The MEAN µ:

https://www.youtube.com/watch?v=bfQLNyiDPsk&list=PLTNMv857s9WVStKLco6ZBOsfSGXzJ1L0f&index=1

$$\bar{x}: \mu = \frac{\sum X}{n}$$

The total of the samples over the number of samples of the data

Example:
10, 28, 28, 33, 54

$$\bar{x} = \mu = \frac{10+28+28+33+54}{5} = \frac{153}{5} = 30.6$$

Weighted Mean:
Use the unique values in a table counting the frequency, so we weighted each of the number with the frequency that occurs

X	10	28	33	54
F(X)	1	2	2	1

$$\bar{x}: \mu = \frac{\sum XF(X)}{\sum F(X)} = \frac{10(1) + 28(2) + 33(1) + 54(1)}{1 + 2 + 1 + 1}$$

Example:
Find the mean of the categorical dataset: [Male, Female, Female, Female, Male, Female]
Equivalent to: [0, 1, 1, 1, 0, 1]
Mean = (0+1+1+1+0+1)/6 = 0.666

The MEDIAN: (used to get the average if data set contains big difference between each other)

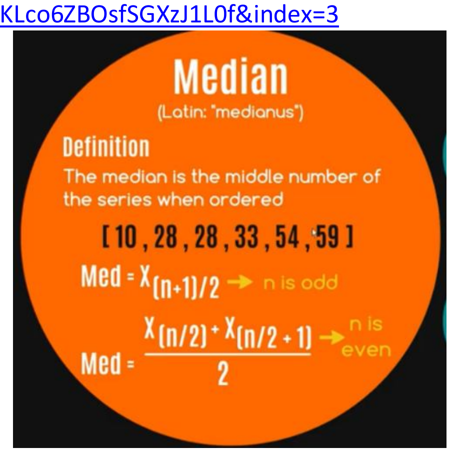
https://www.youtube.com/watch?v=rvBqEEGtJY4&list=PLTNMv857s9WVStKLco6ZBOsfSGXzJ1L0f&index=3

The middle number of the series when ordered, the center of the data, so if odd it will be the middle number, if even it will be the average of the 2 middle numbers.
Example:
10, 28, 28, 33, 54

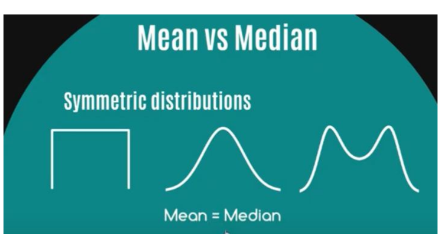
M or MED = 28 , mean as calculated before for the same is 30.6

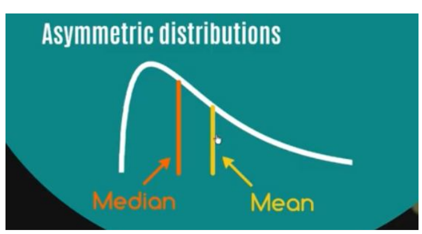
Example:
10, 28, 28, 33, 54, 59

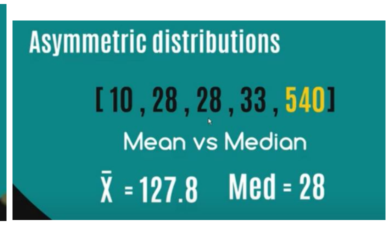
MED = 28+33/2=



Mean VS Median







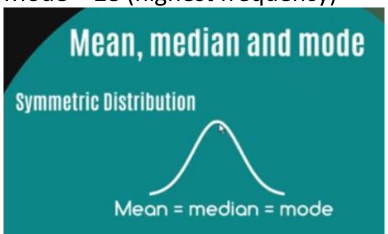
The MODE:

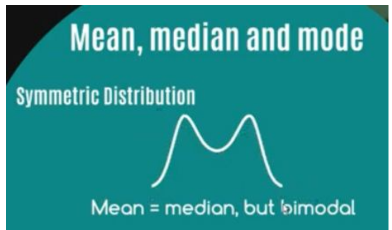
The observation of the highest frequency, more suitable for large data set

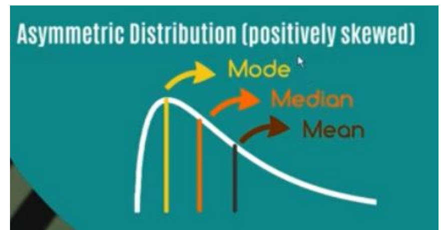
Example:

10, 28, 28, 33, 54

Mode = 28 (highest frequency)







Example:

Find Mean, Median, mode, and range for 10, 7, 14, 23, 15, 7, 32

Solution: first arrange the numbers in orders >> 7, 7, 10, 14, 15, 23, 32

Mean = sum/n = 7+7+10+14+15+23+32/7 = 108/7 = 15.4

Median = 14

Min = 7, max = 32,

range = 32-7 = 25

Mode = 7

Example:

Find Mean, Median, mode, and range for 15,21,59,15,37,59,11,41

Solution: first arrange the numbers in orders: 11, 15, 15, 21, 37, 41, 59, 59

Mean = sum/n = 32.25

Median = 21+37/2 = 58/2 = 29

Mode = 15, 59 (the most frequent number)

Range = 59 – 11 = 48

The Quantiles:

The median:

The observation that is halfway through the ordered dataset

How do you describe the location of the dataset (Max, Min, Mean, Median, Mode)

Percentile: (0.01)

split the dataset in to hundreds (100)

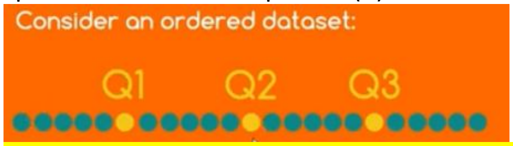
The **first percentile** is the observation that is **one percent** of the way through the ordered dataset

Decile: (0.10)

split the dataset in to tens (10)

The **first decile** is the observation that is **one tenth** of the way through the ordered dataset

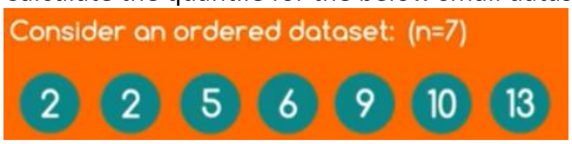
Quartile: (0.25)
split the dataset into quarter (4)



The **first Quartile** is the observation that is **one Quarter** of the way through the ordered dataset

Example:

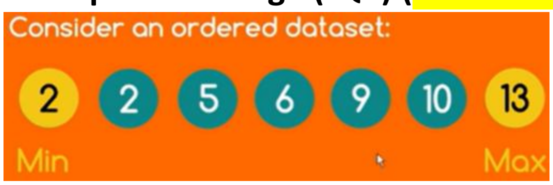
Calculate the quantile for the below small dataset (quartile, decile, percentile)

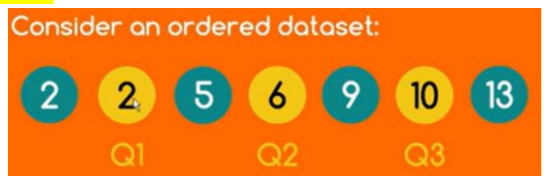


Find the five number summary (Min, Q1, Q2, Q3, Max) for the series above.
Solution:
Min = 2, Q1 = the median of the left side from 6 = 2
Med = 6, Q3 = the median of the right side from 6 = 10 , Max= 13
In R:

```
dt <- c(-3,5,7,9,20)
quantile(dt,c(0.25,0.5,0.75))
```

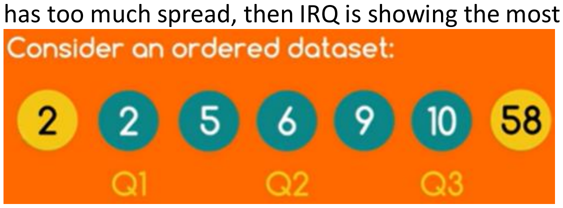
Interquartile Range (IQR) (50 % of the data set)



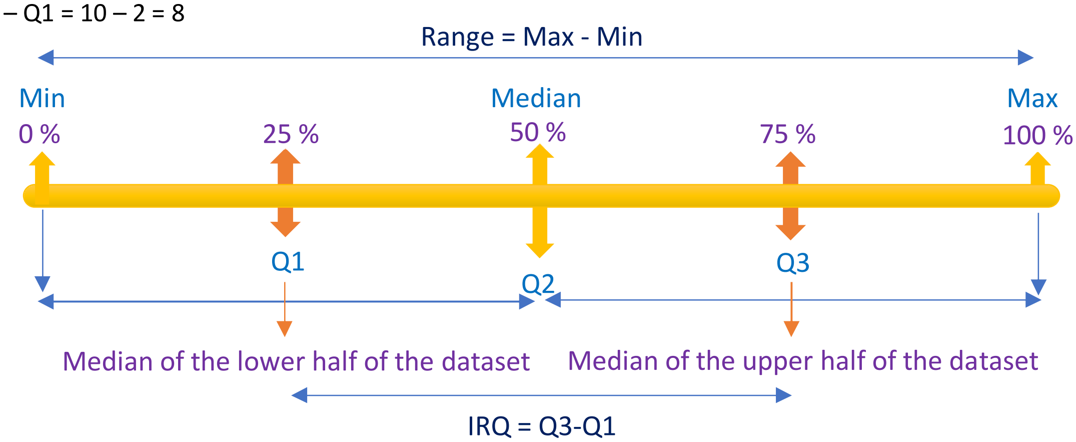


Range = max – min = 13 – 2 = 11 **IRQ** = Q3 – Q1 = 10 – 2 = 8

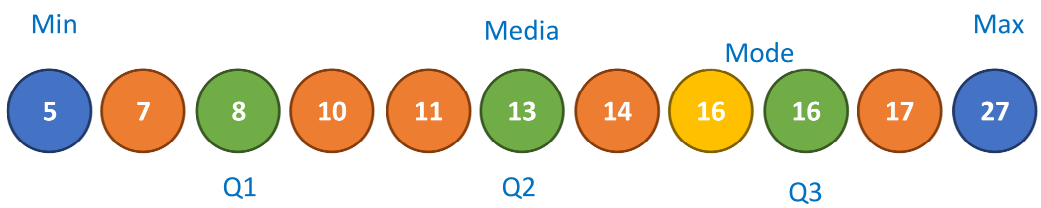
Suppose max value is too far from the range of the numbers then the Range value will be too high as this data set has too much spread, then IRQ is showing the most common value



Range = max – min = 58 – 2 = 56
IRQ = Q3 – Q1 = 10 – 2 = 8

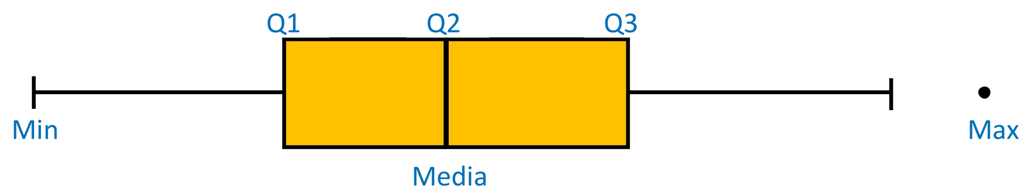


Example:
The numbers: 7,11,14,5,8,27,16,10,13,17,16 (find Q1, Q2, Q3, Median, Min, Max)
Solution: rearrange the numbers: 5,7,8,10,11,13,14,16,16,17,27

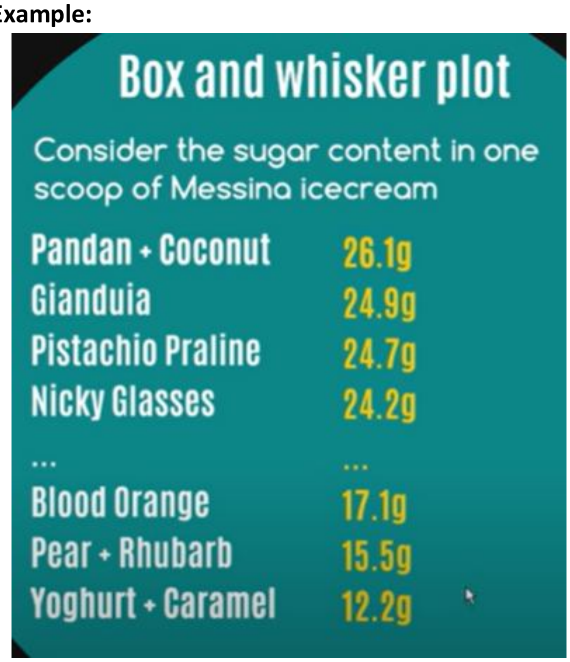


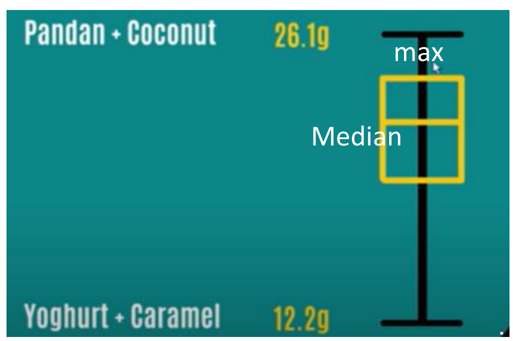
Median = Q2 = 13 (to find Q1, Q3, will not count Q2 which split the dataset to 2 equal set of 5 numbers)
Q1: the median of the lower half of the dataset (5,7,**8**,10,11) = 8
Q3: the median of the upper half of the dataset (14,16,**16**,17,27) = 16
IQR = Q3 – Q1 = 16 – 8 = 8

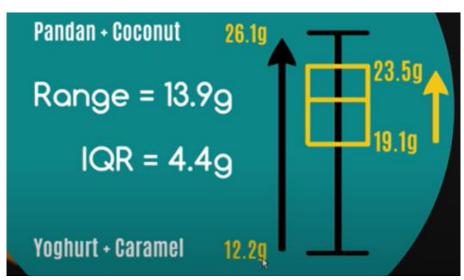
Box and whisker plot



Example:







IQR

ECDF : Empirical Cumulative Distribution Function (ECDF)

1- Arrange our data set
2- A: {15,16,16,16,16,16,16,17,17,18}
3- we have 10 records (student grads) 1/10 <= 15, 7/10 <= 16, 9/10 <=17, 10/10 <= 18.

i	x_i	n_i	f_i	$\sum_{j=1}^{i} f_i$
1	15	1	0.1	0.1
2	16	6	0.6	0.7
3	17	2	0.2	0.9
4	18	1	0.1	1.0

$$F_n(t) = \frac{\sum_{i=1}^{n} I_{x_i \le t}}{n}$$

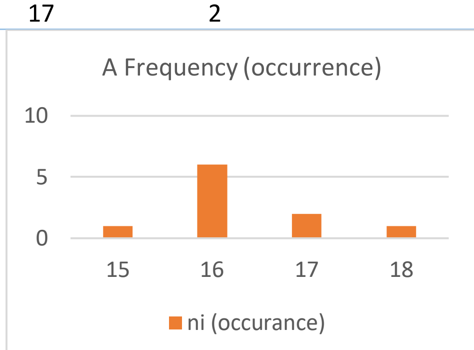
n_i : the number of occurrence ,
f_i : n_i/n the number of occurrence / the total number
the sum of f_i the total > f2 = f2 + f1 , f3 = f3+f2+f1

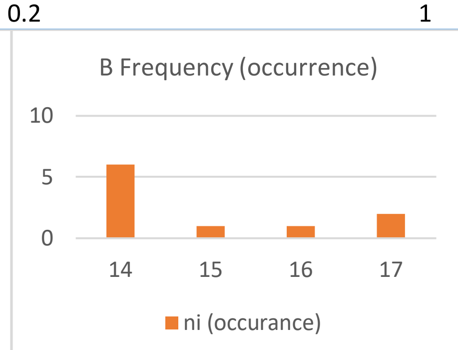
A

i	grades	n_i (occurrence)	f_i (frequency) n_i/total n	total frequency (f_i+f_i-1+f_i-2+…+f1)
1	15	1	0.1	0.1
2	16	6	0.6	0.7
3	17	2	0.2	0.9
4	18	1	0.1	1

B

i	grades	ni (occurrence)	fi (frequency) ni/total n	total frequency (fi+fi-1+fi-2+…+f1)
1	14	6	0.6	0.6
2	15	1	0.1	0.7
3	16	1	0.1	0.8
4	17	2	0.2	1





[Introduction to Statistics - YouTube]

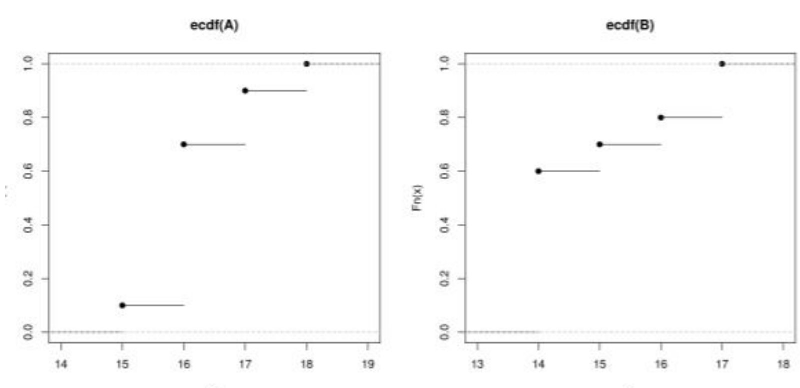


Figure: ECDF of the student's grades

ECDF (t) : fraction of values which are <= t

ECDF properties:

- $0 \leq F_n(t) \leq 1, \quad t \in \mathcal{R}$
- non decreasing
- right continuous
- $F_n(-\infty) = 0,\ F_n(+\infty) = 1$

Example real lab: case study – stat genetics

Suppose we have

n = 80K individuals with

k = 1 million LOCI Genetic code for each

this data missing some values as Null

i	G01	G02	G03	G04	..	G 1 million
1	A/T					
2			NULL			
80 K	. . .			NULL		G/G

We need to evaluate the data and find over all missing rate

So missing rate by LOCI, will rotate the table, and calculate the missing rate per individual loci

i	1	2	3	4	80 k	Missing Rate MR for each LOCI
G01	A/T					0 (no missing)
G02						0.01
		NULL				0.50
G 1 million	. . .				G/G	

Apply in R:

First install library as below:

```
library(tidyverse)
MISSING_RATE_CUTOFF <- 0.01
```

Then import the data

```
Gdata <- read_tsv("C:/my/R_training/genotyping_missing_rates.tsv",
         col_types = cols(chromosome = 'c', SNP_ID = 'c', missing_rate='d')) %>%
 mutate(chromosome = factor(chromosome))
```

To check the summary of the data

summary(Gdata)

```
> summary(Gdata)
   chromosome        SNP_ID           missing_rate
 2      : 55544   Length:670176      Min.   :0.0000321
 6      : 51207   Class :character   1st Qu.:0.0013960
 1      : 50279   Mode  :character   Median :0.0021500
 3      : 48779                      Mean   :0.0062765
 4      : 45430                      3rd Qu.:0.0041880
 5      : 42621                      Max.   :0.4054000
 (Other):376316
```

We can see average of missing data (mean 0.006) and max missing data (0.405)

To see the structure of the data set: str(Gdata)

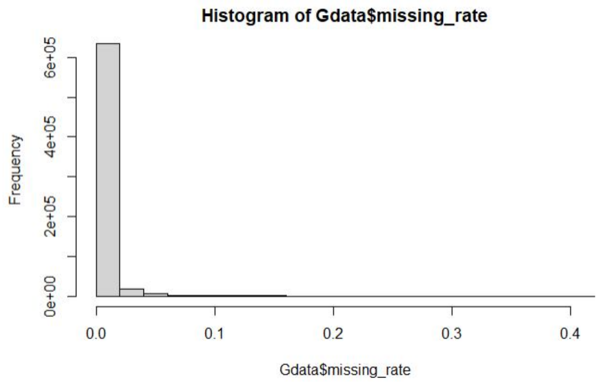
````
```{r}
str(Gdata)
```

spec_tbl_df [670,176 x 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ chromosome  : Factor w/ 24 levels "1","10","11",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ SNP_ID      : chr [1:670176] "SNP1" "SNP2" "SNP3" "SNP4" ...
 $ missing_rate: num [1:670176] 0.00252 0.00687 0.01248 0.00366 0.00204 ...
 - attr(*, "spec")=
  .. cols(
  ..   chromosome = col_character(),
  ..   SNP_ID = col_character(),
  ..   missing_rate = col_double()
  .. )
 - attr(*, "problems")=<externalptr>
````

It tells us that, there is 3 variables or columns X 670,176 record or rows

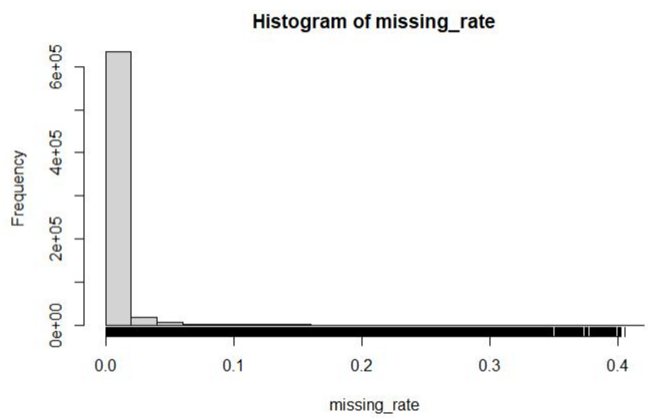
To draw Histogram the missing rates:

hist(Gdata$missing_rate)



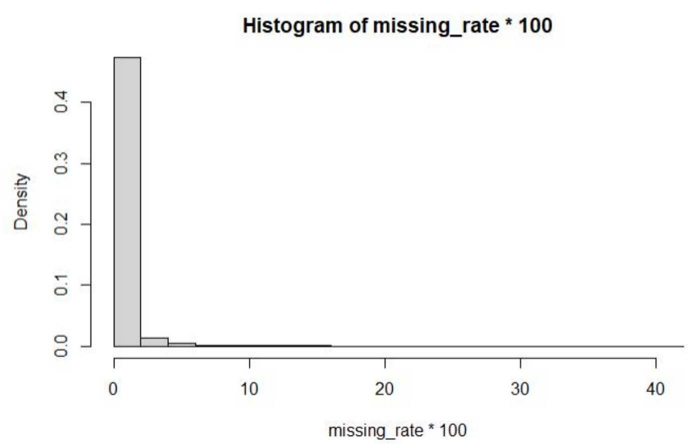
We can see that distributed between 0 and 0.4, and data heavily concentrated around the small values. And we can see the missing rate at maximum is more than 40% and the histogram here is completely flat but not zero, there is data, so we can use rug which adding small vertical fragment for each single observation

```
with(Gdata, { hist(missing_rate)  rug(missing_rate) })
```



