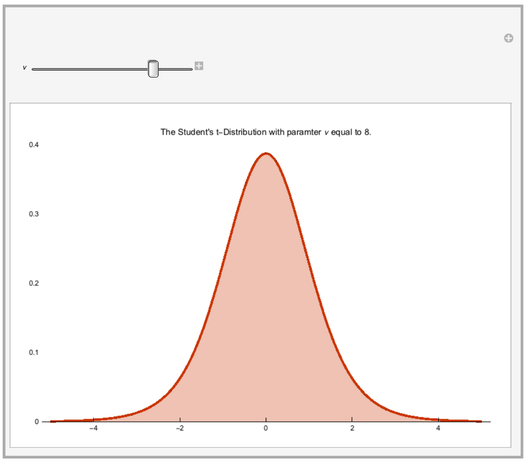
Student's t-Distribution

If $X \sim N(0,\ 1)$ and $Y \sim \chi^2(\nu)$ and X and Y are independently distributed, then the distribution of $\frac{X}{\sqrt{\frac{Y}{r}}}$ is called the Student's t-distribution with ν degrees of freedom, written as $\frac{X}{\sqrt{\frac{Y}{r}}} \sim t(\nu)$.

EXAMPLES

VISUALIZATION OF THE STUDENT'S T-DISTRIBUTION

In the figure below, the Student's t-distribution is shown for different degrees of freedom ν.



SHAPE AND FAT TAILS

The Student's t-distribution is, as we can see above, bell-shaped and symmetric. However, we see that the smaller the degrees of freedom, the lower the 'peak' of the PDF and the more mass there is at the right and left part of the PDF. These parts are called the tails of the distribution. Remember that for $\nu \rightarrow \infty$, the Student's t-distribution converges to a normal distribution. The mass of the tails is larger for a small ν. Also, the mass in the tail of a Student's t-distribution with a small ν high compared to the mass in the tails of a normal distribution. Therefore we say that the Student's t-distribution has fat tails. Fat tails indicate that it is more likely that an observation far from the true mean μ is observed. Student's t-distributions are used in risk management, where extreme scenarios should be more likely than in the normal case.

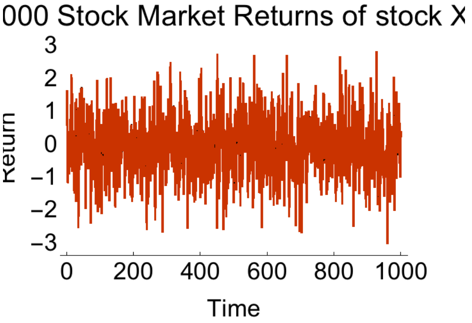
APPLICATIONS I

IGNORING FAT TAILS: LONG-TERM CAPITAL MANAGEMENT L.P. (LTCM)

An American based hedge fund, called Long-Term Capital Management L.P. (LTCM), collapsed in the late 1990's and billions of dollars disappeared overnight by making the wrong bet on the outcome of a trade. LTCM did not consider large low-probability events, because they did considered very large losses unlikely. They based the likelilood of very large losses on past performance of the stock market. They simply ignored fat tails and an unlikely but possible event happened, which made billions of dollars disappear overnight.

STOCK MARKET RETURNS: IDENTIFYING FAT TAILS I

Consider the histogram of 10.000 stock market returns of stock XYZ below.