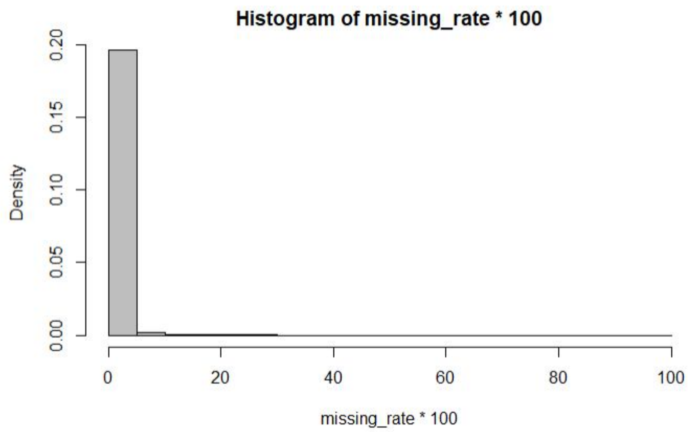
Change the scale with specifying to use density instead of frequency by enabling prob:

```
with(Gdata, hist(missing_rate * 100.0, prob = TRUE)
```



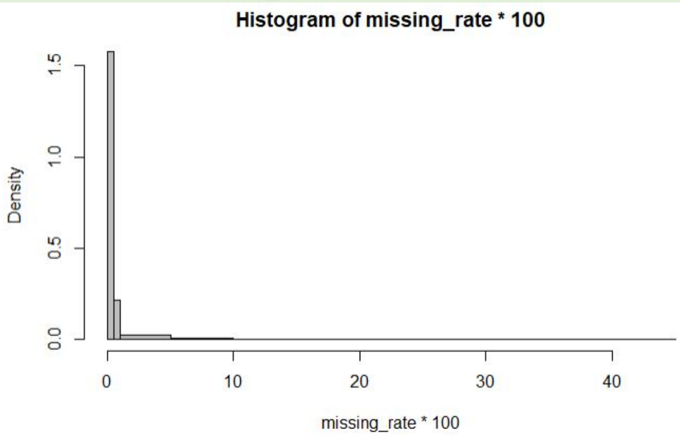
By break using breaks to be from 0 to 100 steps by 5 (for the density):, and colors to be grey

```
with(Gdata, hist(missing_rate * 100.0, prob = TRUE, breaks = seq(0, 100, by = 5), col="grey"))
```



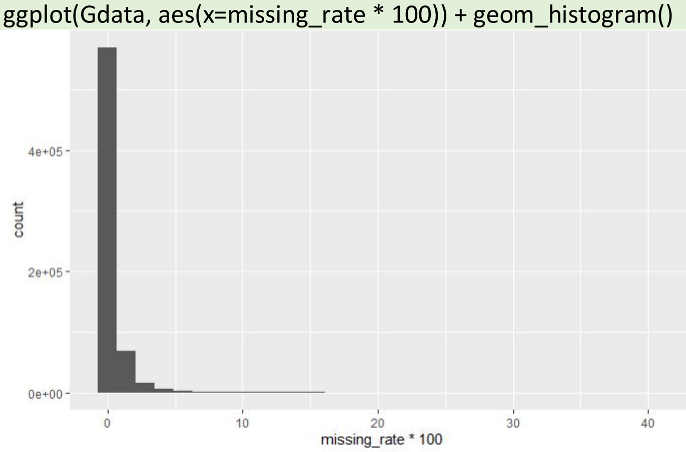
I know that the data concentrated near to the 0, so I can make a break to show the first part near to 0 using smaller scale from 0 to 0.5 and from 0.5 to 1 and then from 5 to 45 step 5 (for the density):

```
with(Gdata, hist(missing_rate * 100.0, prob = TRUE, breaks = c(0,0.5,1, seq(5, 45, by = 5)), col="grey"))
```



Using better plotting (ggplot) and not important for the vertical access information

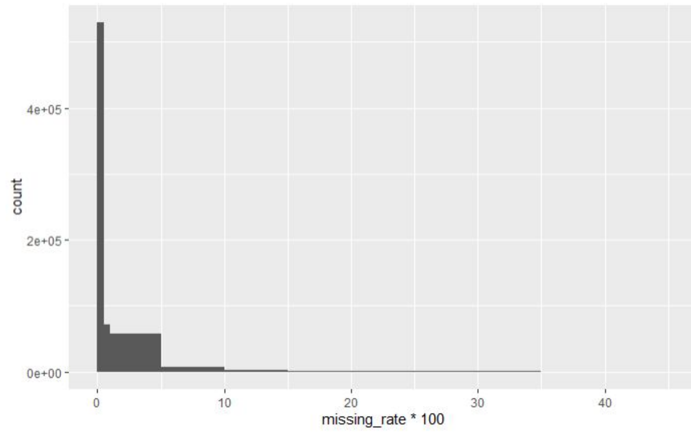
```
ggplot(Gdata, aes(x=missing_rate * 100)) + geom_histogram()
```



To apply same breaks we applied before

```
ggplot(Gdata, aes(x=missing_rate * 100)) + geom_histogram(breaks = c(0,0.5,1, seq(5, 45, by = 5)))
```

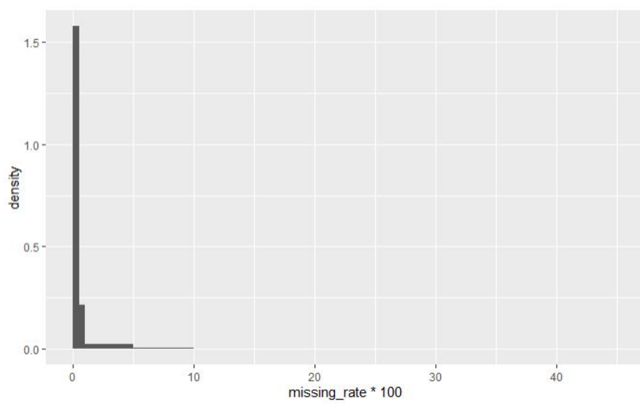
It looks wrong not like the previous histograme



The count of missing rates between .5% and 1% is similar to that between 1% and 5% but the interval width is very different!

Here is better visual representation using **density for vertical axis:**

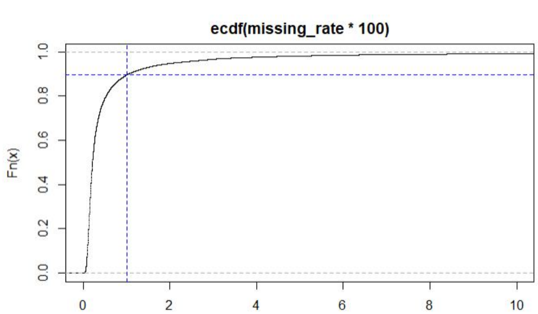
```
ggplot(Gdata, aes(x=missing_rate * 100,y=..density..)) + geom_histogram(breaks = c(0,0.5,1, seq(5, 45, by = 5)))
```



i	class$_i$	n_i	f_i	h_i
1	(0, 50]	29	0.60	0.6/50
2	(50, 100]	6	0.12	0.12/50
3	(100, 1000]	6	0.12	0.12/900
4	(1000, 10000]	5	0.10	0.10/9000
5	(10000, 20000]	2	0.04	0.04/10000

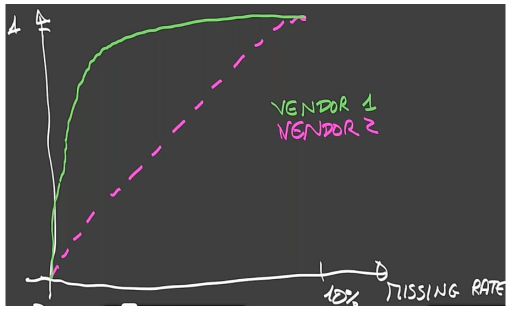
To draw the ECDF

```
with(Gdata, plot(ecdf(missing_rate * 100), xlim=c(0,10)))
abline(v = MISSING_RATE_CUTOFF * 100, lty=2,col="blue")
abline(h=mean(Gdata$missing_rate <= MISSING_RATE_CUTOFF), lty=2, col="blue")
```



From this drawing I can note that 90% of the LOCI have a missing rate of less than 1%

Example ECDF for Comparisons between 2 vendors



Number 1 have low missing rate which make it the best choice.

PMF Probability Mass Function **& CDF** Cumulative Distribution Function

https://www.youtube.com/watch?v=3xAIWiTJCvE
https://www.youtube.com/watch?v=YXLVjCKVP7U
http://www.zstatistics.com/videos

PMF Probability Mass Function:

tells us the probability that a discrete random variable takes on a certain value.

Example:

Suppose we roll a dice one time. If we let x denote the number that the dice lands on, then the probability that the x is equal to different values can be described as follows:

P(X=1): 1/6
P(X=2): 1/6
P(X=3): 1/6
P(X=4): 1/6
P(X=5): 1/6
P(X=6): 1/6

$$p_X(x) = \begin{cases} 1/6 & \text{if } x = 1 \\ 1/6 & \text{if } x = 2 \\ 1/6 & \text{if } x = 3 \\ 1/6 & \text{if } x = 4 \\ 1/6 & \text{if } x = 5 \\ 1/6 & \text{if } x = 6 \\ 0 & \text{otherwise} \end{cases}$$

There is an equal chance that the dice could land on any number between 1 and 6.

One characteristic of a probability mass function is that all of the probabilities must add up to 1. You'll notice that this PMF satisfies that condition:

Sum of probabilities = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1.

Example:

I toss a coin twice. Let X be the number of observed heads. Find the CDF of X.

Solution:

We care only about the Head, so by toss a coin the twice, probability of showing the head as below:

1- HH : 2 heads
2- HT : 1 head
3- TH: 1 head
4- TT: 0 head

So number of probabilities 4 and the outcome (0 or 1 or 2)
Px(0) = P(X=0) = 1/4 (1 time in the 4 possibilities)
Px(1) = P(X=1) = 2/4 =1/2 (2 times in the 4 possibilities)
Px(2) = P(X=2) = 1/4 (1 time in the 4 possibilities)

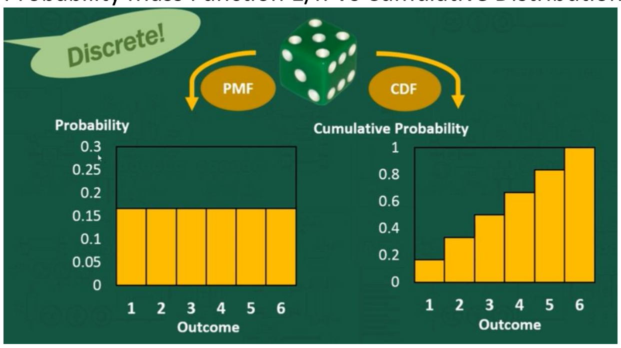
PMF (Probability Mass function) CDF

H	P() probability number of occurrence /n for each	CDF	CDF
0	¼ = 0.25	¼ = 0.25	p(1)
1	½ = 0.5	¾= 0.75	p(1)+p(2)
2	¼ = 0.25	1	P(1)+P(2)+P(3)

Cumulative Distribution Function (CDF):
PMF is one way to describe the distribution of a discrete random variable, PMF cannot be defined for continuous random variables.
(**CDF**) of a random variable is another method to describe the distribution of random variables. The advantage of the CDF is that it can be defined for any kind of random variable (discrete, continuous, and mixed).

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{4} & \text{for } 0 \le x < 1 \\ \frac{3}{4} & \text{for } 1 \le x < 2 \\ 1 & \text{for } x \ge 2 \end{cases}$$

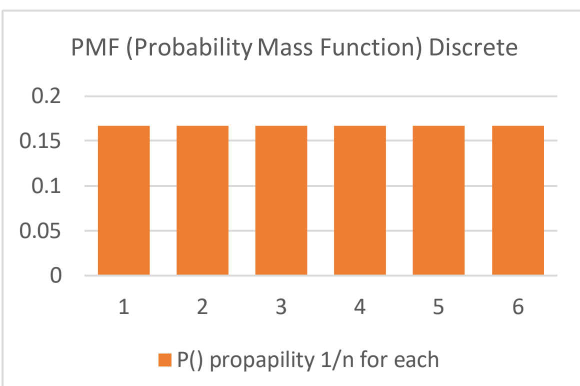
Probability Mass Function 1/n Vs Cumulative Distribution Function

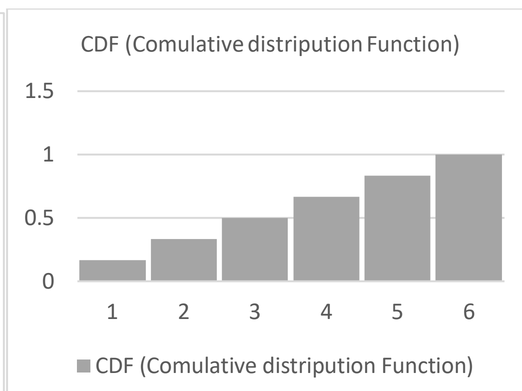


So, in CDF final bar need to be 1

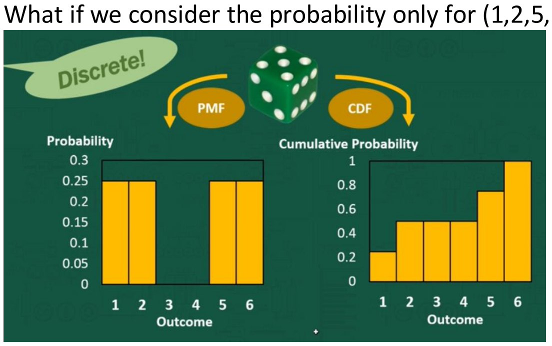
PMF (Probability Mass function) CDF (Comulative Distripution Function)

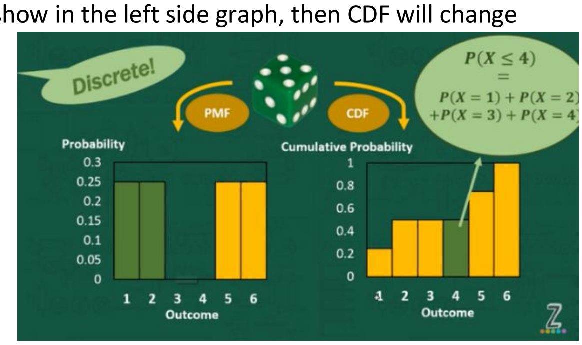
dice	P() propapility 1/n for each	CDF (Comulative distripution Function)	CDF
1	0.166666667	0.166666667	p(1)
2	0.166666667	0.333333333	p(1)+p(2)
3	0.166666667	0.5	P(1)+P(2)+P(3)
4	0.166666667	0.666666667	P(1)+P(2)+P(3)+P(4)
5	0.166666667	0.833333333	P(1)+P(2)+P(3)+P(4)+P(5)
6	0.166666667	1	P(1)+P(2)+P(3)+P(4)+P(5)+P(6)





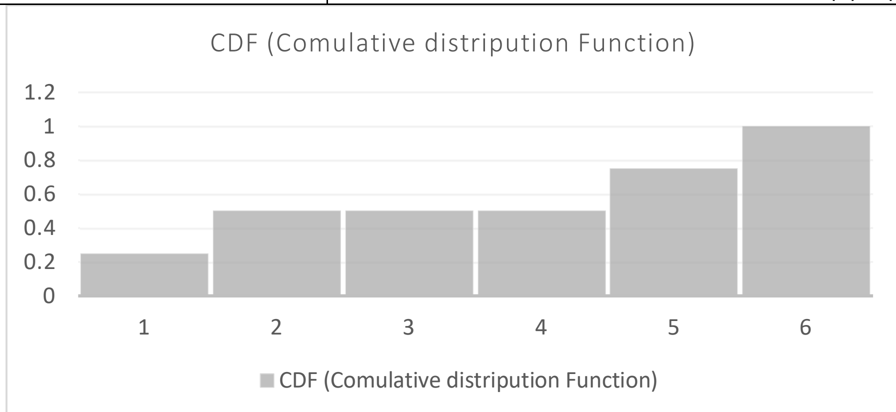
What if we consider the probability only for (1,2,5,6) as show in the left side graph, then CDF will change

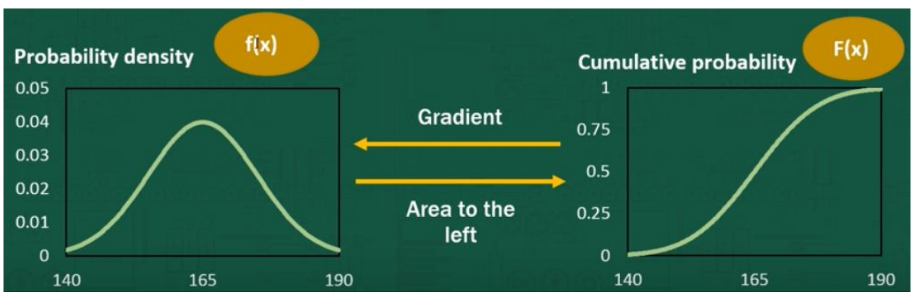




PMF (Probability mass function) CDF (Comulative Distripution Function)

dice	P() propapility 1/4 for each	CDF (Comulative distripution Function)	CDF
1	0.25	0.25	p(1)
2	0.25	0.5	p(1)+p(2)
3	0	0.5	
4	0	0.5	
5	0.25	0.75	P(1)+P(2)+P(5)
6	0.25	1	P(1)+P(2)+P(5)+P(6)





Multivariate distribution: two components

Education and wages: consider the following data collection, for each individual we have multiple information (wages, education, age, etc.)

i	wages	education	age	sex	language
1	10.56	15.00	40	Male	English
2	11.00	13.20	19	Male	English
		...			
7424	11.85	11.00	47	Female	English
7425	23.00	14.00	30	Male	English

$$(x, y) = (x_i, y_i)$$
$$x = (x_i) \qquad i = 1, \ldots, n$$
$$y = (y_i)$$

We will use the pairs like wages & age as (xi,yi)

The contingency table: two categorical variables

By frequencies and by percentages

sex	language: English	French	Other	Total
Female	2999	262	564	3825
Male	2717	235	527	3479
Total	5716	497	1091	7304

absolute frequencies

sex	language: English	French	Other	Total
Female	78.4%	6.85%	14.7%	100%
Male	78.1%	6.75%	15.1%	100%
Total	78.3%	6.80%	14.9%	100%

row percentages

Histogram & boxplots (one numerical, one categorical)

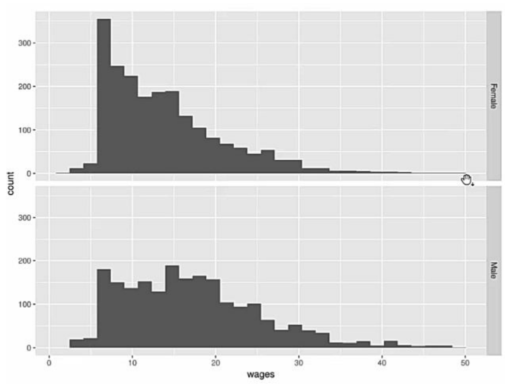


Figure: Hourly wages (horizontal axis) by sex (vertical panels)

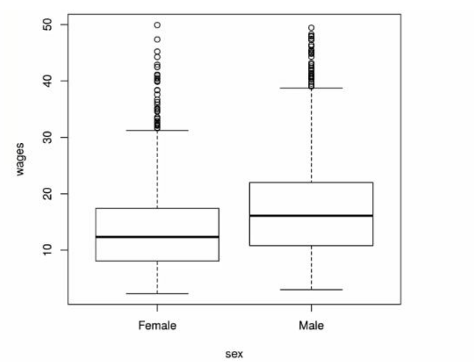


Figure: Hourly wages (vertical axis) and sex (horizontal axis)

Graphical representation: the scatterplot

two numerical components

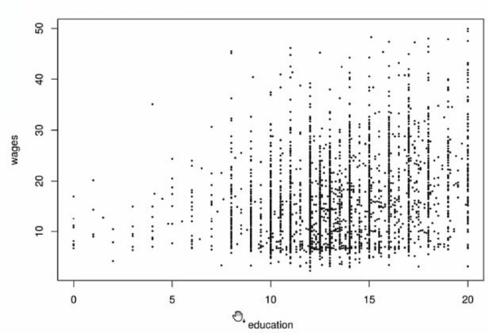


Figure: Hourly wages (vertical axis) and years of schooling (horizontal axis)

Here each dot is a person (years of education, wages)

Covariance: gives the direction if positive or negative.