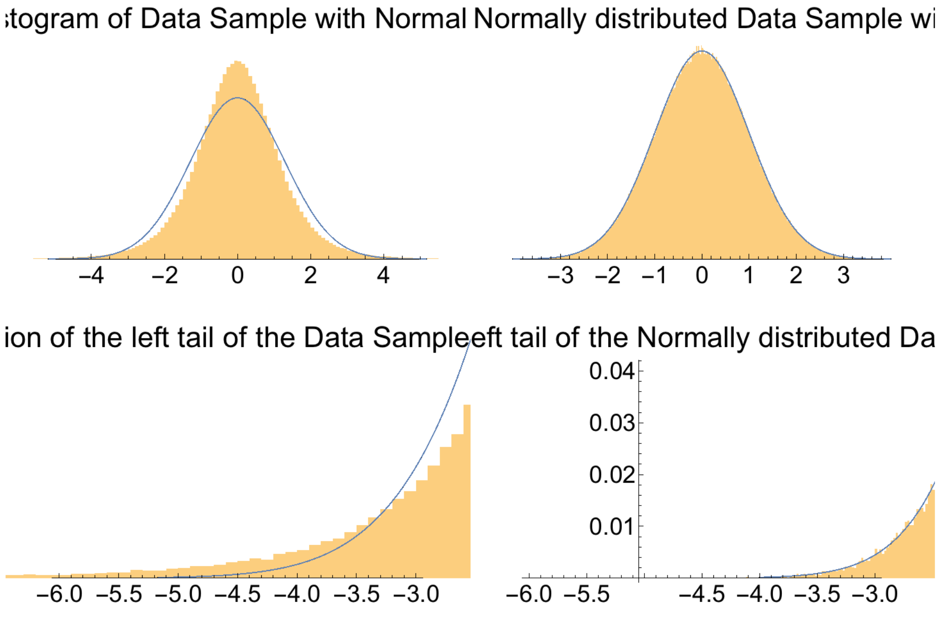


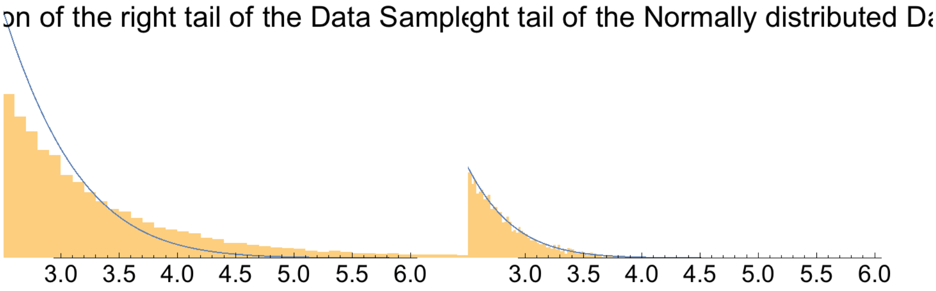
In order to calculate for example the $(1-\alpha)\ \%$ Value-at-Risk, we need to be able to capture the distribiton of the stock market returns. As we want to check how likely relatively extreme returns are, we first try to fit the empirical distribution with a normal distribution. In the figure below, we have on the left side:

- **First figure: A histogram with a normal fit**
- **Second figure: A zoomed in version of the histogram on the left tail of the distribution**
- **Third figure: A zoomed in version of the histogram on the right tail of the distribution**

In order to compare the histogram with an empirical distribution that is certainly normally distributed, we generated artificial normally distributed returns. Therefore in the figure below, we have on the right side:

- **First figure: A histogram with a normal fit**
- **Second figure: A zoomed in version of the histogram on the left tail of the distribution**
- **Third figure: A zoomed in version of the histogram on the right tail of the distribution**





We can clearly see in the figures above on the left that the tails of our stock market returns are much fatter than the figures above on the right. Therefore it is unlikely that the stock market returns are normally distributed.

Q-Q PLOTS

INTRODUCTION

We saw in the application "identifying fat tails" that the stock market returns are clearly not normally distributed, because the distribution of the returns has fat tails. We want to investigate which distribution the returns follow. One graphic tool to do this is the so called Q-Q plot.