MEAN OF X

MEAN OF Y

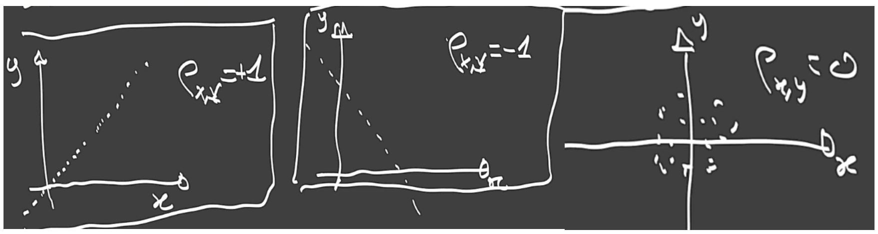
$$\sigma_{x,y} = \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \mu_x)\cdot(y_i - \mu_y)\right]$$

Example: x= years of education, y=wage

Years of education of person xi

Linear correlation coefficient: (-1 or +1)

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \cdot \sigma_y}$$



Variance and Standard Deviation

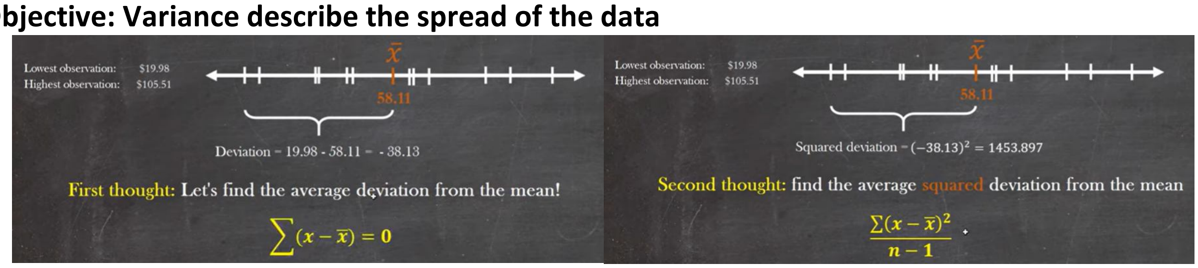
Week	Weekly expenditure on Golden Gaytimes
1	$48.50
2	$87.40
3	$19.98
4	$59.74
5	$40.87
6	$105.51
7	$40.80
8	$23.10
9	$98.10
10	$60.54
11	$64.81
12	$48.01

$$Mean = \bar{x} = \frac{\sum x}{n} = \frac{\$48.50 + \$87.40 + \cdots}{12} = \$58.11$$

$$Variance = s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{(\$48.50 - \$58.11)^2 + (\$87.40 - \$58.11)^2 \ldots}{11} = \$748.01$$

$$Std\ dev = s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\$748.01} = \$27.35$$

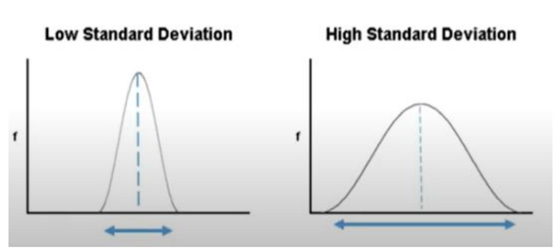
Objective: Variance describe the spread of the data

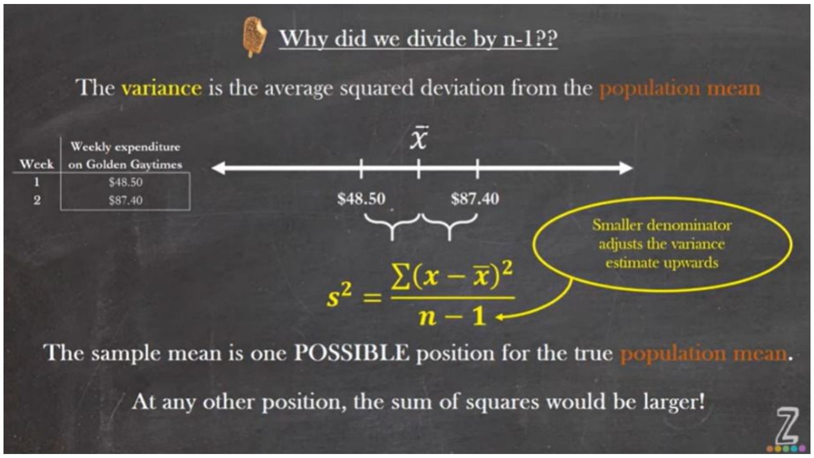


Standard deviation:

متوسط انحرافات القيم عن المتوسط

Standard deviation will be low when all dataset distributed near to the mean, and very high when all dataset distributed and spread far from the mean



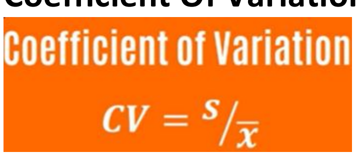


Example:

Week	Expenditure(x)	(x-mean)	(x-mean) ^2
3	19.98	-38.13333333	1454.15111
8	23.1	-35.01333333	1225.93351
7	40.8	-17.31333333	299.751511
5	40.87	-17.24333333	297.332544
12	48.01	-10.10333333	102.077344
1	48.5	-9.613333333	92.4161778
4	59.74	1.626666667	2.64604444
10	60.54	2.426666667	5.88871111
11	64.81	6.696666667	44.8453444
2	87.4	29.28666667	857.708844
9	98.1	39.98666667	1598.93351
6	105.51	47.39666667	2246.44401
		total	8228.12867

Min	19.98	Minimum
Max	105.51	Maximum
Median	54.12	average of the 2 numbers in the middle if dataset is even
Mean	58.11333333	Average
Range	85.53	Maximum - Minimum
Variance	748.011697	**Variance** = total(x-mean) ^2/n-1
std Deviation	27.3498025	**std deviation** = SQRT(Variance)

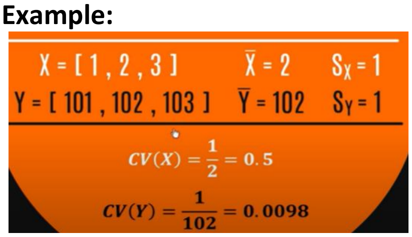
Coefficient Of Variation



Coefficient of Variation = Standard deviation / Mean

$std = \sqrt{\frac{\sum(X-mean)^2}{n-1}}$, CV(x)= std(x) /mean

Example:



$$CV(x) = \sqrt{\frac{(1-2)^2+(2-2)^2+(3-2)^2}{3-1}}/2 = \frac{1}{2}$$

$$CV(y) = \sqrt{\frac{(101-102)^2+(102-102)^2+(103-102)^2}{3-1}}/102 = \frac{1}{102}$$

Example:

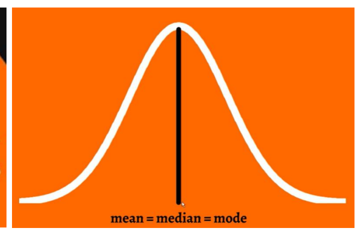
Fuel prices (per gallon) were surveyed every week for 5 weeks in the US and in Vietnam. Which country experiences the greatest fuel price fluctuations?

USA	USA - mean	(USA-Mean)^2
2.70	(2.70-2.806)= -0.11	0.0112
3.06	(3.06-2.806)= 0.25	0.0645
2.87	(2.87-2.806)= 0.06	0.0041
2.69	(2.69-2.806)= -0.12	0.0135
2.71	(2.71-2.806)= -0.10	0.0092
Coefficient of Variation	Variation = (std)^2	0.0256
	std deviation	0.1601
	CV = (std deviation)/mean	0.0571

Vietnam	Vietnam - mean	(Vietnam - mean)^2
11,612	(11,612-12,385)= -773	597,219.840
12,138	(12,138-12,385)= -247	60,910.240
12,980	(12,980-12,385)= 595	354,263.040
13,110	(13,110-12,385)= 725	525,915.040
12,084	(12,084-12,385)= -301	90,480.640
Coefficient of Variation	variation	407,197.200
	std deviation	638.120
	CV = (std deviation)/mean	0.052

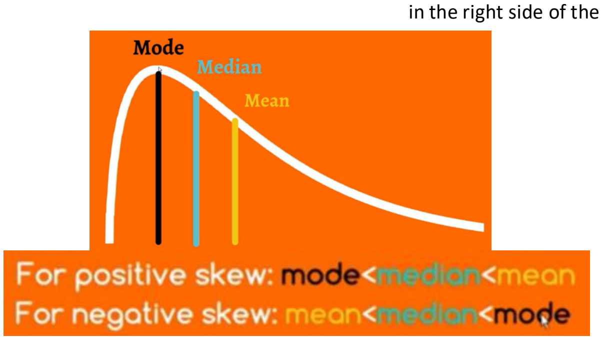
mean	2.806
Min	2.69
Max	3.06
range	0.37

mean	12,385
Min	11612
Max	13110
range	1498



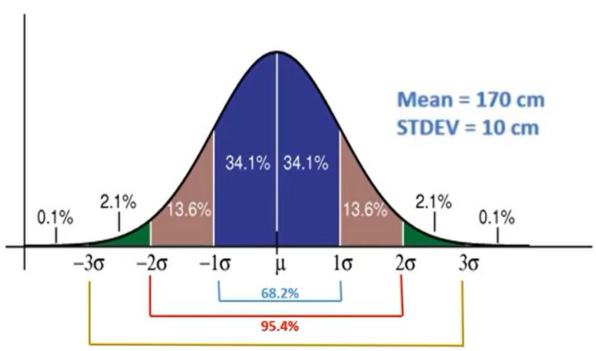
No Skew

left (negative) skew: more observations in the right side of the mode



Standard deviation law:

68.2% of the data distributed around the mean if standard deviation between 1xSTD and -1xSTD



68 % of observations lies between mean ± 1σ → 170 ± 10 cm
95 % of observations lies between mean ± 2σ → 170 ± 20 cm
99% of observations lies between mean ± 3σ → 170 ± 30 cm

If we have 1000 of observation:

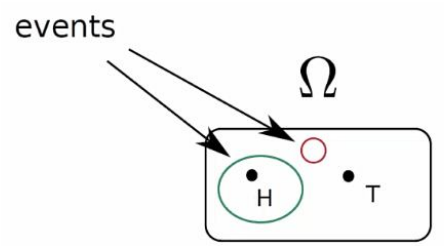
Probability:
Probability of an event is a *chance* that this event will happen.
Probability of an event is the limit of the frequency of successes, n -> ∞, in a sequence of trials all performed in the same condition.

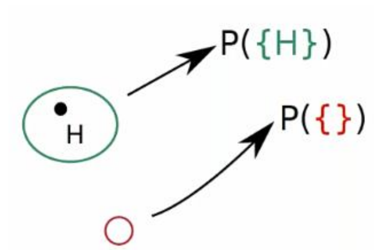
Example.
If there are 5 communication channels in service, and a channel is selected at random when a telephone call is placed, then each channel has a probability $1/5 = 0.2$ of being selected

Sample space
A collection of all elementary results, or **outcomes** of an experiment
Event
Any set of outcomes is an **event**. Thus, **events** are subsets of the sample space

Sample Space Omega, is a set of all possible outcomes of an experiment that can occur, it has the 2 elements (which is the 2 possibilities of throwing a coin Head or Tail).

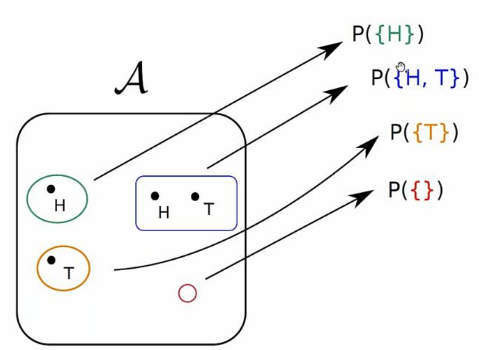
Event: is a subset of Omega as example the empty set subset of sample space, also other set which have H is a subset of Omega sample space.

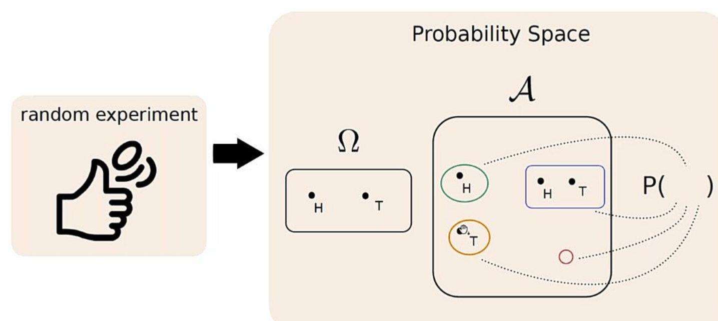




To each event we sign a probability which is a function takes the input of event and gives output a number. So, probability of event Head we can write as **P({H})**
Domain probability function: is a family of events, is a set of all possible subsets of Omega. name Sigma A of algebra, then we can calculate probability on it.

$$\sigma(\Omega) = \mathcal{A} : \sigma\text{-algebra of events}$$





If A, B in the sample space, then A U B need to be in the sample space.
If A in sample space, then A compliment need to be in the sample space.
If infinite A is in sample space an infinite Union of A in sample space.

- $A, B \in \mathcal{A} \implies A \bigcup B \in \mathcal{A}$
- $A \in \mathcal{A} \implies A^C \in \mathcal{A}$
- $A_i \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Probability is a function from sigma algebra to the numbers between 0 and 1 with some properties:
(If 0 means never happen) $0 \le P(A) \le 1$ (if 1 means always happen)

1- **For any event in Sigma algebra** is non negative but it can be zero or positive.
2- The probability of the full sample space is 1
3- For any pair of events not overlapping then the probability of the union of the 2 events is equal to the sum of the probabilities.
4- If we have an infinite sequence of events that are pairwise not overlapping then the probability of the infinite union needs to be equal to the infinite summation of the probabilities

Probability is a function $P : \mathcal{A} \mapsto [0,1]$ such that:

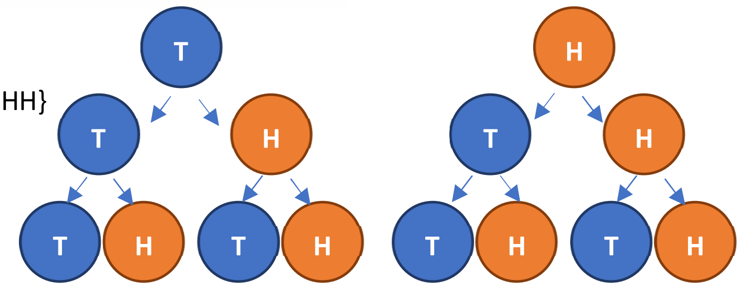
1. $\forall A \in \mathcal{A}, P(A) \geq 0$
2. $P(\Omega) = 1$
3. $A \cap B = \emptyset \implies P(A \cup B) = P(A) + P(B)$
4. $A_i \cap A_j = \emptyset \implies P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

$(\Omega, \mathcal{A}, P)$ is a probability space

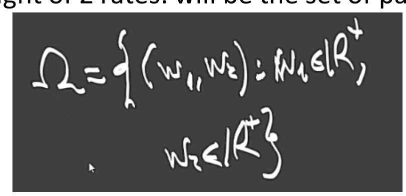
Examples:

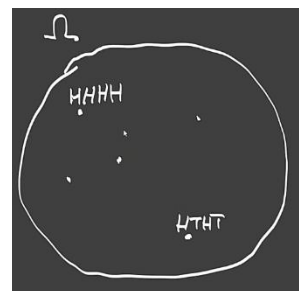
- Toss a coin 3 times >>

Ω = {TTT, TTH, THT, THH, HTT, HTH, HHT, HHH}



- Toss a coin 4 times >> sample space elements: HHHH, HTHT, HTTT, TTT, etc.
- Weight of 2 rates: will be the set of pairs such that w1, w2 is real numbers





Ω={(W1,W2) : ∀W1∈R+, W2∈R+

- Proportion of defectives in a shipment of electronic components. Ω = [0,1]

Counting techniques:

1- Multiplication (product role) in case of sequence of tasks:

The total number of ways to complete the operation is

$$n_1 \times n_2 \times \cdots \times n_k$$

Example1:

How many sample points are there in the sample space when a pair of dice is thrown once?

Solution:

dice can land in $n_1 n_2 = (6)(6) = 36$ possible ways.

Example3:

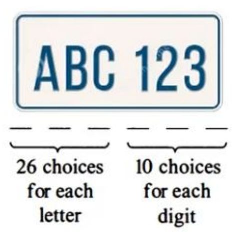
In how many different ways can a true-false test consisting of 10 questions be answered?

Solution: Each of the 10 questions can be chosen in two ways, because each question is either true or false. Therefore, the product rule shows there are:

$2 \times 2 \times \cdots \times 2 = 2^{10} = 1024$ ways to answer the test.

Example6:

How many different license plates are available if each plate contains a sequence of *three letters* followed by *three digits*.



Solution:

There are 26 choices for each of the three letters and ten choices for each of the three digits. Hence, by the product rule there are a total of $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 = 17{,}576{,}000$ possible license plates.

Permutations (1/3):

Another useful calculation finds the number of ordered sequences of the elements of a set. Consider a set of elements, such as $S = \{a, b, c\}$.

A permutation of the elements is an ordered sequence of the elements. For example, abc, acb, bac, bca, cab, and cba are all of the permutations of the elements of S.

$$3 \times 2 \times 1 = 6$$

The number of **permutations** of n different elements is $n!$ where

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

Ex. The number of permutations of the four letters a, b, c, and d will be $4! = 24$.

Permutations (3/3):

In some situations, we are interested in the number of arrangements of only some of the elements of a set.

The number of permutations of subsets of $\boldsymbol{r}$ elements selected from a set of $\boldsymbol{n}$ different elements is

$$P_r^n = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

$$_nP_r = \frac{n!}{(n-r)!}$$

Example2:

In one year, three awards (research, teaching, and service) will be given to a class of 25 graduate students in a statistics department. If each student can receive at most one award, how many possible selections are there?

<u>**Solution:**</u> $\boldsymbol{r = 3},\ \ \boldsymbol{n = 25}$

$$P_r^n = {}_nP_r = \frac{n!}{(n-r)!} = \frac{25!}{(22)!}$$

$$= \frac{25 \times 24 \times \cdots \times 3 \times 2 \times 1}{22 \times 21 \times \cdots \times 2 \times 1} = 13,800$$

Permutations of Similar Objects:

The number of permutations of $n = n_1 + n_2 + \cdots + n_r$ objects of which n_1 are of one type, n_2 are of a second type, … , and n_r are of an rth type is

$$\frac{n!}{n_1!n_2!n_3!\ldots n_r!}$$

Example3:

In a Statistics class, the teacher needs to have 20 students standing in a row. Among these 20 students, there are 12 boy, and 8 girl. How many different ways can they be arranged in a row if only their class level will be distinguished?

<u>**Solution:**</u> $\boldsymbol{n = 20},\ \ \boldsymbol{n_1 = 12},\ \ \boldsymbol{n_2 = 8}$

$$= \frac{n!}{n_1!\,n_2!} = \frac{20!}{12!\,8!} = 125,970$$

Combinations (1/2):

Another counting problem of interest is the number of subsets of r elements that can be selected from a set of n elements. Here, order is not important. These are called *combinations*.

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example1:

How many possible selections of 3 balls from a box contains 10 colored balls?

Solution: $\boldsymbol{n} = \mathbf{10}$, $\boldsymbol{r} = \mathbf{3}$

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} = \frac{10!}{3!\,7!} = 120$$

Example3: even = 2 $\boxed{\text{odd}}$

A dice is loaded in such a way that an even number is twice as likely to occur as an odd number. If E is the event that a number less than 4 occurs on a single toss of the dice, find $P(E)$.

We assign a probability of w to each odd number and a probability of $2w$ to each even number. Since the sum of the probabilities must be 1, we have $9w = 1$ or $w = 1/9$.

Solution:

w + 2w + w + 2w + w + 2w =

$S = \{1, 2, 3, 4, 5, 6\}$,

$E = \{1, 2, 3\}$

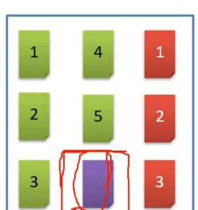
$$P(E) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$$

Example4:

(a) the dictionary is selected?

If 3 books are picked at random from a box containing 5 novels, 3 books of mathematics, and a dictionary, what is the probability that

Solution:



#of members of $S = \binom{9}{3} = \frac{9!}{3!\,6!} = 84$,

#of members of $E = \boxed{\binom{1}{1}}\binom{8}{2}$

n r select

$$\#\text{of members of } E = \binom{1}{1}\binom{8}{2}$$

$$= 1 * \frac{8!}{2!\ 6!} = 28$$

The probability that the dictionary is selected

$$= \frac{28}{84} = \frac{1}{3} = 0.333$$

(b) 2 novels and 1 book of mathematics are selected?

$$\#\text{of members of } E = \binom{5}{2}\binom{3}{1}$$

$$= \frac{5!}{2!\,3!} * \frac{3!}{1!\,2!} = 10 * 3 = 30$$

The probability that 2 novels and 1 book of mathematics are selected

$$= \frac{30}{84} = \frac{5}{14} = 0.357$$

Recap:

Counting techniques:

1- Multiplication (product role) in case of sequence of tasks:

The total number of ways to complete the operation is

$$n_1 \times n_2 \times \cdots \times n_k$$

2- Permutation of elements n and care about order is **n!**

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$$

3- Permutation of subset of r elements selected from a set of n different elements is:

$$P_r^n = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}$$

4- Permutation of a group of similar groups (a set of n as 2 groups of boys n1 and girls n2)

The number of permutations of $n = n_1 + n_2 + \cdots + n_r$ objects of which n_1 are of one type, n_2 are of a second type, … , and n_r are of an rth type is

$$\frac{n!}{n_1!n_2!n_3!\ldots n_r!}$$

5- Combination of groups not care about order like select 3 students from a class of 20 students.

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Example:

(a) If two fair coins are flipped, what is the probability of getting at least one head? (b) if three coins are flipped what is the probability of getting at least two tails? (c) if three coins are flipped, what is the possibility of getting exactly one tails?

Solution:

(A) Probability of getting at least one head
Sample space: Ω = {HH, HT, TH, TT} (4 elements)
A = {HH, HT, TH} (3 elements which have the H)
P(A) = ¾ = 0.75 = 75%

(B) if three coins are flipped what is the probability of getting at least two tails?
Sample Space: Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} (8 elements)
B = {HTT, THT, TTH, TTT} (4 elements)
P(B) = 4/8 = 4.1/4.2 = ½

(C) if three coins are flipped, what is the possibility of getting exactly one tails?
Sample Space: Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT} (8 elements)
C = {HHT, HTH, THH}
P(C) = 3/8 = 0.375 = 37.5%

Example:

Tossing a fair coin: Ω = {H, T} and all $\mathcal{A} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ possible events

By fair, we mean P(H)=P(T) >> probability of H = probability of T
We completely defined the probability space (Ω, A, P)!

$P(\Omega) = P(\{H\} \cup \{T\}) = P(\{H\}) + P(\{T\}) = 1$
$P(H) = P(T) = 1/2$

Note how we explicitly covered all the events in A:

$P(\emptyset) = 0$
$P(\{H\}) = 1/2$, $P(\{T\}) = 1/2$
$P(\Omega) = 1$

Example:

Of 2 dice then Ω= {all the possible likely items like [(.,.), (.,..),all the 36] and Calligraphic A will be the events all possibility we want to measure the probabilities)]

Example:

Tossing a fair dice , Ω= {1,2,3,4,5,6}
A & B elements of event for instance A= {1,2}, B= {4,5}
Calligraphic **A** = {all subsets of Ω}, for instance {{}, {1,6}, {1,2,3}, etc.} so probability for each will be number of items over 6 {0/6, 2/6, 3/6, etc.}
P(A)=|A|/6 (the number of elements in the event A divided by 6)
In this case, we defined a general rule to design a probability to any set in **A** (event) is (Ω, **A**, P) a valid probability space.

1- $|A| \geq 0$ then --> $P(A) \geq 0$ $\forall$**A,** $|B| \geq 0$ then --> $P(B) \geq 0$ $\forall$**A**
2- $P(\Omega) = |\Omega|/6 = 6/6 = 1$, ($|\Omega|$ is number of elements in Omega)
3- $A \cap B = \emptyset \rightarrow |A \cup B| = |A| + |B|$:
$P(A \cup B) = |A \cup B|/6 = |A|/6 + |B|/6 = P(A) + P(B)$, P(A) is probability of A, P(B) probability of B

Theorem:

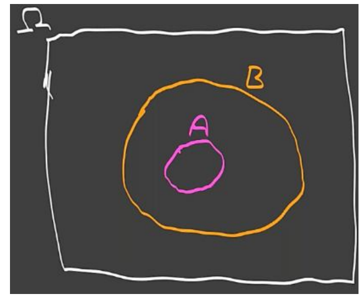
1- $P(\emptyset) = 0$
2- $P(A^C) = 1 - P(A)$, A^C is A complement (negation), $\bar{A} \cong A^C$
3- $P(A) \leq 1$

Proof:

1- $P(\Omega) = P(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega) = 1$, thus $P(\emptyset) = 0$
2- $P(\Omega) = P(A \cup A^C) = P(A) + P(A^C) = 1$, thus $P(A^C) = 1 - P(A)$
3- $P(A) + P(A^C) = 1$, $P(A) \geq 0$ $P(A^C) \geq 0$ thus $P(A) = 1 - P(A^C)$ >> $P(A) \leq 1$

Theorem:

1- $P(B \cap A^C) = P(B) - P(A \cap B)$
2- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3- $A \subset B \rightarrow P(A) \leq P(B)$, (if any event A included in B then the probability of that event A is less than or equal the probability of B)

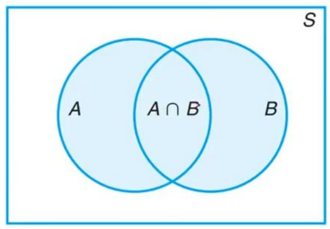


Proof:

1- $P(B) = P(\{B \cap A\} \cup \{B \cap A^C\}) = P(B \cap A) + P(B \cap A^C)$
$P(B \cap A^C) = P(B) - P(A \cap B)$
2- We see that $A \cup B = A \cup \{B \cap A^C\}$ as the two components are disjoint: $P(A \cup B) = P(A) + P(B \cap A^C) = P(A) + P(B) - P(A \cap B)$
3- If $A \subset B$, then $A \cap B = A$ therefore, using point 1:
$0 \leq P(B \cap A^C) = P(B) - P(A)$

If A and B are two events, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

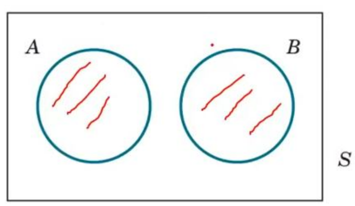


For three events A, B, and C,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

If A and B are mutually exclusive, then

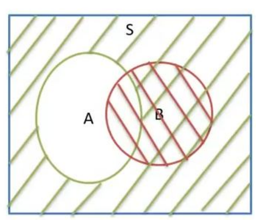
$$P(A \cup B) = P(A) + P(B).$$

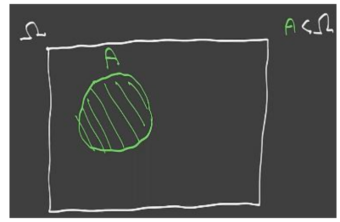


Example4:

If $P(A) = 0.3$, $P(B) = 0.2$, $P(A \cap B) = 0.1$ determine the following probabilities:

3) $P(A' \cap B)$
$= P(B) - P(A \cap B) = 0.2 - 0.1$
$= 0.1$





Some definition from set theory:
A **set** is bag of things, set of sets for instance Van diagram.

Set of Sample space Omega Ω include other set A of event then $A \subset \Omega$

$A \subset B \Leftrightarrow (\omega \in A \Rightarrow \omega \in B)$

(A included in B then if we have an element ω belong to A then necessarily the element ω is belong to B

$A = B \Leftrightarrow A \subset B$ and $B \subset A$

(A & B sets are equal then necessarily A included in B and B included in A)

$A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\}$

(Can be read as A or B: The set of elements ω that either in A or in B)

$A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\}$

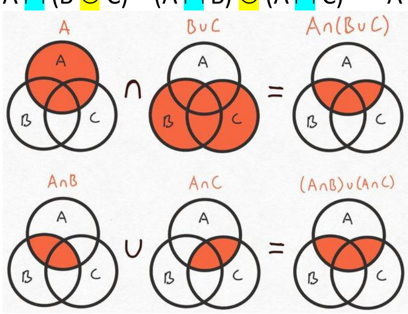
(Can be read as A and B: The set of elements ω that are in A and in B)

$A^C = \{\omega : \omega \notin A\}$

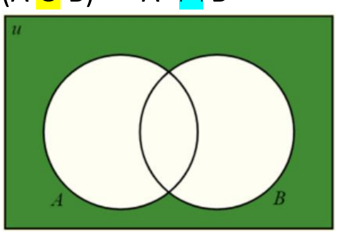
(The set of elements that are not in A)

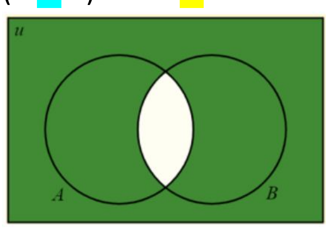
Theorem:

1- **Commutativity**:
   $A \cup B = B \cup A$ $\quad$ $A \cap B = B \cap A$
2- **Associativity:**
   $A \cup (B \cup C) = (A \cup B) \cup C$ $\quad$ $A \cap (B \cap C) = (A \cap B) \cap C$
3- **Distributive laws:**
   $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $\quad$ $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$



4- **De Morgan's Law:**
   $(A \cup B)^C = A^C \cap B^C$ $\quad$ $(A \cap B)^C = A^C \cup B^C$





RECAP:

1- Sample space Ω : is a set of all possible outcomes of random experiment.
2- Event is a subset of Ω.
3- Probability is a function over events (4 axioms most important **for any event in Sigma algebra** is non negative but it can be zero or positive.)
4- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ always

Conditional Probability:

Definition
Given $A, B \in \mathcal{A},\ P(B) > 0$:

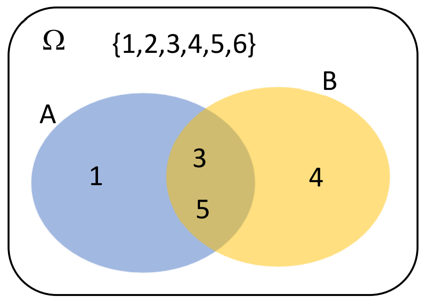
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

"Probability on A conditional on B is defined by Probability of the intersection between A & B divided by the probability of B whenever probability of B is positive", represent that the probability of event A will occur given that event B has already occurred.

Example:

Rolling a dice one time, sample space $\Omega = \{1,2,3,4,5,6\}$, P (Ω)=1/6
Giving event A odd numbers when rolling the dice, A= {1,3,5} and event B= {3,4,5}

What is the probability for A given that B has already occurred?
Simple way: how much of A is in B?
the elements in A which are in B is {3,5},
2 elements of A in the 3 elements of B, so P(A|B) = 2/3



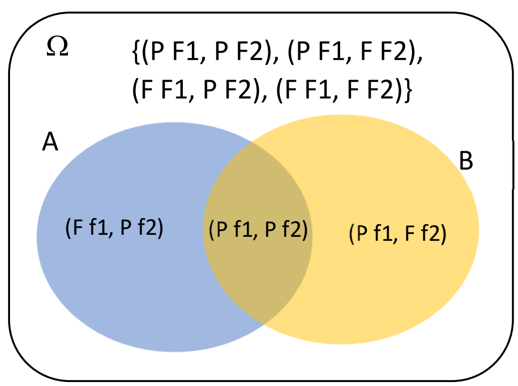
P(A|B) = P(A and B)/P(B) = P(A∩B)/P(B)
B has 3 elements out of 6 of the sample space, so P(B)=3/6
A∩B = {3,5} which is 2 element over 6 of the sample space, so P(A∩B)=2/6
P(A|B) = (2/6) / (3/6) = 2/3

Example:

Relating 2 events:

- You will take exams for 2 classes (FSML-1, FSML-2).
- You can pass or fail either of them
- Let's consider 2 events
  A = "Passing FSML-2 exam"
  B = "Passing FSML-1 exam"
- Knowledge of B should influence our assessment of P(A)?

Ω is made of 4 points each point is the pair stands for (pass or fail for each class)
like (Pass A, Pass B), (Pass A, Fail B), (Fail A, Pass B), (Fail A, Fail B)
A = (P f1, P f2), (F f1, P f2)



P(A) = 1/2

B= (P f1, P f2), (P f1, F f2)

P(B) = ½

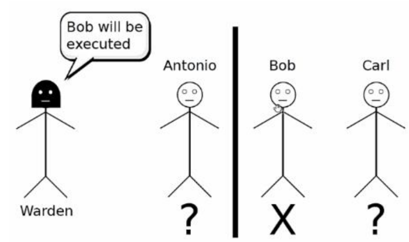
A ∩ B = {(P f1, P f2)}

P(A ∩ B) = ¼ (one over the sample space of 4)

Logically P(A|B) probability of A when B occurred,
There is only one element from A is in B while B has 2 elements
So P(A|B) = ½
P(A|B) = P(A and B)/P(B) = P(A∩B)/P(B)
= (1/4) / (1/2) = 1/2

Suppose that you already pass first Module that mean the other probability (F,P) and (F,F) will not considered, then we will have only (P,P) or (P,F)

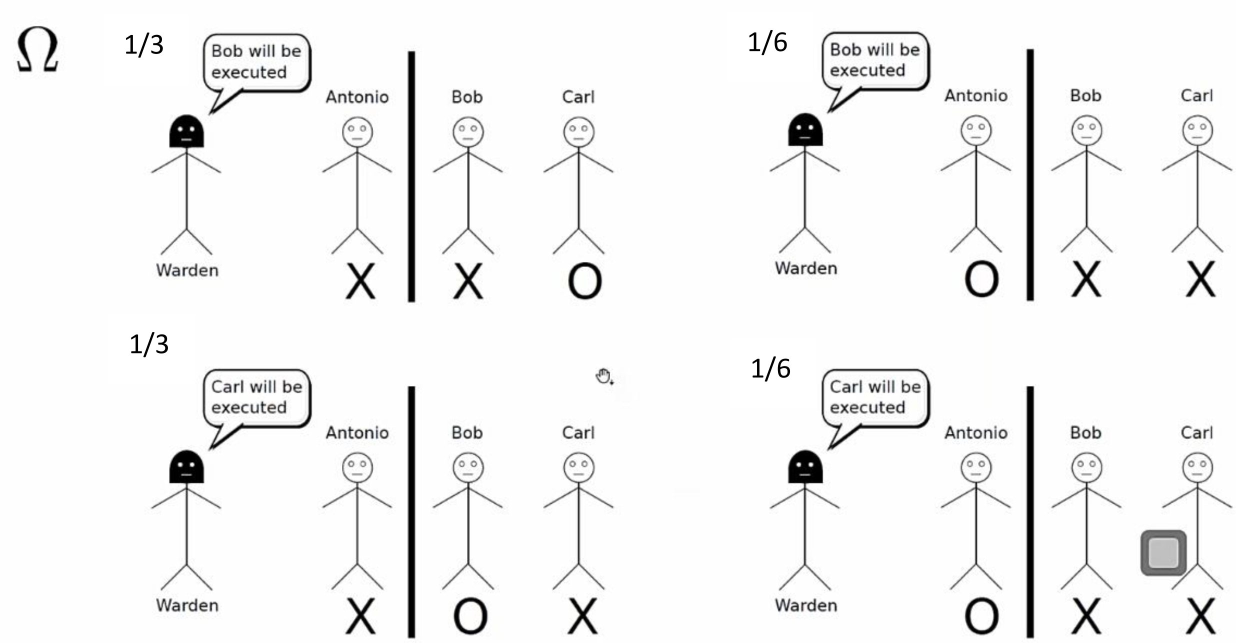
Example:



Example: problem

- 3 persons A, B, C are on death row.
- Governor decides to pardons one of the prisoners.
- The information we have that B will
- So, A thinks his chances are now 1/2

Sample space (Ao,Bx,Cx) , (Ax,Bo,Cx), (Ax, Bx, Co)
Other way >>
Warden : Antonie Executed >> (Ax,Bx,Co) , (Ax,Bo,Cx)
Warden : Bob Executed >> (Ao,Bx,Cx) , (Ax,Bx,Co)
Warden : Carl Executed >> (Ao,Bx,Cx) , (Ax,Bo,Cx)
A∩B = (Ao,Bx,Cx) , (Ax,Bx,Co) , P(A∩B) = 2/3
B = (Ao,Bx,Cx) , (Ax,Bx,Co), P(B)=1/2
P(A|B) = (2/3)/(1/2) =2/6 = 1/3



lets describe the sample space first:

i	ω_i = (pardoned, revealed)	$P(\omega_i)$
1	(A, B)	1/6
2	(A, C)	1/6
3	(B, C)	1/3
4	(C, B)	1/3

Lets call $W = \{\text{'B is revealed'}\} = \{(A,B),(C,B)\}$
What's $P(A|W)$?

$$\begin{aligned} P(A|W) &= P(A\cap W)/P(W) \\ &= \frac{P(\{(A,B),(A,C)\}\cap\{(A,B),(C,B)\})}{P(W)} \\ &= P(\{(A,B)\})/P(W) \\ &= 1/6/(1/6+1/3) \\ &= 1/3 \end{aligned}$$

In other words, $P(A|W) = P(A) = 1/3$, and the warden has not provided any useful information to prisoner A.

Example:

Example

Monty Hall Problem

- 3 identically looking doors: A, B, C
- there's a prize behind one door
- first, the guest player chooses one door, say A
- second, the host opens one of the remaining 2 doors which does not contain the prize, and offers the player the option to change his/her mind and choose the other, remaining door
- question: should the player switch door?

Note: we assume the host always opens one of the remaining doors, in any given game

Example

Strategy 1. Stick with your initial choice of, say, door A

$$\Omega = \{\text{'prize is behind A','prize is behind B','prize is behind C'}\}$$

$$P(\text{win}) = 1/3$$

Strategy 2, Switch to different door than first choice A, switch to the remaining closed door.

Ω= {W1: "prize behind A, Host opened B, I switched to C",
W2: "prize behind A, Host opened C, I switched to B",
W3: "prize behind B, host opened C, I switched to B",
W4: "prize behind C, Host opened B, I switched to C"}

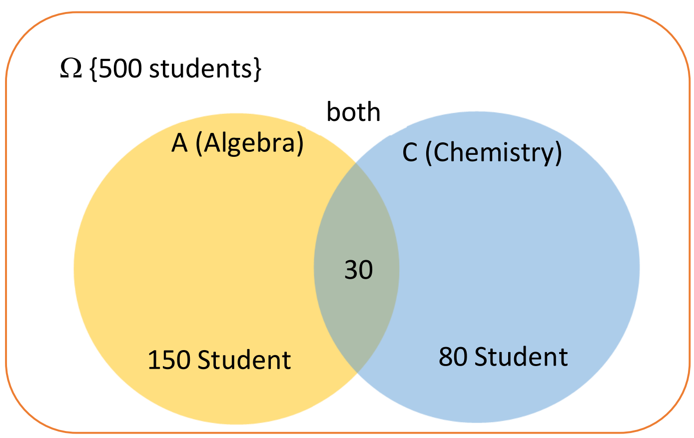
W1, W2 are same where prize behind A and I selected different door

P(win) = P ({w3, w4}) = 2/3

Example:

There are 500 students in a certain school, 150 students are enrolled in Algebra course and 80 students are enrolled in a Chemistry course. There are 30 students who are taking both Algebra and Chemistry. If a student is chosen at random,

(a) What is the probability that the student is taking Algebra?
(b) What is the probability that the student is taking Chemistry given that the student is also taking Algebra?
(c) What is the probability that the student is taking Algebra given that the student is also taking Chemistry?



P(A) = 150/500 = 3/10 = 0.3 or 30%

P(C) = 80/500

P(C|A) = logical way How many from C in A
= 30 / 150 = 1/5

C∩A = 30, P(C∩A) – 30/500

P(C|A) = P(C∩A)/P(A) = (30/500)/ (150/500)
= 30/150 = 1/5

P(A|C) = logical way How many from A in C
= 30/80 = 3/8

P(C|A) = P(A∩C)/P(C) = ((30/500) / (80/500) = 30/80 = 3/8

Example:

There are 200 birds in a zoo, 70 birds are male with brown eyes and 100 birds are female with brown eyes, 20 of the birds are male with blue eyes and 10 birds are female with blue eyes. Construct a contingency table. If a bird is selected at random, what is the probability that the bird is

(a) Female? (b) a male with brown eyes? (c) a female given that it has brown eyes? (d) a male given that it has blue eyes? (e) a creature with blue eyes given that it's a female.

Contingency table:

Gender	Brown eyes	Blue eyes	total
Male	70	20	90
Female	100	10	110
total	170	30	200

(a) Probability that the bird is female: P(F) = 110/200 = 11/20
(b) Probability that a male with brown eyes not conditional yet: P(M&Br) = 70/200 = 7/20
(c) Probability that a female given that it has brown eyes: P(F|BR) = 100/170 = 10/17 = 0.588
(d) Probability that a male given that it has blue eyes: P(M|BL) = 20/30 = 2/3
(e) Probability that a creature with blue eyes given that it's a female: P(BL|F) = 10/110 = 1/11

Independence:

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

Note: When A and B Independent, means:

$$P(A|B) = P(A)\ ,\ P(B|A) = P(B)$$

Theorem:

If A, B are independent, then also:

- A, B^C - A^C, B A^C, B^C

Are Independent.

Important Notes:

Let's not confuse independence with incompatibility!

A, B are said to be incompatible if $A \cap B = \varnothing$

A, B are said to be independent if $P(A \cap B) = P(A)P(B)$

Independence:

S is a sample space, A, B are two events
$A, B \subseteq S$ and A, B are **independent**

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

$$\therefore P(A \cap B) = P(A) * P(B)$$

$$\therefore P(A \cup B) = P(A) + P(B) - P(A) * P(B)$$

Example4:

- If $P(A) = 0.2$, $P(B) = 0.3$ determine the following probabilities:

If A, and B are **disjoint** (mutually exclusive)

Solution:

$P(A \cap B) = 0$

$P(A \cup B) = P(A) + P(B) = 0.2 + 0.3 = 0.5$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

Example5:

- If $P(A) = 0.2$, $P(B) = 0.3$ determine the following probabilities:

If A, and B are **independent**

Solution:

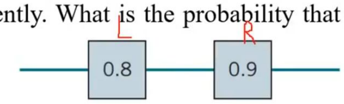
$P(A \cap B) = P(A) * P(B) = 0.2 * 03 = 0.06$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 - 0.06 = 0.44$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) * P(B)}{P(B)} = P(A) = 0.2$$

Example6:

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?

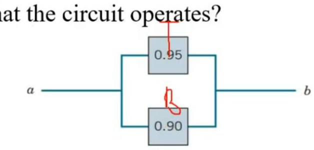


Let L and R denote the events that the left and right devices operate, respectively.

$$P(L \cap R) = P(L)P(R) = 0.80(0.90) = 0.72$$

Example7:

What is the probability that the circuit operates?

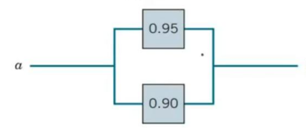


Let T and B denote the events that the top and bottom devices operate, respectively.

$$P(T \cup B) = P(T) + P(B) - P(T)P(B)$$
$$= 0.95 + 0.90 - (0.95)(0.90) = 0.995$$

Example7: Another Solution

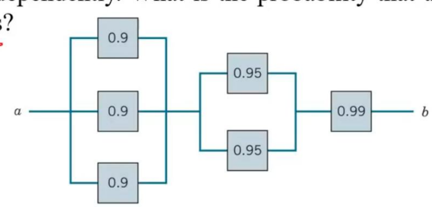
What is the probability that the circuit operates?



$$P(T \cup B) = 1 - P(T \cup B)'$$
$$= 1 - P(T' \cap B') = 1 - P(T')P(B')$$
$$= 1 - (0.05)(0.10) = 0.995$$

Example8:

The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently. What is the probability that the circuit operates?



$P(L1 \cup L2 \cup L3) \cap P(M1 \cup M2) \cap P(R1)$

$P(L1 \cup L2 \cup L3) = 1 - P(L1 \cup L2 \cup L3)^C = 1 - P(L1^C \cap L2^C \cap L3^C)$

$= 1 – (P\ L1^C) * (P\ L2^C) * (P\ L3^C) = 1 – (0.1)*(0.1)*(0.1) = 0.999$

$P(M1 \cup M2) = 1\ P(M1 \cup M2)^C = 1 - P(M1^C \cap M2^C)$

$= 1 - P(M1^C)*P(M2^C) = 1 – (0.05)*(0.05) = 0.9975$

$P(L1 \cup L2 \cup L3) \cap P(M1 \cup M2) \cap P(R1) = (0.999) *(0.9975) *(0.99)$

$= 0.9865$

Example:

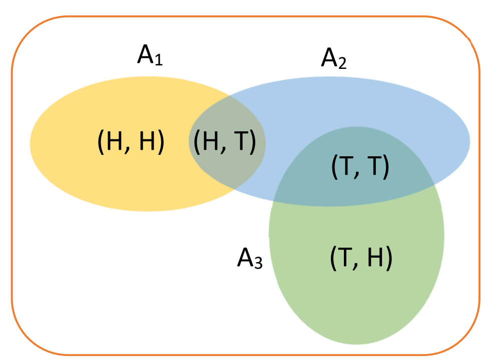
Tossing a fair coin twice:

Ω = {(H, H), (H, T), (T, H), (T, T)}

A1: the first result is H >> = {(H, H), (H, T)}

A2: the second result is T >> = {(H, T), (T, T)}

A3: the first result is T >> = {(T, H), (T, T)}



What is the relation between A1, A2, A3?