Q-Q PLOT

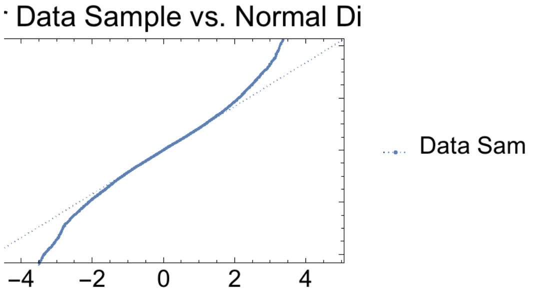
Q-Q Plot
In statistics, a Q-Q plot (abbreviated from a Quantile-Quantile plot) is a probability plot and is a graphical method for comparing two probability distributions by plotting their quantiles against each other. If the two distributions are similar, then all the points in the Q-Q plot will lie roughly on the line $y = x$ (45° line).

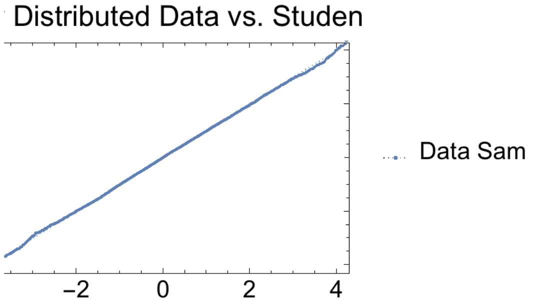
APPLICATIONS II

STOCK MARKET RETURNS: IDENTIFYING FAT TAILS II

Next we try to see if the stock market returns follow a Student's $t\ (5)$-distribution, meaning Student's t-distribution with 5 degrees of freedom. This distribution allows for much fatter tails. In order to check if the stock market returns indeed follow a Student's $t\ (5)$-distribution, we create a Q-Q plot.

In the top figure below, the Q-Q plot is shown for the stock market returns vs. a Student's $t\,(5)$-distribution. In the bottom figure below, the Q-Q plot is shown for the stock market returns vs. a normal distribution.





The Q-Q plot strongly indicates that the stock market returns of stock XYZ indeed follow a Student's $t\,(5)$-distribution and not a normal distribution. The Student's $t\,(5)$-distribution is clearly capable of capturing the fat tails that are present in the stock market returns.

TAKE AWAY

NORMAL DISTRIBUTION

The normal distribution is a continuous probability distribution that is centered around the mean, bell-shaped, symmetric and is completely determined by two input paramters: the mean μ and the variance σ^2.

Notation: A normally distributed random variable X with mean μ and variance σ^2 is denoted by $X \sim N\left(\mu,\ \sigma^2\right)$.

STANDARD NORMAL DISTRIBUTION

The standard normal distribution is a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. Any normally distributed random variable with mean μ and variance σ^2 can be rewritten as a standard normal random variable Z in the following way:

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

STUDENT’S T-DISTRIBUTION

In probability and statistics, the Student’s t-distribution is a symmetric and bell-shaped continuous probability distribution. The Student’*s* t-distribution arises when estimating the mean of a normally distributed population in situations where the sample size is small and the population standard deviation is unknown. A Student’s t-distribution is characterized by one parameter ν, which is called the degrees of freedom.

NORMAL DISTRIBUTION VS. STUDENT’S T-DISTRIBUTION

A normal distribution describes a full population with known parameters μ and σ. When only a sample of the full population is drawn, μ and σ have to be estimated. The Student’s t-distributions describe samples drawn from a full population; accordingly, the Student’s *t*-distribution for each sample size is different, and the larger the sample, the more the distribution resembles a normal distribution.

DEGREES OF FREEDOM

The degrees of freedom parameter ν in a Student’s *t*-distribution is equal to $n - 1$, where n is the number of obervations in the sample. If $\nu \to \infty$ (infinity), the Student’s *t*-distribution converges to the normal distribution with its true parameters μ and σ.

QUANTILE

The p^{th} quantile (or percentile) of a probability distribution is the value x such that $P\,(X \leq x) = p$. If X is normally distributed, $X \sim N\,(\mu,\ \sigma)$, then the x value such that $P\,(X \leq x) = p$ is

$$x = InverseCDF\left[NormalDistribution[\mu,\ \sigma],\ p\right].$$