Probability for each event:

$P(A_1) = 1/2$ $P(A_2) = 1/2$ $P(A_3) = 1/2$

Probability of the intersection between each two events:

$A1 \cap A2$ = {HT} 1 of 4 >> not $\varnothing$ (not incompatible)

$P(A1 \cap A2) = 1/4$

$P(A \cap B) = P(A)P(B)$ >> $P(A1 \cap A2) = 1/2 * 1/2 = 1/4$

First check the probability of the result of the intersection

Then apply the theorem if both are equal then A1, A2 are independent and if no intersection then both are not incompatible

A1, A2 are independent and not incompatible

$A1 \cap A3 = \varnothing$ >> (incompatible)
$P(A1 \cap A3) = 0$
$P(A \cap B) = P(A)P(B)$ >> $P(A1 \cap A3) = 1/2 * 1/2 = 1/4$
not equal to the probability of the intersection
A1, A3 are incompatible and not independent

Probability of the intersection between each two events:
$A2 \cap A3 = \{TT\}$ 1 of 4 >> not $\varnothing$ (not incompatible)
$P(A2 \cap A3) = 1/4$
$P(A \cap B) = P(A)P(B)$ >> $P(A2 \cap A3) = 1/2 * 1/2 = 1/4$
A2, A3 are independent and not incompatible

Independence: more than 2 events

Definition
$A_1, \ldots, A_n$ are *mutually independent* iff, for any collection $A_{i_1}, \ldots, A_{i_k}$, we have:

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$$

A1, A2, A3 are mutually independent iff
$P(A1 \cap A2)=P(A1)P(A2)$
$P(A2 \cap A3)=P(A2)P(A3)$
$P(A1 \cap A3)=P(A1)P(A3)$
$P(A1 \cap A2 \cap A3)=P(A1)P(A2)P(A3)$

Example:

Tossing 2 fair dice:
Ω = ((1,1), (1,2), …. (6,6)} >> 36 elements
A = {doubles appear} = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)} >> 6 elements
B = {the sum is between 7 and 10} =
B= {(1, 6), (2, 5), (3, 4), (4,3), (5,2), (6,1), (2,6), (3,5), (4,4), (5,3), (6,2), (3,6), (4,5) ,(5,4), (6,3), 4,6), (5,5), (6,4)}
C = {the sum is either 2, 7, or 8} = {(1, 1), (1, 6), (2, 5), (3, 4), (4,3), (5,2), (6,1), (2,6), (3,5), (4,4), (5,3), (6,2)}
P(A) = 6/36 = 1/6
P(B) = 18/36 = 1/2
P(C) = 12/36 = 1/3
$A \cap B$ = {(4, 4), (5, 5)}
$B \cap C$ = {(1, 6), (2, 5), (3, 4), (4,3), (5,2), (6,1), (2,6), (3,5), (4,4), (5,3), (6,2)}
$A \cap C$ = {(1, 1),(4, 4)}
$A \cap B \cap C$ = {(4,4)}

Ω = {
(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)
{

$P(A \cap B)$ = 2/36 = 1/18
$P(B \cap C)$ =11/36
$P(A \cap C)$ = 2/36 = 1/18
$P(A \cap B \cap C)$ = 1/36

By calculation:
$P(A \cap B)$ =P(A)P(B)= 1/6 * 1/2 = 1/12
$P(B \cap C)$ =P(B)P(C) = 1/2 * 1/3 = 1/6
$P(A \cap C)$ = P(A)P(C) = 1/6 * 1/3 = 1/18 (independent)
$P(A \cap B \cap C)$ =P(A)P(B)P(C) = 1/6 * 1/2 * 1/3 = 1/36 (independent)
But not mutually independent after all. 2 ways independent was not enough to guarantee pairwise independent.

Example

triplets of letters

Ω = all permutations of letters a, b, c including triples (9 elements)

$$\Omega = \left\{ \begin{array}{ccc} abc, & acb, & cab, \\ bac, & bca, & cba, \\ aaa, & bbb, & ccc \end{array} \right\}$$

with $P(\omega_i) = 1/9 \quad \forall i$

Consider the events:

$$A_i = \{\text{the i-th letter is an 'a'}\}$$

Are the A_i mutually independent?

A1 = first letter is an a = {abc, acb, aaa} P(A1)=3/9=1/3
A2 = second letter is an a = {cab, bac, aaa} P(A2) = 3/9 = 1/3
A3 = third letter is an a = {bca, cba, aaa} P(A3) = 3/9 = 1/3

A1 ∩ A2 = {aaa}, P(A1 ∩ A2) = 1/9 = P(A1)P(A2) = 1/3*1/3 = 1/9 (independent)
A2 ∩ A3 = {aaa}, P(A2 ∩ A3) = 1/9 = P(A2)P(A3) = 1/3*1/3 = 1/9 (independent)
A1 ∩ A3 = {aaa}, P(A1 ∩ A3) = 1/9 = P(A1)P(A3) = 1/3*1/3 = 1/9 (independent)
A1 ∩ A2 ∩ A3 = {aaa}, P(A1 ∩ A2 ∩ A3) = 1/9 not equal <> P(A1)P(A2)P(A3) = 1/3*1/3*1/3 = 1/27
Pairwise independence was not enough to guarantee 3-way independence, Ai not mutually independent

Example:

3 fair coins tosses

Ω = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Ai= {the i-th toss is a H}

Are A1, A2, A3 mutually independent?

A1 = {HHH, HHT, HTH, HTT} P(A1) = 4/8 = 1/2
A2 = {HHH, HHT, THH, THT} P(A2) = 4/8 = 1/2
A3 = {HHH, HTH, THH, TTH} P(A3) = 4/8 = 1/2

A1 ∩ A2 = {HHH, HHT} >> P (A1 ∩ A2) = 2/8 = 1/4 = P (A1 ∩ A2) = P(A1) P(A2) =1/2*1/2=1/4
A2 ∩ A3 = {HHH, THH} >> P (A2 ∩ A3) = 2/8 = 1/4 = P (A2 ∩ A3) = P(A2) P(A3) =1/2*1/2=1/4
A1 ∩ A3 = {HHH, THH} >> P (A1 ∩ A3) = 2/8 = 1/4 = P (A1 ∩ A3) = P(A1) P(A3) =1/2*1/2=1/4
They are pairwise independent
A1 ∩ A2∩ A3={HHH} >> P (A1 ∩ A2∩ A3) =1/8 = P (A1 ∩ A2∩ A3) =1/2*1/2*1/2=1/8

The same result holds for A1, A3 and A2, A3, yes, they are mutually independent.

- $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
- $P(A \cap B) = P(A|B)\,P(B)$
- $P(A^c|B) = 1 - P(A|B)$

Multiplication Rule:

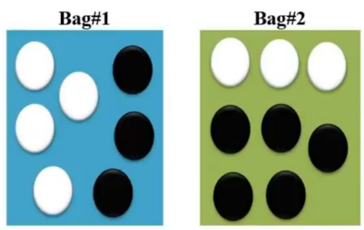
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad , \qquad for \;\; P(B) > 0$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad , \qquad for \;\; P(A) > 0$$

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

Example9: B_1 and B_2 or W_1 and B_2

One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?



Solution:

B_1: Black from bag#1

W_1: White from bag#1

B_2: Black from bag#2

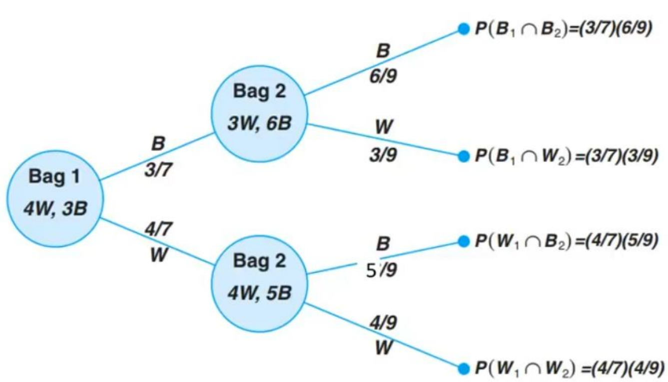
W_2: White from bag#2

$$P(B_1 \cap B_2) = P(B_2|B_1)P(B_1) = \left(\frac{6}{9}\right)\left(\frac{3}{7}\right)$$

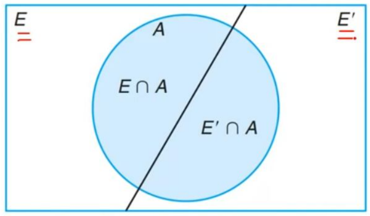
$$P(W_1 \cap B_2) = P(B_2|W_1)P(W_1) = \left(\frac{5}{9}\right)\left(\frac{4}{7}\right)$$

What is the probability that a ball now drawn from the second bag is black $= \left(\frac{6}{9}\right)\left(\frac{3}{7}\right) + \left(\frac{5}{9}\right)\left(\frac{4}{7}\right) = \frac{38}{63}$

Example9: B_1 and B_2 or W_1 and B_2 **Disjoint**



Total Probability Rule:



$$P(A) = P(E \cap A) \cup P(E' \cap A)$$
$$= P(E \cap A) + P(E' \cap A)$$
$$= P(A|E)P(E) + P(A|E')P(E')$$

Example10:

Consider the information about contamination in the following table.

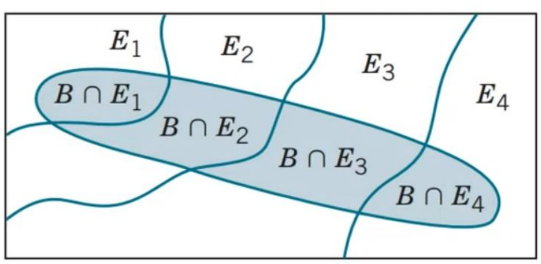
$P(F|H) = 0.1$
$P(F|H') = 0.005$
$P(H) = 0.2$
$P(H') = 0.8$
$P(F)$?

Probability of Failure	Level of Contamination	Probability of Level
0.1	High	0.2
0.005	Not high	0.8

Let F denote the event that the product fails, and let H denote the event that the chip is exposed to high levels of contamination. What is the probability of the product fails?

$$P(F) = P(F|H)P(H) + P(F|H')P(H')$$
$$= (0.1)(0.2) + (0.005)(0.8) = 0.024$$

Total Probability Rule (Multiple Events):



$$B = (B \cap E_1) \cup (B \cap E_2) \cup (B \cap E_3) \cup (B \cap E_4)$$

$$P(B) = P(B \cap E_1) + P(B \cap E_2) + P(B \cap E_3) + P(B \cap E_4)$$

Example11:

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

$P(B_1) = 0.3, P(B_2) = 0.45, P(B_3) = 0.25$

$P(D|B_1) = 0.02,$
$P(D|B_2) = 0.03,$
$P(D|B_3) = 0.02.$

Applying the total probability rule , we can write

$$P(D)$$
$$= P(D|B_1)P(B_1) + P(D|B_2)P(B_2) + P(D|B_3)P(B_3)$$
$$= 0.02(0.3) + 0.03(0.45) + 0.02(0.25) = 0.0245$$

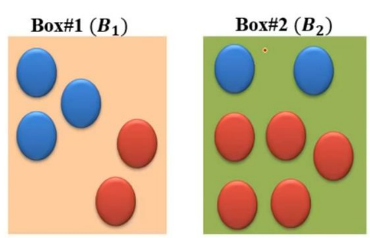
Example12:

Box#1 contains 2 red balls and 3 blue balls; Box#2 contains 5 red balls and 2 blue balls. **If the selection of two boxes is equally likely**, and you selected one ball, what is the probability that it is red?

$P(B_1) = P(B_2) = 0.5$

$R: read, \quad B: blue$

Find $P(R)$?



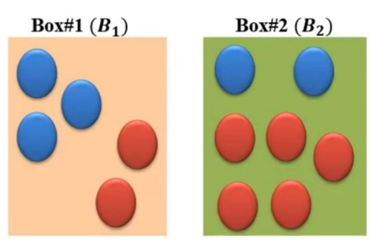
Example12:

$P(B_1) = P(B_2) = 0.5$

$R: read, \quad B: blue$

$P(R|B_1) = {}^2/_5 = 0.4$

$P(R|B_2) = {}^5/_7 = 0.7143$



$$P(R) = P(R|B_1)P(B_1) + P(R|B_2)P(B_2)$$
$$= (0.4)(0.5) + (0.7143)(0.5) = 0.55715$$

Bayes Role:

From the definition of conditional probability,

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

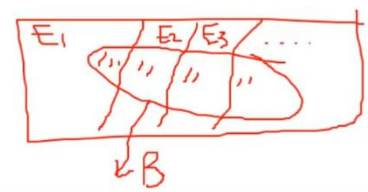
Now, considering the second and last terms in the preceding expression, we can write

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad \text{for} \quad P(B) > 0$$

If $E_1, E_2, \dots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P(E_1 \mid B) = \frac{P(B \mid E_1)P(E_1)}{P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_k)P(E_k)}$$

for $P(B) > 0$



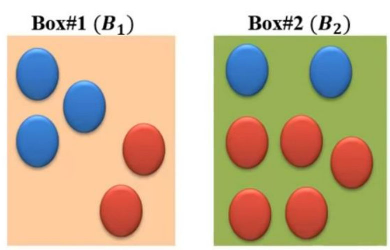
Example1:

Box#1 contains 2 red balls and 3 blue balls; Box#2 contains 5 red balls and 2 blue balls. If the selection of two boxes is equally likely, and the selected ball was red, what is the probability that it is from Box#1?

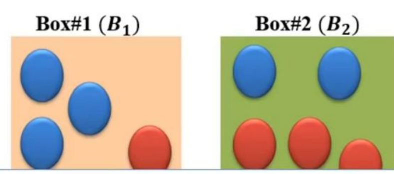
$P(B_1) = P(B_2) = 0.5$

$R: read, \quad B: blue$

Find $P(B_1|R)$ **?**



Example1:



$P(B_1) = P(B_2) = 0.5$

$R: read, \quad B: blue$

$P(R|B_1) = 2/5$

$P(R|B_2) = 5/7$

$P(R) = P(R|B_1)P(B_1) + P(R|B_2)P(B_2)$

$\quad = (0.4)(0.5) + (0.7143)(0.5) = 0.55715$

$$P(B_1|R) = \frac{P(R|B_1)P(B_1)}{P(R)} = \frac{(0.4)(0.5)}{P(R)}$$

Example2:

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. If a product was chosen randomly and found to be defective, what is the probability that it was made by machine B_3?

***D*: the product is defective. Find $P(B_3|D)$?**

Example2: Applying the total probability rule , we can write

$$P(D)$$
$$= P(D|B_1)P(B_1) + P(D|B_2)P(B_2) + P(D|B_3)P(B_3)$$
$$= 0.02(0.3) + 0.03(0.45) + 0.02(0.25) = 0.0245$$

Using Bayes' rule to write

$$P(B_3|D) = \frac{P(D|B_3)P(B_3)}{P(D)} = \frac{(0.02)(0.25)}{0.0245} = 0.2041$$

The Bayes Theorem

Theorem

Let $A_1, A_2, \dots$ be a partition of the sample space, and let B be any event. Then:

$$\forall i, \quad P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

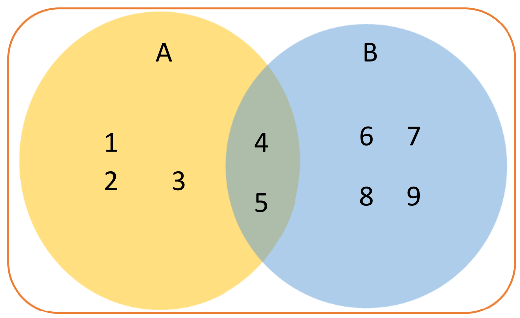
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

We know that:

- P(A|B) = P(A and B) / P(B) = P(A ∩ B) / P(B)
- P(B|A) = P(B and A) / P(A) = P(B ∩ A) / P(A)
- P(A ∩ B) = P(B ∩ A)

Thus

- P(A|B) P(B) = P(B|A) P(A)
- P(A|B) = P(B|A) P(A) / P(B)

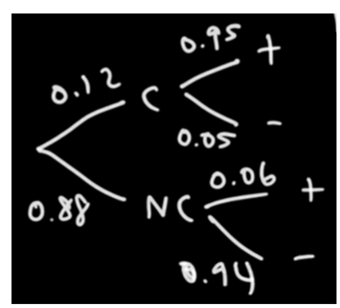


Example:
Let's have A= {1,2,3,4,5} and B= {4,5,6,7,8,9}
So Ω = {1,2,3,4,5,6,7,8,9}
P(A) = 5/9, P(B) = 6/9
P(A|B) = 2/6, P(B|A) = 2/5
P(A|B) = P(B|A) P(A) / P(B) = (2/5. 5/9) / (6/9) = 2/6 = 1/3

Example:

A particular study showed that 12% of men will likely develop prostate cancer at some point in their lives. A man with prostate cancer has a 95% chance of a positive test result from a medical screening exam. A man without prostate cancer has a 6% chance of getting a false positive test result. What is the probability that a man has cancer given that has a positive test result?

P(C) = 0.12 (probability of a men to have prostate cancer)
P(+|C) = 0.95 (probability of Positive test for a man has given Prostate cancer)
P(+|NC) = 0.06 (probability of false Positive test for a man has given without Prostate cancer)
P(C|+) =? What is the probability that a man has cancer given that has a positive test result?
P(C|+) = P(+|C) P(C)/P(+)
We have P(+|C) and P(C) but we need to get P(C|+) and P(+). Let's do the tree diagram
P(+) depends on 2 events (the probability that a person has cancer and positive test result) or (the probability that person doesn't has cancer and has positive test result)
P(+) = P(C and +) + P(NC and +) = (0.12)(0.95) + (0.88)(0.06)=0.1668
P(C|+) = (0.95)(0.12)/(0.1668) = 0.68345
So, 68.34% chance that a man has cancer given that has a positive test result.
P(NC|+) = 100% - 68.35% = 31.71 %
So, 31.71% chance that a man doesn't have cancer even if he given that he has a positive test result.



Example:

Case Study: breast cancer screening

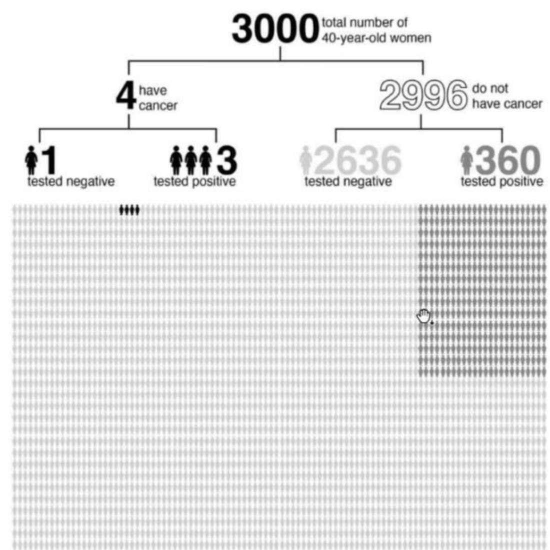
- In 2009, the US Preventive Services Task Force recommended that forty-year-old women should not get annual mammograms.
- A 40 years old woman undergoes breast cancer screening, and the test comes positive.
- For a typical 40 years old woman, the probability of getting breast cancer in the next year is about 1/700 P(D)=1/700
- According to the Breast Cancer Surveillance Consortium (BCSC), the sensitivity of mammograms for 40 years old women is 73%: P(T|D) =0.73
- According to BCSC, false positive rate of mammograms for 40 years old women is 12%: $P(T|D^c) = 0.12$
- P(D|T) =? (D: disease, T: Positive screening test)
-

$P(T) = P(T|D) \times P(D) + P(T|D^C) \times P(D^C) \approx 12.1\ \%$

$P(D|T) = P(T|D) \times P(D) / P(T) = 0.73 \times 1/700/0.121 \approx 0.0086 < 1\%$

Case study: breast cancer screening (cont.)

3 out of 363 40yo women who test positive for breast cancer actually have the disease



from: 'The Book of Why', by J. Pearl and D. Mackenzie, 2018. Data based on false-positive and false-negative rates provided by the Breast Cancer Surveillance Consortium. Infografic by Maayan Harel

Example: Morse code transmission with noise:

In a typical Morse code transmission '.' And '- 'appear with a proportion 3 :4

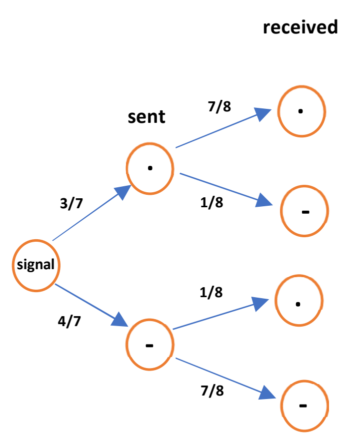
P(. sent) = 3/7 , P(- sent) = 4/7

Suppose there's interference:

P(- received but . sent) = 1/8 P(. received but - sent) = 1/8

P(. received & . sent) = 7/8 P(. received & . sent) = 7/8

We just received a . what's the probability that a . was actually sent?



P(. sent | . received) = P (. sent ∩ . received) / P(. received)

P(. received ∩ - sent) = 1/8 >> P(. received ∩ . sent) = 7/8

P(- received ∩ . sent) = 1/8 >> P(- received ∩ - sent) = 7/8

P(. sent ∩ . received) = (3/7) x (7/8) = 3/8 = 0.375

P(. received) = P(. sent ∩ . received) ∪ P(- sent ∩ . received)

P(. received) = 0.375 + (4/7 x 1/8) = 0.375 + 0.0714 = 0.446

P(. sent | . received) = 0.375 / 0.446 = 0.84 = 84%

Example:

Case Study: diagnosing HIV

A positive result means antibodies to HIV were found in your blood. This means you have HIV infection. You are infected and can spread HIV to others.

Illinois Department of Public Health

What does it mean to get a positive test?

- outcome: $E =$ the test is positive
- scenarios:
  - A_1: the subject is infected with HIV
  - A_2: the subject is not infected with HIV

Note: $A_1 \cup A_2 = \Omega$, and $A_1 \cap A_2 = \emptyset$: A_1, A_2 form a partition of the sample space

- $P(E|A_1) = 99.9\%$ (aka *sensitivity*)
- $P(E|A_2) = 0.01\%$ (aka 1 - *specificity*)

the 2 quantities above are sometimes called likelihoods of the events A_1 and A_2
What we're interested in, though, is the following:

$$P(A_1|E) = ?$$

Pay close attention: $P(A_1|E) \neq P(E|A_1)$!!!

By definition:

$$P(A_1|E) = \frac{P(A_1 \cap E)}{P(E)}$$

We don't directly know $P(A_1 \cap E)$, however, from the definition of conditional probability:

$$P(A_1 \cap E) = P(E \cap A_1) = P(E|A_1)P(A_1)$$

$P(A_1)$ is the *base rate* of the disease. In a general population of young men with no risk behavior, in a developed western country, we can assume e.g. $P(A_1) = 0.01\%$, so we have all we need for $P(A_1 \cap E)$
For $P(E)$, note that we can express it as:

$$P(E) = P((E \cap A_1) \cup (E \cap A_2)) = P(E \cap A_1) + P(E \cap A_2)$$

and we also know how to compute $P(E \cap A_2)$ (do we?)

Putting it all together:

$$P(A_1|E) = \frac{P(E|A_1)P(A_1)}{\sum_{i=1}^{2} P(E|A_i)P(A_i)}$$

Now we just plug in the numbers:

$$P(A_1|E) = \frac{99.9 \times 0.01}{99.9 \times 0.01 + 0.01 \times 99.99} \simeq 50\%$$

Let this sink in for a moment: $P(E|A_1) = 99.9\%$ (you could say, essentially sure). Still, $P(A_1|E) \simeq 50\%$. Yet, it is a common fallacy to confuse $P(E|A_1)$ with $P(A_1|E)$!

Not convinced yet? Think of it this way:

- out of 10000 men, 1 is infected (base rate of 0.01%)
- this infected man will have a 99.9% chance of getting a positive test (i.e., practically sure)
- out of the 9999 non infected men, about 1 will have a positive test ($\simeq$ 0.01% false positive rate)

Now, same question as before: a man tests positive. What's his chance of being infected? Out of 10000 men, 2 will test positive. Of these 2, 1 is really infected. So, given a positive test, there's about a 50% chance of being actually infected

Random Variable

- Is a function that assigns a real number to each outcome in the sample space of random experiment. Denoted by an uppercase letter such as X

A Discrete Random Variable

- Is a random variable with a finite (or countable infinite) range.
- The possible values of X may be listed as $x_1, x_2, \ldots$

Example1

- Flipping a coin of two times. Let X is the number of heads.

Answer:

$$S = \{HH, HT, TH, TT\}$$

$$\quad 2 \quad 1 \quad 1 \quad 0$$

$x = 0, 1, 2$

$$P(0) = \frac{1}{4}, \qquad P(1) = \frac{2}{4}, \qquad P(2) = \frac{1}{4}$$

Probability Mass Function

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$

x_i	x_1	x_2	x_3	x_4	x_5
$f(x_i) = P(x_i)$	$P(x_1)$	$P(x_2)$	$P(x_3)$	$P(x_4)$	$P(x_5)$

Example1

Verify that the function is a probability mass function:

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Answer:

$\sum P(x_i) = 1, \qquad P(x_i) \geq 0$

Example2

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Find:

a. $P(X \leq 2)$ b. $P(X > -2)$

c. $P(-1 \leq X \leq 1)$ d. $P(X \leq -1 \text{ or } X = 2)$

Answer:

a. $P(X \leq 2) = 1$

b. $P(X > -2) = \frac{7}{8}$

c. $P(-1 \leq X \leq 1) = \frac{6}{8}$

d. $P(X \leq -1 \text{ or } X = 2) = \frac{3}{8} + \frac{1}{8} = \frac{4}{8}$

Example3

Two balls are drawn in succession without replacement from a box containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y, where y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Example4

A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Note: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. **Then x can only take the numbers 0, 1, and 2.**

Solution:

For each x value calculate the probability.

- **For 0 means (select 0 from defective of 3, and [∩] 2 from working of 17)**
- **For 1 means (select 1 from defective of 3, and [∩] 1 from working of 17)**
- **For 2 means (select 2 from defective of 3, and [∩] 0 from working of 17)**

Find total number of sample points in sample space

For all possibilities to select 2 from 20 (combination of the 20 selecting 2)

-

The number of combinations, subsets of r elements that can be selected from a set of n elements, is denoted as $\binom{n}{r}$ or C_r^n and

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{136}{190}$$

$$f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$$

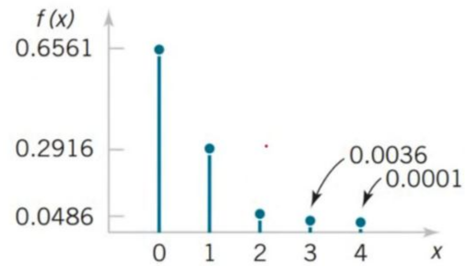
$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$$

Example5

There is a chance that a bit transmitted through a digital transmission channel is received in error. Let X equal the number of bits in **error** in the next **four bits** transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$.

Suppose that the probabilities are

$$P(X = 0) = 0.6561 \quad P(X = 1) = 0.2916$$
$$P(X = 2) = 0.0486 \quad P(X = 3) = 0.0036$$
$$P(X = 4) = 0.0001$$



Random Variable X

(R.V.) X Define it as a real function of the sample space.

X: $\Omega \longmapsto \mathbb{R}$ (from Omega to the real numbers) $\mathcal{A} \mapsto \mathcal{B}$

Note: in order for X to be a proper RV, it must be *measurable*, i.e.:

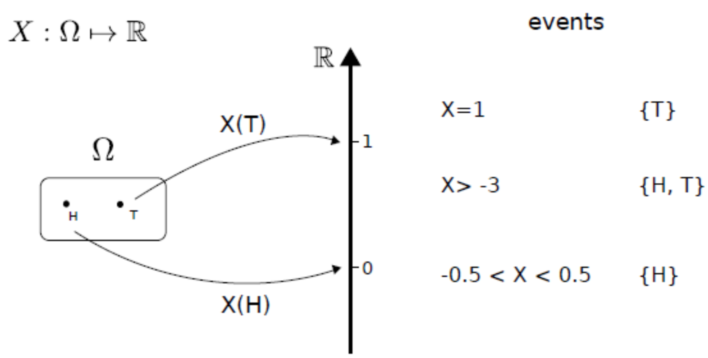
$$\forall B \in \mathcal{B}, \quad X^{-1}(B) \in \mathcal{A}$$

Take any element from the sample space, assign it a number

Example:

tossing the coin, so sample space Ω= {H, T}

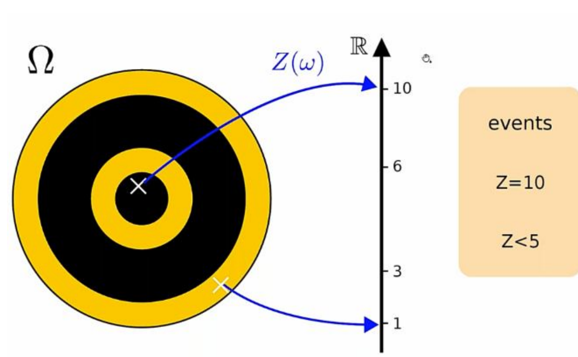
we assign a number to each
X(T) = 1, X(H) = 0
In this situation we assign events and define them in terms of the domain of sample space.
X=1 correspond to the set {T}
X > -3 correspond to the set {H, T}



Example:

suppose we are shooting at the target and random experiment throwing and see what we hit.
And the sample space is the infinite set of points inside the circle.
We define a function Z that assign to every point at the sample space as a score (number)

Z(w) w is an argument of an element in the sample space, assign 10 to all the points in the center circle, and 1 to all the points in the outer circle. And 3 to the next circle and so on



We can define events in terms of Z:
Z=10 (is the full set of all points in the center circle)
Z<5.

Distribution Function

We have random variable X, and the distribution function which assigns probability to possible values to the random variables.

To each possible value of the random variable, we assign a probability.

Equal notation for $P_X(x)$ distribution function of the random variable x calculated at the single point x

$P_X(x)$ = P(X(w)=x) = P ({w: X(w)=x})

(Is the probability of the sample space such that X of this element is x)

= P($X^{-1}(x)$) = P (X=x) (probability of the inverse image of this specific point x)

Distribution function of the random variable B calculated at a bigger set of points.

$P_X(B)$ = P(X(w) ∈ B) = P ({w: X(w) ∈ B}) = P($X^{-1}(B)$) anti image of this set B -> $P_X(B)$

Example

Experiment: toss 2 dice; X = sum of scores

Sample space 36 of all possible pair of dice, a random variable assign to each pair of dice that the sum of the scores. So go from the 36 to an integer number between 2 and 12

Experiment: toss a fair coin 3 times; X = total numbers of heads, so what is the distribution function of X

The random variable could be the count of how many times we got the head.

i	1	2	3	4	5	6	7	8
ω_i	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$P(\omega_i)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
$X(\omega_i)$	3	2	2	2	1	1	1	0

P(wi) the probability of that event

X(wi) how many times we got the head H

The domain or support of X is x= {0,1,2,3} (the set of possible values that the random variable)

The distribution function of X is:

The probability X is equal 3 or probability of the anti-image of the single point 3 is the probability of w such that the set of omega X(w)=3

$$P_X(3) = P(X^{-1}(3)) = P(\{\omega : X(\omega) = 3\}) = P(\{HHH\}) = 1/8$$

$$P_X(2) = P(X^{-1}(2)) = P(\{\omega : X(\omega) = 2\}) = P(\{HHT, HTH, THH\}) = 3/8$$

The distribution function of X:

x	0	1	2	3
P(X=x)	1/8	3/8	3/8	1/8

The cumulative Distribution function (CDF)

The cumulative distribution function (cdf), denoted by $F(x)$, measures the probability that the random variable X assumes a value less than or equal to x, that is,

If X is discrete, then

$$F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$$

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8
$F(x) = P(X \le x)$	**1/8**	**3/8**	**5/8**	**7/8**	**8/8**

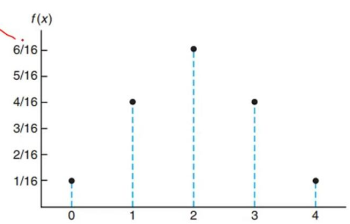
Example1

x	0	1	2	3	4
$f(x) = P(X = x)$	0.6561	0.2916	0.0486	0.0036	0.0001
$\boldsymbol{F(x) = P(X \le x)}$	**0.6561**	**0.9477**	**0.9963**	**0.9999**	**1**

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.6561 & 0 \le x < 1 \\ 0.9477 & 1 \le x < 2 \\ 0.9963 & 2 \le x < 3 \\ 0.9999 & 3 \le x < 4 \\ 1 & 4 \le x \end{cases}$$

Example2

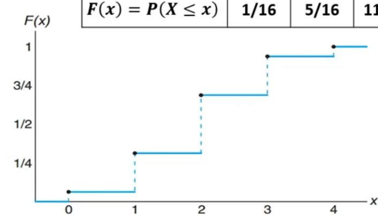
x	0	1	2	3	4
$f(x) = P(X = x)$	1/16	4/16	6/16	4/16	1/16



Probability mass function plot.

Example2

x	0	1	2	3	4
$f(x) = P(X = x)$	1/16	4/16	6/16	4/16	1/16
$\boldsymbol{F(x) = P(X \le x)}$	**1/16**	**5/16**	**11/16**	**15/16**	**16/16**



Discrete cumulative distribution function.

Cumulative Distribution Function of a RV (CDF)

Given a RV X, $F_X(t) = P(X \le t),\ t \in \mathbb{R}$

CDF is the probability that random variable is less than or equal some threshold t for t being any real number (the probability that X $\le$ t)

Example

Tossing a coin 3 times, X = number of heads, what CDF

CDF >> FX(t) = P (X$\le$t)

The CDF of X:

x	0	1	2	3
P(X=x)	1/8	3/8	3/8	1/8
CDF	1/8	1/8+3/8=4/8	1/8+3/8+3/8=7/8	1/8+3/8+3/8+1/8=1

$$F_X(t) = \begin{cases} 0 & t < 0 \\ 1/8 & 0 \le t < 1 \\ 1/2 & 1 \le t < 2 \\ 7/8 & 2 \le t < 3 \\ 1 & 3 \le t \end{cases}$$

Notes

- F_X is defined $\forall t$, not just $\{0, 1, 2, 3\}$. E.g., $F_X(2.5) = P(X \le 2.5) = P(X \in \{0, 1, 2\}) = 7/8$
- F_X *jumps* at x_i, and the jump height is $P(X = x_i)$
- $F_X(t) = 0$ for $t < 0$, as X cannot be negative
- $F_X(t) = 1$ for $t \ge 3$, as $X \le 3$ a.s.

Theorem:

F(x) is a CDF iff:

1- $\lim\limits_{t \to -\infty} F(t) = 0,\ \lim\limits_{t \to \infty} F(t) = 1$

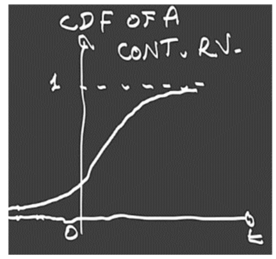
2- F is non-decreasing

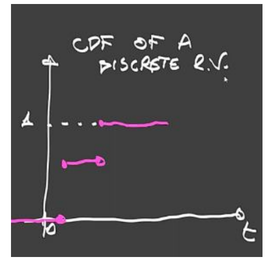
3- F is right continuous: $\lim\limits_{t \to x_0^+} F(t) = F(x_0)$

Definition:
We said a Random Variable (RV) is:
continuous iff F(t) is continuous
discrete iff F(t) is a step function

X, Y are identically distributed (we write: $X \sim Y$) iff $\forall A \in B$, $P(X \in A) = P(Y \in A)$

Note: $X \sim Y$ does not imply $X = Y$
E.g.:
3 coins tosses, (identically distributed)
X = number of heads,
Y = number of tails
We have
$X \sim Y$, but $X \neq Y$ for all $w \in \Omega$





Theorem:
$X \sim Y$ iff $F_X = F_Y$ (if they have the same CDF then they are identically distributed)
Definition:
Probability Mass Function PMF of a discrete RV of X:

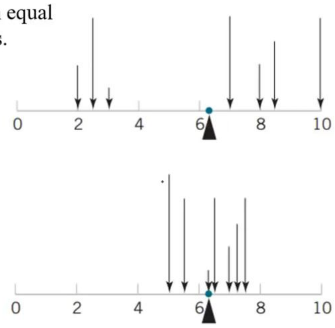
$f_X(x) = P(X = x)\ \forall x \in \chi$ for any x of the domain χ

(The probability of single point of discontinuity point which is the heights of the jumps of the CDF at the discontinuity points)

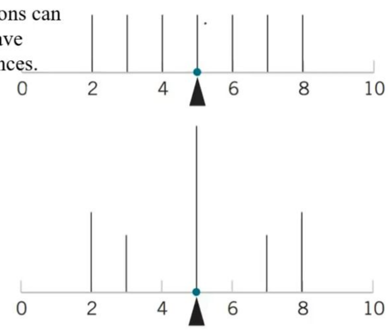
Mean & Variance:

Two numbers are often used to summarize a probability distribution for a random variable X. The mean is a measure of the center or middle of the probability distribution, and the variance is a measure of the dispersion, or variability in the distribution.

Probability distributions with equal means but different variances.



Two probability distributions can differ even though they have identical means and variances.



Mean, Variance, and Standard deviation

The mean or expected value of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x x f(x)$$

The variance of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X-\mu)^2 = \sum_x (x-\mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

$$\boxed{E(X^2) - (E(X))^2}$$

The standard deviation of X is $\sigma = \sqrt{\sigma^2}$.

Example1

x	-2	-1	0	1	2
$f(x) = P(X = x)$	1/8	2/8	2/8	2/8	1/8

Find:

Determine the mean and variance of the random variable X

Answer: (1/2)

$E(X) =$

$$\sum x_i P(x_i) = (-2)\left(\frac{1}{8}\right) + (-1)\left(\frac{2}{8}\right) + (0)\left(\frac{2}{8}\right) + (1)\left(\frac{2}{8}\right) + (2)\left(\frac{1}{8}\right)$$

$$= 0$$

$V(X) = E(X^2) - \left(E(X)\right)^2$

$E(X) = 0$

$E(X^2)$

$$= \sum x_i^2 P(x_i) = (4)\left(\frac{1}{8}\right) + (1)\left(\frac{2}{8}\right) + (0)\left(\frac{2}{8}\right) + (1)\left(\frac{2}{8}\right) + (4)\left(\frac{1}{8}\right) = 1.5$$

$V(X) = 1.5 - (0)^2 = 1.5,$ Standard Deviation $(\sigma) = \sqrt{1.5}$

Example2:

A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

Let X represent the number of good components in the sample. **Then x can only take the numbers 0, 1, 2 and 3.**

The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$f(0) = P(X = 0) = \frac{\binom{4}{0}\binom{3}{3}}{\binom{7}{3}} = \frac{1}{35} \qquad f(1) = P(X = 1) = \frac{\binom{4}{1}\binom{3}{2}}{\binom{7}{3}} = \frac{12}{35}$$

$$f(2) = P(X = 2) = \frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}} = \frac{18}{35} \qquad f(3) = P(X = 3) = \frac{\binom{4}{3}\binom{3}{0}}{\binom{7}{3}} = \frac{4}{35}$$

x	0	1	2	3
$f(x) = P(X = x)$	1/35	12/35	18/35	4/35

$$E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7$$

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

$$E(X^2) = \sum x_i^2 P(x_i) = 0\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (4)\left(\frac{18}{35}\right) + (9)\left(\frac{4}{35}\right) = \frac{120}{35} = 3.43$$

$V(X) = 3.43 - (1.7)^2 = 0.54,$ Standard Deviation $(\sigma) = \sqrt{0.54} = 0.74$

For any constants *a* and *b*:

Mean

1. $E(a) = a\,, \quad a \in \mathbb{R}$
2. $E(aX + b) = aE(X) + b\,, \quad a, b \in \mathbb{R}$

Variance

1. $V(a) = 0\,,\ a \in \mathbb{R}$
2. $V(aX + b) = a^2 V(X)\,, \quad a, b \in \mathbb{R}$

Example3 – Answer

A discrete random variable with $V(X) = 2.5$
Evaluate $V(2X + \cancel{1})$

$V(aX + b) = a^2 V(X)\,, \quad a, b \in \mathbb{R}$

$V(2X + 1) = 4V(X) = 4 \times 2.5 = 10$

Example4 – Answer

A discrete random variable with $E(X) = 2.5$
Evaluate $E(2X + 1)$

$E(aX + b) = aE(X) + b\,, \quad a, b \in \mathbb{R}$

$E(2X + 1) = 2E(X) + 1$

$E(2X + 1) = 2 \times 2.5 + 1 = 6$

Example5:

Let X is a random variable with mean 6 and variance 100.
Consider another random variable Y such that
$Y = 3X + 6$, evaluate the mean and variance of Y?

$E(X) = 6 \ ,\ V(X) = 100$

$E(Y) = E(3X + 6) = 3E(X) + 6 = 3(6) + 6 = 24$

$V(Y) = V(3X + 6) = 9V(X) = 9(100) = 900$

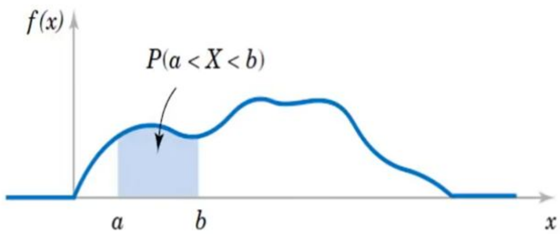
Continuous Random Variable:

If the range space R_X of the random variable X is an interval or a collection of intervals, X is called a *continuous random variable*.

A continuous random variable has a probability of **0** of assuming *exactly* any of its values. Consequently, its probability distribution cannot be given in tabular form.

Example:

If we talk about the probability of selecting a person who is at least 163 centimeters but not more than 165 centimeters tall. Now we are dealing with an interval rather than a point value of our random variable.



If X is a **continuous random variable**, for any x_1 and x_2,

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2)$$

Probability Density Function

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \ge 0$

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

(3) $P(a \le X \le b) = \int_a^b f(x)\,dx =$ area under $f(x)$ from a to b

for any a and b

Definite Integral:

$$\boxed{\int_a^b f(x)\,dx = F(b) - F(a)}$$

$$\int_1^3 x^2\,dx$$

$$\int_1^3 x^2\,dx = \left[\frac{x^3}{3}\right]_1^3 = \left[\frac{3^3}{3}\right] - \left[\frac{1^3}{3}\right] = \left[9 - \left(\frac{1}{3}\right)\right] = \frac{26}{3}$$

Example1:

Suppose that $f(x) = e^{-x}$ for $x > 0$
Check the probability density function, then determine the following probabilities:

1. $P(X < 1)$
2. $P(1 \leq X < 2.5)$
3. $P(X = 3)$
4. $P(X \geq 3)$

Check the probability density function:

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

$$\int_{0}^{\infty} e^{-x}\, dx = -e^{-x}\Big|_{0}^{\infty} = (-e^{-\infty}) - (-e^{0}) = 0 + 1 = 1$$

1) $P(X < 1)$

$$P(X < 1) = \int_{0}^{1} e^{-x}\, dx = -e^{-x}\Big|_{0}^{1} = (-e^{-1}) - (-e^{0})$$

$$= -0.367879 + 1 = 0.632121$$

2) $P(1 \leq X < 2.5)$

$$P(1 \leq X < 2.5) = \int_{1}^{2.5} e^{-x}\, dx = -e^{-x}\Big|_{1}^{2.5}$$

$$= (-e^{-2.5}) - (-e^{-1}) = -0.082085 + 0.367879$$

$$= 0.449964$$

3) $P(X = 3)$

$$P(X = 3) = 0$$

4) $P(X \geq 3)$

$$P(X \geq 3) = \int_{3}^{\infty} e^{-x}\, dx = -e^{-x}\Big|_{3}^{\infty} = (-e^{-\infty}) - (-e^{-3})$$

$$= 0 + 0.049787 = 0.049787$$

Example2:

Suppose that the error in the reaction temperature, in °C (Celsius), for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

Check the probability density function:

$$\int_{-1}^{2} \frac{x^2}{3}\, dx$$

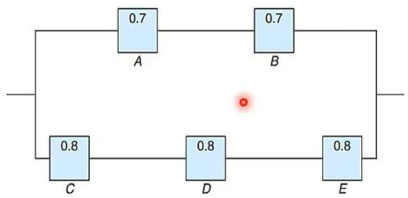
$$\int_{-1}^{2} \frac{x^2}{3}\, dx = \frac{x^3}{9}\Big|_{-1}^{2} = \left(\frac{(2)^3}{9}\right) - \left(\frac{(-1)^3}{9}\right) = \frac{8}{9} + \frac{1}{9} = 1$$

2) $P(\mathbf{0} < \boldsymbol{X} \leq \mathbf{1})$

$$P(\mathbf{0} < \boldsymbol{X} \leq \mathbf{1}) = \int_0^1 \frac{x^2}{3}\, \boldsymbol{dx} = \frac{x^3}{9}\Big|_0^1$$

$$= \left(\frac{(1)^3}{9}\right) - \left(\frac{(0)^3}{9}\right) = \frac{1}{9} + \frac{0}{9} = \mathbf{\frac{1}{9}}$$

1. A circuit system is given in the following figure. Assume the components fail independently. What is the probability that the entire system works?



UP: (0.7) (0.7) = 0.49

Down: (0.8) (0.8) (0.8) = 0.512

$P(W) = 1 - [(0.51)\,(0.488)] = 0.75112$

2. If a multiple-choice test consists of 6 questions, each with 4 possible answers of which only 1 is correct. In how many ways can a student check off one answer to each question and get all the answers wrong?

Ways to get all the answers wrong $= (3)^6 = 729$

5. The likely location of a mobile device in the home is as follows:

Adult bedroom: 0.10, Child bedroom: 0.20,
Office: 0.40, Other rooms: 0.30

(a) What is the probability that a mobile device is in a bedroom?

$= P(Adult\ bedroom \cup Child\ bedroom) = 0.10 + 0.20 = 0.30$

(b) What is the probability that it is not in a bedroom?

$= 1 - 0.30 = 0.70$

6. How many distinct permutations can be made from the letters of the word COMPUTER?

$n! = 8! = 40{,}320$

Rev. Example:

Suppose that the error in the reaction temperature, in °C (Celsius), for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

Check the probability density function:

$$\int_{-1}^{2} \frac{x^2}{3}\, \boldsymbol{dx}$$

$$\int_{-1}^{2} \frac{x^2}{3}\, \boldsymbol{dx} = \frac{x^3}{9}\bigg|_{-1}^{2} = \left(\frac{(2)^3}{9}\right) - \left(\frac{(-1)^3}{9}\right) = \frac{8}{9} + \frac{1}{9} = \mathbf{1}$$

2) $\boldsymbol{P(0 < X \leq 1)}$

$$\boldsymbol{P(0 < X \leq 1)} = \int_{0}^{1} \frac{x^2}{3}\, \boldsymbol{dx} = \frac{x^3}{9}\bigg|_{0}^{1}$$

$$= \left(\frac{(1)^3}{9}\right) - \left(\frac{(0)^3}{9}\right) = \frac{1}{9} + \frac{0}{9} = \mathbf{\frac{1}{9}}$$

Cumulative Distribution Function

The **cumulative distribution function** of a continuous random variable X is

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

for $-\infty < x < \infty$.

Probability Density Function from the Cumulative Distribution Function

Given $F(x)$,

$$f(x) = \frac{dF(x)}{dx}$$

as long as the derivative exists.

$$P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx},$$

if the derivative exists.

Example1:

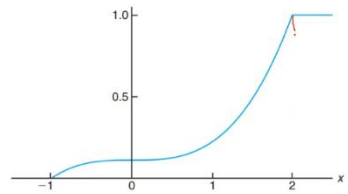
Find the cumulative distribution function $F(x)$ and use it to evaluate $P(0 < X \leq 1)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

$$F(x) = \int_{-\infty}^{x} \frac{u^2}{3}\, \boldsymbol{du} = \int_{-1}^{x} \frac{u^2}{3}\, \boldsymbol{du} = \frac{u^3}{9}\Big|_{-1}^{x} = \left(\frac{(x)^3}{9}\right) - \left(\frac{(-1)^3}{9}\right)$$

$$= \frac{x^3}{9} + \frac{1}{9} = \frac{x^3+1}{9}$$

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



Find the cumulative distribution function $F(x)$ and use it to evaluate $P(0 < X \leq 1)$.

$$\boxed{F(x) = \frac{x^3+1}{9}}$$

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}$$

Mean and Variance

Suppose that X is a continuous random variable with probability density function $f(x)$. **mean** or **expected value** of X, denoted as μ or $E(X)$, is

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

The **variance** of X, denoted as $V(X)$ or σ^2, is

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

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Example1 – Answer (1/3)

Find the expected value of x, $E(x)$ and the variance $V(x)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

$$E(x) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{-1}^{2} \frac{x^3}{3}\, dx = \frac{x^4}{12}\Big|_{-1}^{2}$$

$$E(x) = \int_{-1}^{2} \frac{x^3}{3}\, dx = \frac{x^4}{12}\Big|_{-1}^{2} = \frac{(2)^4}{12} - \frac{(-1)^4}{12} = \frac{15}{12}$$

$$E(x^2) = \int_{-1}^{2} \frac{x^4}{3}\, dx = \frac{x^5}{15}\Big|_{-1}^{2} = \frac{(2)^5}{15} - \frac{(-1)^5}{15} = \frac{33}{15}$$

$$V(x) = E(x^2) - \left(E(x)\right)^2 = \frac{33}{15} - \left(\frac{15}{12}\right)^2 = 0.6375$$

Standard Deviation (σ) $= \sqrt{0.6375} = 0.798$

Definition:

If X and Y are two random variables, the probability distribution that defines their simultaneous behavior is called a joint probability distribution.

If X and Y are discrete random variables:

Joint Probability Mass Function

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

(1) $f_{XY}(x, y) \geq 0$
(2) $\sum_X \sum_Y f_{XY}(x, y) = 1$
(3) $f_{XY}(x, y) = P(X = x, Y = y)$

Example1:

A coin is tossed two times. Calculate the joint probability mass function of the discrete random variables X and Y, where X denote the number of heads appear and Y denote the number of tails appear.

$S = \{HH, HT, TH, TT\}$
$X \quad 2 \quad 1 \quad 1 \quad 0$
$Y \quad 0 \quad 1 \quad 1 \quad 2$

y \ x	0	1	2
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Example2:

A coin is tossed two times. Calculate the joint probability mass function of the discrete random variables X and Y, where X denote the number of heads appear and Y denote the number of tails appear.

Find:

1) $f_{XY}(1,2) = P(X = 1,\ Y = 2)$
2) $f_{XY}(2,0) = P(X = 2,\ Y = 0)$
3) $P(X = 1,\ Y \leq 2)$
4) $P(Y = 2)$

y \ x	0	1	2
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0

Find:

1) $f_{XY}(1,2) = P(X = 1,\ Y = 2) = 0$
2) $f_{XY}(2,0) = P(X = 2,\ Y = 0) = 1/4$
3) $P(X = 1,\ Y \leq 2) = 0 + \frac{2}{4} + 0 = \frac{2}{4}$
4) $P(Y = 2) = \frac{1}{4} + 0 + 0 = \frac{1}{4}$

Marginal Probability Distributions

The marginal distributions of the random variable X alone is:

$$f_X(x) = \sum_y f_{XY}(x, y)$$

The marginal distributions of the random variable Y alone is:

$$f_Y(y) = \sum_x f_{XY}(x, y)$$

Example3:

A coin is tossed two times. Calculate the joint probability mass function of the discrete random variables X and Y, where X denote the number of heads appear and Y denote the number of tails appear.

Find:

1) $f_X(x)$
2) $f_Y(y)$

y \ x	0	1	2
0	0	0	1/4
1	0	2/4	0
2	1/4	0	0
$f_X(x)$	1/4	2/4	1/4

Find:

1) $f_X(x)$

y \ x	0	1	2	$f_Y(y)$
0	0	0	1/4	1/4
1	0	2/4	0	2/4
2	1/4	0	0	1/4

Find:

2) $f_Y(y)$

Independence:

If X and Y are two random variables, the X and Y are independent if any one of the following properties is true:

1) $f_{XY}(x, y) = f_X(x)f_Y(y)$ for all x and y

2) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

3) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

Example1:

$f_{XY}(x, y)$

y \ x	1	1.5	2.5	3
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Find:

$f_X(x)$

$f_Y(y)$

Are X and Y independent?

Example1 – Answer (1/3):

$f_{XY}(x, y)$

y \ x	1	1.5	2.5	3
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8
$f_X(x)$	1/4	3/8	1/4	1/8

Find:

$f_X(x)$

Example1 – Answer (2/3):

$f_{XY}(x, y)$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**	$f_Y(y)$
1	1/4	0	0	0	**1/4**
2	0	1/8	0	0	**1/8**
3	0	1/4	0	0	**1/4**
4	0	0	1/4	0	**1/4**
5	0	0	0	1/8	**1/8**

Find:

$f_Y(y)$

If:

$$\boxed{f_{XY}(x, y) = f_X(x) f_Y(y) \text{ for all } x \text{ and } y}$$

Example1 – Answer (3/3):

$f_{XY}(x, y)$

$y \backslash x$	**1**	**1.5**	**2.5**	**3**
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8

Same as

x	1	1.5	1.5	2.5	3
y	1	2	3	4	5
$f_{XY}(x, y)$	**1/4**	**1/8**	**1/4**	**1/4**	**1/8**

Example1 – Answer (3/3):

x	1	1.5	1.5	2.5	3
y	1	2	3	4	5
$f_{XY}(x, y)$	**1/4**	**1/8**	**1/4**	**1/4**	**1/8**

x	**1**	**1.5**	**2.5**	**3**
$f_X(x)$	**1/4**	**3/8**	**1/4**	**1/8**

y	**1**	**2**	**3**	**4**	**5**
$f_Y(y)$	**1/4**	**1/8**	**1/4**	**1/4**	**1/8**

$$f_{XY}(1,1) = \frac{1}{4} \quad \neq \quad f_X(1) f_Y(1) = \frac{1}{4} * \frac{1}{4} = \frac{1}{16}$$

Then X and Y are not independent.

Linear relationship

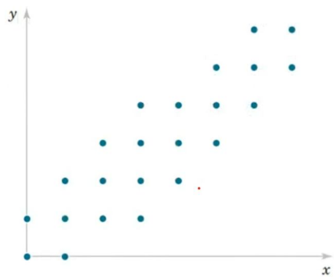
Covariance (1/5):

The covariance between the random variables X and Y is the measure of *linear relationship* between them, denoted as $cov(X, Y)$ or σ_{XY}, where

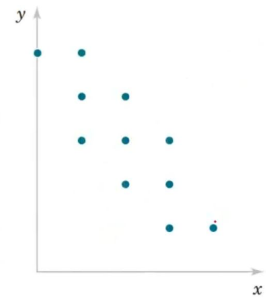
$$cov(X, Y) = \sigma_{XY} = E\Big(\big(X - E(X)\big)\big(Y - E(Y)\big)\Big)$$
$$\boxed{= E(XY) - E(X)E(Y)}$$

Covariance measures the total variation of two random variables from their expected values. Using covariance, we can only standard the direction of the relationship. However, it does not indicate the strength of the relationship.

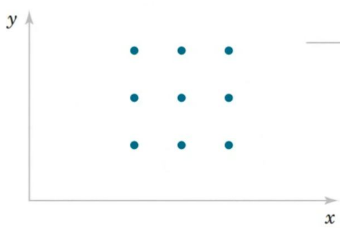
Positive covariance

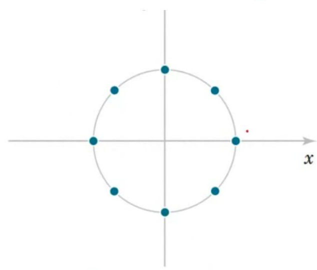


Negative covariance



Zero covariance





Example1:

$f_{XY}(x,y)$

y \ x	1	2	4
3	1/8	0	0
4	1/4	0	0
5	0	1/2	0
6	0	0	1/8

Determine the covariance $\boldsymbol{\sigma_{XY}}$?

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

Example1:

$f_{XY}(x,y)$

y \ x	1	2	4
3	1/8	0	0
4	1/4	0	0
5	0	1/2	0
6	0	0	1/8

Same as

x	1	1	2	4
y	3	4	5	6
$f_{XY}(x,y)$	**1/8**	**1/4**	**1/2**	**1/8**

Example1 – Answer (1/6):

$f_{XY}(x,y)$

x	1	1	2	4
y	3	4	5	6
$f_{XY}(x,y)$	**1/8**	**1/4**	**1/2**	**1/8**

x	y	$f_{XY}(x,y)$	$x\,f_{XY}(x,y)$	$y\,f_{XY}(x,y)$	$xy\,f_{XY}(x,y)$
1	3	1/8			
1	4	1/4			
2	5	1/2			
4	6	1/8			

x	y	$f_{XY}(x,y)$	$x\,f_{XY}(x,y)$	$y\,f_{XY}(x,y)$	$xy\,f_{XY}(x,y)$
1	3	1/8	1/8	3/8	3/8
1	4	1/4	1/4	4/4	4/4
2	5	1/2	2/2	5/2	10/2
4	6	1/8	4/8	6/8	24/8
Sum			**15/8**	**37/8**	**75/8**
			$E(X)$	$E(Y)$	$E(XY)$

$E(X) = 15/8$ $E(Y) = 37/8$ $E(XY) = 75/8$

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

$$cov(X,Y) = \sigma_{XY} = \frac{75}{8} - \frac{(15)(37)}{64} = 0.703125$$

Positive covariance

Correlation (1/2):

The correlation between the random variables X and Y is just scales the covariance by the product of the standard deviation of each variable. Correlation measures the strength of the relationship between variables and denoted as ρ_{XY}, where

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

E(XY) - E(X)E(Y)

$\sqrt{V(X)}$ $\sqrt{V(Y)}$

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

Correlation (2/2):

If X and Y are independent random variables,

$$\rho_{XY} = \sigma_{XY} = 0$$

However, if the correlation between two random variables is zero, we cannot conclude that the random variables are independent.

Example2:

$f_{XY}(x,y)$

y \ x	1	2	4
3	1/8	0	0
4	1/4	0	0
5	0	1/2	0
6	0	0	1/8

Determine the correlation ρ_{XY}?

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

(From the previous example)

Example2 – Answer (2/4):

New

x	y	$f_{XY}(x,y)$	$x\,f_{XY}(x,y)$	$y\,f_{XY}(x,y)$	$xy\,f_{XY}(x,y)$	$x^2\,f_{XY}(x,y)$	$y^2\,f_{XY}(x,y)$
1	3	1/8	1/8	3/8	3/8	1/8	9/8
1	4	1/4	1/4	4/4	4/4	1/4	16/4
2	5	1/2	2/2	5/2	10/2	4/2	25/2
4	6	1/8	4/8	6/8	24/8	16/8	36/8
Sum			**15/8**	**37/8**	**75/8**	**35/8**	**177/8**
			$E(X)$	$E(Y)$	$E(XY)$	$E(X^2)$	$E(Y^2)$

$E(X) = 15/8$ $E(Y) = 37/8$ $E(XY) = 75/8$

$E(X^2) = 35/8$ $E(Y^2) = 177/8$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{E(X^2) - \left(E(X)\right)^2} = \sqrt{0.859375}$$

$$\sigma_Y = \sqrt{V(Y)} = \sqrt{E(Y^2) - \left(E(Y)\right)^2} = \sqrt{0.734375}$$

$$\sigma_{XY} = E(XY) - E(X)E(Y) = 0.703125$$

$$\rho_{XY} = \frac{0.703125}{\sqrt{(0.859375)(0.734375)}} = 0.885079$$

Joint Probability Density Function:

If X and Y are two continuous random variables, the joint probability density fun. is denoted as $f_{XY}(x, y)$, satisfies

1) $f_{XY}(x, y) \geq 0$ for all x, y

2) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\ dxdy = 1$

3) For any region R of two−dimensional space,

$$P\big((X, Y) \in R\big) = \int\int_{R} f_{XY}(x, y)\ dxdy$$

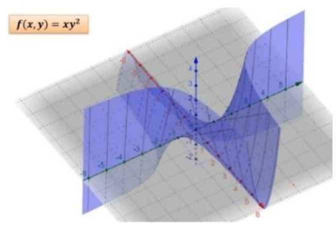
Double Integration:

$$\int_{0}^{1} \int_{1}^{2} \underbrace{xy^2}_{f(x,y)}\ dxdy$$

Same as:

$$\iint_{R} xy^2\ dxdy$$

Where $R = \{(x, y) \mid 1 < x < 2,\ \ 0 < y < 1\}$



$$\int_{1}^{2} xy^2\ dx = \frac{x^2}{2} y^2 \Bigg|_{x=1}^{x=2} = \left(\frac{4}{2}y^2\right) - \left(\frac{1}{2}y^2\right) = \boxed{\frac{3}{2}y^2}$$

$$\int_{0}^{1} \frac{3}{2}y^2 dy = \frac{y^3}{2} \Bigg|_{y=0}^{y=1} = \left(\frac{1}{2}\right) - (0) = \frac{1}{2}$$

Example1:

Let X and Y are two continuous random variables with a joint probability distribution $f_{XY}(x, y)$, where

$$f_{XY}(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Verify that $f_{XY}(x, y)$ is a joint density function.

b) Find $P[(X, Y) \in A]$,

where $A = \{(x, y) \mid 0 < x < \frac{1}{2},\ \frac{1}{4} < y < \frac{1}{2}\}$

$$\int_{0}^{1} \int_{0}^{1} \frac{2}{5}(2x + 3y)\ dxdy$$

$$\int_{0}^{1} \left[\frac{2}{5}(x^2 + 3xy)\right]_{x=0}^{x=1} dy$$

$$\int_{0}^{1} \frac{2}{5}[(1 + 3y) - (0)]\ dy = \int_{0}^{1} \frac{2}{5} + \frac{6y}{5}\ dy = \left[\frac{2}{5}y + \frac{3y^2}{5}\right]_{y=0}^{y=1}$$

$$\left[\frac{2}{5}y+\frac{3y^2}{5}\right]_{y=0}^{y=1}=\left(\frac{2}{5}+\frac{3}{5}\right)-(0)=\frac{5}{5}=1$$

$$\therefore \int_0^1\int_0^1 f_{XY}(x,y)\ dxdy=1$$ is a joint density function

Find $P[(X,Y)\in A]$,

where $A=\{(x,y)\mid 0<x<\frac{1}{2},\ \frac{1}{4}<y<\frac{1}{2}\}$

$$\int_{1/4}^{1/2}\int_0^{1/2}\frac{2}{5}(2x+3y)\ dxdy$$

$$\int_{1/4}^{1/2}\left[\frac{2}{5}(x^2+3xy)\right]_{x=0}^{x=1/2} dy=\int_{1/4}^{1/2}\frac{2}{5}\left(\frac{1}{4}+\frac{3}{2}y\right)-0\ dy$$

$$\int_{1/4}^{1/2}\left(\frac{1}{10}+\frac{3}{5}y\right)dy=\left[\frac{1}{10}y+\frac{3}{10}y^2\right]_{y=1/4}^{y=1/2}$$

$$=\left(\frac{1}{20}+\frac{3}{40}\right)-\left(\frac{1}{40}+\frac{3}{160}\right)=0.125-0.04375=\boxed{0.08125}$$

Example2:

Determine the value of $\boldsymbol{c}$ that makes the function $f_{XY}(x,y)=\boldsymbol{c}(x+y)$ a joint probability density function over the range $0<x<3$ and $x<y<x+2$.

$$\int_0^3\int_x^{x+2}\boldsymbol{c}(x+y)\ dydx=1$$

$$\int_0^3\left[\boldsymbol{c}\left(xy+\frac{y^2}{2}\right)\right]_{y=x}^{y=x+2} dx$$

$$\int_0^3\boldsymbol{c}\left[\left(x^2+2x+\frac{(x+2)^2}{2}\right)-\left(x^2+\frac{x^2}{2}\right)\right]dx$$

$$\int_0^3\boldsymbol{c}(4x+2)\ dx=[\boldsymbol{c}(2x^2+2x)]_{x=0}^{x=3}$$

$$[\boldsymbol{c}(2x^2+2x)]_{x=0}^{x=3}=\boldsymbol{c}[(18+6)-(0)]=24\,\boldsymbol{c}$$

$$24\,\boldsymbol{c}=1\quad\boxed{\therefore \boldsymbol{c}=\frac{1}{24}}$$

Marginal Probability Distributions (Continuous):

The marginal distributions of the random variable X alone is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dy$$

The marginal distributions of the random variable Y alone is:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dx$$

Example3:

Let X and Y are two continuous random variables with a joint probability distribution $f_{XY}(x,y)$, where

$$f_{XY}(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Find $f_X(x)$
b) Find $f_Y(y)$

Find $f_X(x)$

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dy = \int_0^1 \frac{2}{5}(2x+3y)\ dy$$

$$\int_0^1 \frac{2}{5}(2x+3y)\ dy = \left[\frac{2}{5}\left(2xy + \frac{3y^2}{2}\right)\right]_{y=0}^{y=1} = \left(\frac{4x}{5} + \frac{3}{5}\right) - (0)$$

$$f_X(x) = \frac{4x}{5} + \frac{3}{5}$$

Find $f_Y(y)$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\ dx = \int_0^1 \frac{2}{5}(2x+3y)\ dx$$

$$\int_0^1 \frac{2}{5}(2x+3y)\ dx = \left[\frac{2}{5}(x^2+3xy)\right]_{x=0}^{x=1} = \left(\frac{2}{5} + \frac{6y}{5}\right) - (0)$$

$$f_Y(y) = \frac{2}{5} + \frac{6y}{5}$$

Mean from a Joint Distribution (Continuous):

Mean for the random variable X alone is:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\ dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{XY}(x,y)\ dydx$$

Mean for the random variable Y alone is:

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x,y)\ dydx$$

Variance from a Joint Distribution (Continuous):

Variance for the random variable X alone is:

$$V(X) = \boxed{E(X^2)} - \left(E(X)\right)^2$$

and s.d. of $X = \sigma_X = \sqrt{V(X)}$

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

Variance for the random variable Y alone is:

$$V(Y) = \boxed{E(Y^2)} - \left(E(Y)\right)^2$$

and s.d. of $Y = \sigma_Y = \sqrt{V(Y)}$

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 f_Y(y)\ dy$$

Example4:

Let X and Y are two continuous random variables with a jc
probability distribution $f_{XY}(x,y)$, where

$$f_{XY}(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

a) Find $E(X)$

b) Find $E(Y)$

c) Find σ_X

Example4 – Answer (1/4):

$$f_X(x) = \frac{4x}{5} + \frac{3}{5}$$

Find $E(X)$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\ dx = \int_0^1 \frac{4x^2}{5} + \frac{3x}{5}\ dx$$

$$\int_0^1 \frac{4x^2}{5} + \frac{3x}{5}\ dx = \left[\frac{4x^3}{15} + \frac{3x^2}{10}\right]_{x=0}^{x=1} = \left(\frac{4}{15} + \frac{3}{10}\right) - (0)$$

Example4 – Answer (2/4):

$$\boxed{f_Y(y) = \frac{2}{5} + \frac{6y}{5}}$$

Find $E(Y)$

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\ dy = \int_0^1 \frac{2y}{5} + \frac{6y^2}{5}\ dy$$

$$\int_0^1 \frac{2y}{5} + \frac{6y^2}{5}\ dy = \left[\frac{y^2}{5} + \frac{2y^3}{5}\right]_{y=0}^{y=1} = \left(\frac{1}{5} + \frac{2}{5}\right) - (0)$$

$$E(Y) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5} = 0.6$$

Example4 – Answer (3/4):

$$\boxed{E(X) = 0.5667}$$

Find σ_X

$$\sigma_X = \sqrt{V(X)} = \sqrt{E(X^2) - \left(E(X)\right)^2}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\ dx = \int_0^1 \frac{4x^3}{5} + \frac{3x^2}{5}\ dx$$

$$E(X^2) = \left[\frac{x^4}{5} + \frac{x^3}{5}\right]_{x=0}^{x=1} = \left(\frac{1}{5} + \frac{1}{5}\right) - (0) = \frac{2}{5} = 0.4$$

Example4 – Answer (4/4):

$$\boxed{E(X) = 0.5667}$$

$$\boxed{E(X^2) = 0.4}$$

Find σ_X

$$\sigma_X = \sqrt{V(X)} = \sqrt{E(X^2) - \left(E(X)\right)^2}$$

$$\sigma_X = \sqrt{V(X)} = \sqrt{0.4 - (0.5667)^2} = \sqrt{0.07885111}$$

$$\sigma_X = 0.2808044$$

Probability distributions:

- Discrete Uniform Distribution.
- Binomial Distribution
- Continuous Uniform Distribution
- Normal Distribution.

Discrete Uniform Distribution.

Definition:

A random variable X has a discrete uniform distribution if each of the n values in its range, $x_1, x_2, \dots, x_n$, has equal probability. Then

$$f(x_i) = \frac{1}{n}$$

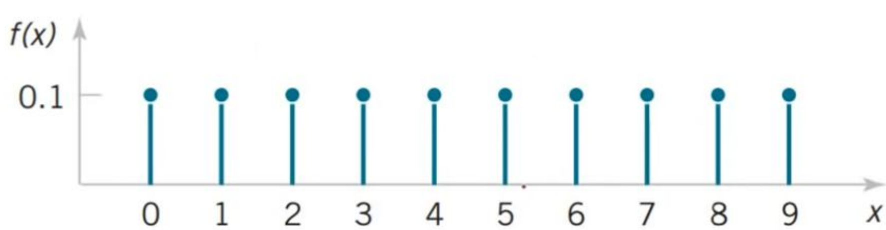
Example1 (1/2):

The first digit of a part's serial number is equally likely to be any one of the digits 0 through 9. If one part is selected randomly from a large batch and X is the first digit of the serial number, X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \dots, 9\}$. That is,

$$f(x) = \frac{1}{10} = 0.1$$

Example1 (2/2):

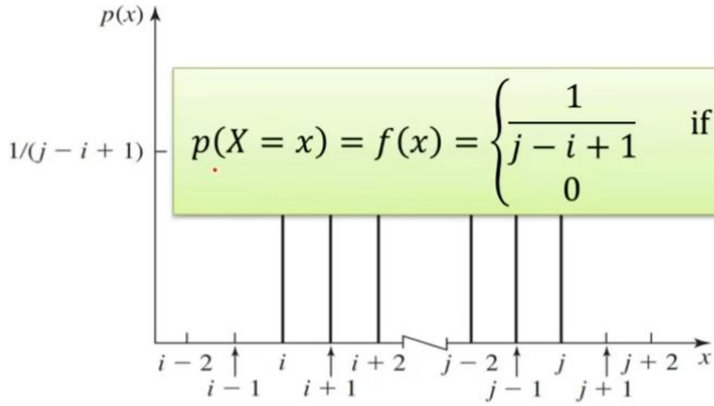
Probability mass function for a discrete uniform random variable



Discrete Uniform $DU(i, j)$:

$X \sim \mathrm{DU}(i, j)$

Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \dots, j$, for $i \le j$

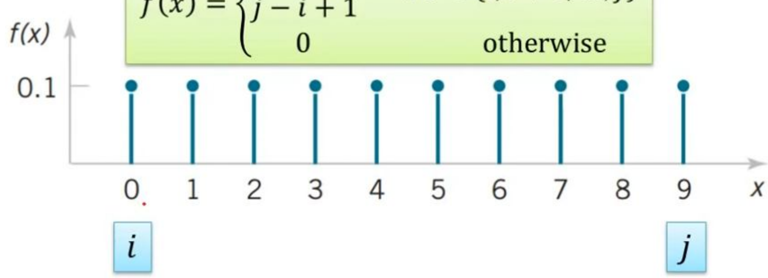


$$p(X = x) = f(x) = \begin{cases} \dfrac{1}{j - i + 1} & \text{if } x \in \{i, i+1, \dots, j\} \\ 0 & \text{otherwise} \end{cases}$$

Recall Example1:

Probability mass function for a discrete uniform random variable

$$f(x) = \begin{cases} \dfrac{1}{j - i + 1} & \text{if } x \in \{i, i+1, \dots, j\} \\ 0 & \text{otherwise} \end{cases}$$



$$f(x) = \begin{cases} \dfrac{1}{9 - 0 + 1} & \text{if } x \in \{0, 1, \dots, 9\} \\ 0 & \text{otherwise} \end{cases}$$

$$f(x) = \begin{cases} 1/10 & \text{if } x \in \{0, 1, \dots, 9\} \\ 0 & \text{otherwise} \end{cases}$$

Mean and Variance:

Suppose that the range of the discrete random variable X equals the consecutive integers $i, i+1, \dots, j$, for $i \le j$

$$\mu = E(X) = \frac{j + i}{2}$$

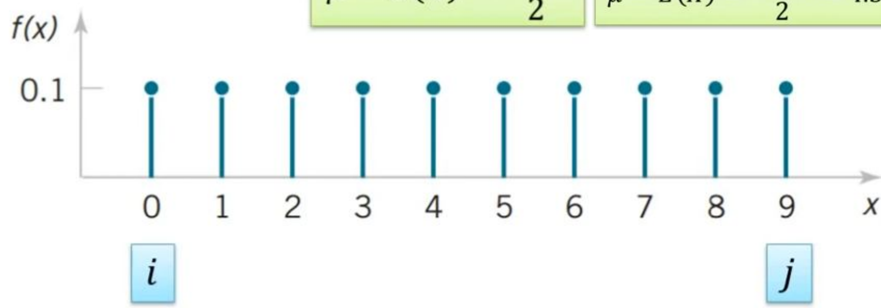
$$\sigma^2 = V(X) = \frac{(j - i + 1)^2 - 1}{12}$$

Example2 (1/4):

$f(x)$ is a probability mass function for a discrete uniform random variable X

$$\mu = E(X) = \frac{j+i}{2}$$

$$\mu = E(X) = \frac{9+0}{2} = 4.5$$



$$\sigma^2 = V(X) = \frac{(j-i+1)^2 - 1}{12}$$

$$\sigma^2 = V(X) = \frac{(9-0+1)^2 - 1}{12} = 8.25$$

The Bernoulli Process:

1. The experiment consists of ***repeated*** trials and each trial is called a ***Bernoulli trial***.
2. Each trial results in an outcome that may be classified as a ***success*** or a ***failure***.
3. The probability of success, denoted by $\boldsymbol{p}$, remains constant from trial to trial.
4. The repeated trials are ***independent***.

The Bernoulli Process: (Examples)

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A multiple-choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
3. In the next 20 births at a hospital, let X = the number of female births.

Binomial Distribution $b(n, p)$:

$X \sim b(n,p)$ $X \sim \text{Bin}(n,p)$

A random experiment consists of n Bernoulli trials. The random variable X that equals the number of trials that result in a success is a *binomial random variable* with parameters $0 < p < 1$ and finite $n = 1, 2, \ldots$.

The probability mass function of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

Example 1 (2/5):

Flip a coin 10 times, what is the probability that 3 heads occurs?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$f(3) = \binom{10}{3} (0.5)^3 (0.5)^{10-3}$$

$$= 120\,(0.5)^3 (0.5)^7$$

$$= 0.1171875$$

Example 2 (1/5): $X \sim \text{Bin}(10, 0.5)$

Flip a coin 10 times, what is the probability that:

a) No head occurs.

b) At least 2 heads occurs.

a) No head occurs.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$f(0) = \binom{10}{0} (0.5)^0 (0.5)^{10-0}$$

b) At least 2 heads occurs.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(x) = \binom{10}{x} (0.5)^x (0.5)^{10-x}$$

$$P(X \geq 2) = \cancel{f(2) + f(3) + \cdots + f(10)}$$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [f(1) + f(0)]$$

$$f(0) = \binom{10}{0} (0.5)^0 (0.5)^{10-0}$$

$$f(1) = \binom{10}{1} (0.5)^1 (0.5)^{10-1}$$

Mean and Variance $b(n, p)$:

If X is a binomial random variable with parameters p and n

$$\mu = E(X) = np$$

$$\sigma^2 = V(X) = np(1-p)$$

Example 3 (5/5):

Flip a coin 10 times. Let X = number of heads obtained, what are the expected value and the variance of X?

- Bernoulli trial → Flip a coin
- Number of trials ($n = 10$)
- Success → Head occur ($p = 0.5$)
- Let X = number of heads obtained
  - Values $x = 0, 1, 2, \ldots, 10$

$$\mu = E(X) = np = 10 \times 0.5 = 5$$

$$\sigma^2 = V(X) = np(1-p) = 10 \times 0.5 \times 0.5 = 2.5$$

Example 4 (2/5):

The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) exactly 5 survive, and (b) at least 3 survive?

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
    - Values $x = 0, 1, 2, \dots, 15$

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(x) = \binom{15}{x} (0.4)^x (0.6)^{15-x}$$

(a) exactly 5 survive $x = 5$

$$f(5) = \binom{15}{5} (0.4)^5 (0.6)^{15-5}$$

$$= 0.1859$$

(b) at least 3 survive

$$P(X \geq 3) = 1 - P(X < 3)$$
$$= 1 - [f(2) + f(1) + f(0)]$$

Example 5 (2/4):

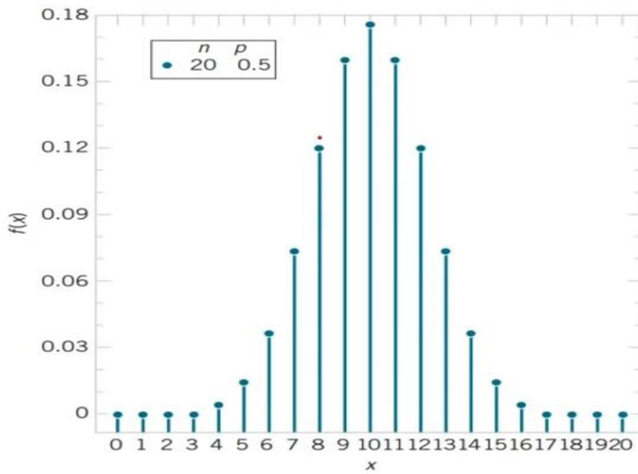
The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what are the expected value and the variance of the survive?

- Bernoulli trial → Patient
- Number of trials ($n = 15$)
- Success → Patient recover ($p = 0.4$)
- Let X = the number of people who survive
    - Values $x = 0, 1, 2, \dots, 15$

$$\mu = E(X) = np$$
$$= 15 \times 0.4 = 6$$

$$\sigma^2 = V(X) = np(1-p)$$
$$= 15 \times 0.4 \times 0.6$$
$$= 3.6$$

Binomial distribution for selected values of $n = 20$ and $p = 0.5$



Example 6 (2/4):

$X \sim \text{Bin}(10, 0.8)$

The probability that a patient recovers from a COVID-19 virus is 0.8. What is the probability that at least 3 of the next 10 patients having this virus will survive?

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- Bernoulli trial → Patient
- Number of trials ($n = 10$)
- Success → Patient recover ($p = 0.8$)
- Let X = the number of people who survive
  - Values $x = 0, 1, 2, \ldots, 10$

$$f(x) = \binom{10}{x} (0.8)^x (0.2)^{10-x}$$

$$P(X \geq 3) = 1 - P(X \leq 2)$$
$$= 1 - [f(2) + f(1) + f(0)]$$

Example 7 (3/4):

$X \sim \text{Bin}(9, 0.25)$

A traffic control engineer reports that 75% of the vehicles passing through a checkpoint are from within the state. What is the probability that fewer than 4 of the next 9 vehicles are from out of state?

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(x) = \binom{9}{x} (0.25)^x (0.75)^{9-x}$$

- Bernoulli trial → Vehicle
- Number of trials ($n = 9$)
- Success → vehicle is from out of state ($p = 0.25$)
- Let X = the number of vehicles are from out of state?
    - Values $x = 0, 1, 2, \dots, 9$

$P(X < 4) = P(X \le 3)$

$P(X \le 3)$, $n = 9$, $p = 0.25$

You can find $P(X \le 3)$ from the **Table**.

$P(X \le 3) = 0.8343$

Continuous Uniform Distribution

A continuous random variable X with probability density function

$$f(x) = 1/(b-a), \quad a \le x \le b$$

is a **continuous uniform random variable**.

Mean and Variance

If X is a continuous uniform random variable over $a \le x \le b$,

$$\mu = E(X) = \frac{a+b}{2}$$

$$\sigma^2 = V(X) = \frac{(b-a)^2}{12}$$

Normal Distribution

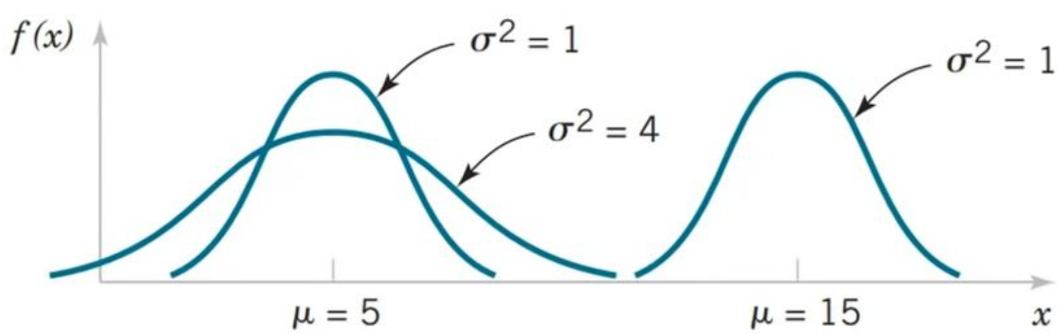
A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

is a **normal random variable** with parameters μ where $-\infty < \mu < \infty$ and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution.



Multinomial Experiments:

The binomial experiment becomes a multinomial experiment if we let each trial have **more than two** possible outcomes.

Multinomial Distribution:

If a given trial can result in the k outcomes $E_1, E_2, \ldots, E_k$ with probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $X_1, X_2, \ldots, X_k$, representing the number of occurrences for $E_1, E_2, \ldots, E_k$ in n independent trials, is

$$f(x_1, x_2, \ldots x_k) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

$$= \frac{n!}{x_1!\,x_2!\cdots x_k!}$$

with

$$\sum_{i=1}^{k} x_i = n \quad \text{and} \quad \sum_{i=1}^{k} P_i = 1$$

Multinomial Distribution – Example 1 (1/5):

The complexity of arrivals and departures of planes at an airport is such that computer simulation is often used to model the "ideal" conditions. For a certain airport with three runways, it is known that in the ideal setting the following are the probabilities that the individual runways are accessed by a randomly arriving commercial jet:

Runway1: $p_1 = 2/9$,
Runway2: $p_2 = 1/6$,
Runway3: $p_3 = 11/18$.

$$f(x_1, x_2, x_3) = \binom{n}{x_1, x_2, x_3} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$$f(2,1,3) = \binom{6}{2,1,3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3$$

What is the probability that 6 randomly arriving airplanes are distributed in the following fashion?

Runway1: 2 airplanes,
Runway2: 1 airplane,
Runway3: 3 airplanes.

$$x_1 = 2$$
$$x_2 = 1$$
$$x_3 = 3$$

If p1+p2+p3 =1 then this is a multinomial distribution.

$$f(2,1,3) = \binom{6}{2,1,3} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3$$

$$= \frac{6!}{2!\,1!\,3!} \left(\frac{2}{9}\right)^2 \left(\frac{1}{6}\right)^1 \left(\frac{11}{18}\right)^3 = 0.1127$$

Multinomial Distribution – Example 2 (1/4):

The probabilities are 0.4, 0.2, 0.3, and 0.1, respectively, that a delegate to a certain convention arrived by air, bus, automobile, or train. What is the probability that among 9 delegates randomly selected at this convention, 3 arrived by air, 3 arrived by bus, 1 arrived by automobile, and 2 arrived by train.

$p_1 = 0.4$
$p_2 = 0.2$
$p_3 = 0.3$
$p_4 = 0.1$

$x_1 = 3$
$x_2 = 3$
$x_3 = 1$
$x_4 = 2$

$$f(x_1, x_2, x_3, x_4) = \binom{n}{x_1, x_2, x_3, x_4} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4}$$

$$f(3,3,1,2) = \binom{9}{3,3,1,2} (0.4)^3 (0.2)^3 (0.3)^1 (0.1)^2$$

$$f(3,3,1,2) = \binom{9}{3,3,1,2} (0.4)^3 (0.2)^3 (0.3)^1 (0.1)^2$$

$$= \frac{9!}{3!\,3!\,1!\,2!} (0.4)^3 (0.2)^3 (0.3)^1 (0.1)^2 = 0.0077$$

Negative Binomial Distribution:

$X \sim nb(k, p)$

If repeated independent trials can result in a success with probability p, then the probability distribution of the random variable X, the number of the trial on which the kth success occurs, is

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$$

with

$$k = 1.2. \ldots \quad \text{and} \quad x = k.k+1.k+2. \ldots$$

Negative Binomial Distribution – Example (3/8):

In an NBA (National Basketball Association) championship series, the team that wins **four games** out of seven is the winner. Suppose that teams A and B face each other in the championship games and that **team A has probability 0.55 of winning a game over team B**.

What is the probability that team A will win the series in $\boxed{6}$ **games**?

$p = 0.55$, $x = 6$, $k = 4$ → $x = 4, 5, \boxed{6}, 7$

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$$

$$f(6) = \binom{5}{3} (0.55)^4 (1-0.55)^{6-4} = 0.1853$$

What is the probability that team A will win the series?

$p = 0.55,$

1) $x = 4, \quad k = 4$
2) $x = 5, \quad k = 4$
3) $x = 6, \quad k = 4$
4) $x = 7, \quad k = 4$

$$f(4) + f(5) + f(6) + f(7) = 0.6083$$

$$f(4) = \binom{3}{3} (0.55)^4 (1-0.55)^{4-4} = 0.0915$$

$$f(5) = \binom{4}{3} (0.55)^4 (1-0.55)^{5-4} = 0.1647$$

$$f(6) = \binom{5}{3} (0.55)^4 (1-0.55)^{6-4} = 0.1853$$

$$f(7) = \binom{6}{3} (0.55)^4 (1-0.55)^{7-4} = 0.1668$$

Negative Binomial Distribution: (Mean and Variance)

If X is a negative binomial random variable with parameters p and k,

$$\mu = E(X) = k/p \text{ and}$$

$$\sigma^2 = V(X) = k(1-p)/p^2$$

Geometric Distribution:

$$\boxed{X \sim g(p)}$$

If repeated independent trials can result in a success with probability p, then the probability distribution of the random variable X, the number of the trial on which the **first** success occurs, is

$$f(x) = p(1-p)^{x-1}$$

with

$$x = 1, 2, 3, \ldots$$

Geometric Distribution – Example (3/5):

For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the **fifth item inspected** is the **first** defective item found?

$p = 0.01, \quad x = 5, \quad k = 1$

$f(x) = p(1-p)^{x-1}$

$f(5) = (0.01)(1-0.01)^{5-1} = 0.0096$

Geometric Distribution: (Mean and Variance)

If X is a geometric random variable with parameter p,

$$\mu = E(X) = 1/p \text{ and}$$

$$\sigma^2 = V(X) = (1-p)/p^2$$

Poisson Distribution:

$X \sim p(\lambda t)$ | $X \sim \text{Poisson}(\lambda t)$

The mean number of outcomes is computed from $\mu = \lambda t$, where t is the specific "time," "distance," "area," or "volume" of interest. Since the probabilities depend on λ, the rate of occurrence of outcomes.

$$f(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

with

$$\lambda > 0 \text{ and } \quad x = 0, 1, 2, \ldots$$

Poisson Distribution – Example 1 (1/4):

During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

$\lambda t = 4$
$x = 6$

$$f(6) = \frac{e^{-4}(4)^6}{6!} = 0.1042$$

Poisson Distribution – Example 2 (1/6):

Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

$\lambda = 10,\ x > 15$

$\lambda = 10,\ x > 15$

$$f(x) = \frac{e^{-\lambda}(\lambda)^x}{x!}$$

$$f(X > 15) = 1 - \boxed{f(X \leq 15)}$$

$$\boxed{\begin{matrix} f(X=0) \\ f(X=1) \\ f(X=2) \\ \ldots \\ f(X=15) \end{matrix}}$$

$\lambda = 10,\ x > 15$

$$f(X > 15) = 1 - 0.9513 = 0.0487$$

Both the mean and the variance of the Poisson distribution are

$$E(X) = V(X) = \lambda t$$

Poisson Distribution: (From Binomial)

Let X be a binomial random variable with probability distribution $b(n, p)$.

When $n \to \infty$, $p \to 0$,

and $np \xrightarrow{n\to\infty} \mu$ remains constant,

$b(n, p) \xrightarrow{n\to\infty} p(\mu)$.

Poisson Distribution: (From Binomial): Example (2/6)

In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that, on average, **1 in every 1000** of these items produced has one or more bubbles. What is the probability that a random **sample of 8000** will yield **fewer than 7** items possessing bubbles?

$p = 0.001$, and $n = 8000$

$$f(x < 7) = f(x \leq 6) = \sum_{x=0}^{6} Bin(x; n, p)$$

$p = 0.001$, and $n = 8000$

$$f(x < 7) = f(x \leq 6) = \sum_{x=0}^{6} Bin(x; 8000, 0.001)$$

Since p is very close to 0 and n is quite large, we shall approximate with the Poisson distribution using

$$\mu = np = 8$$

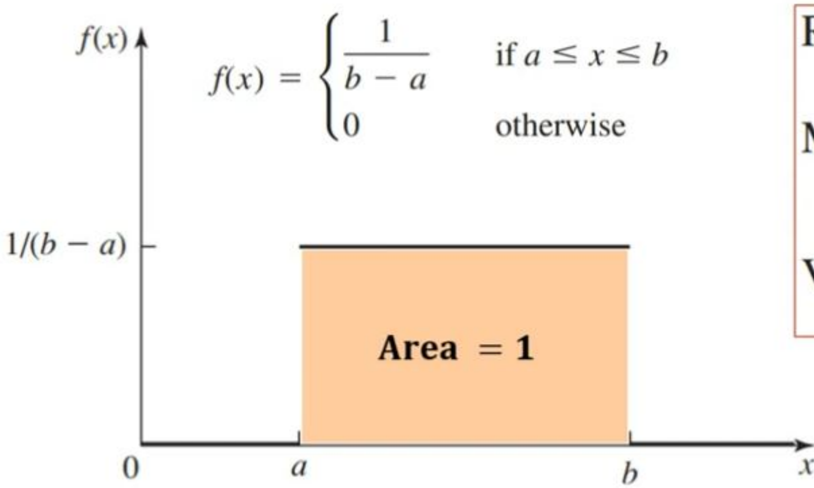
$$\sum_{x=0}^{6} p(x; 8) = 0.3134 \quad \boxed{\text{From the Table}}$$

- Continuous Uniform Distribution.

Probability Density Function:

$\boldsymbol{U(a, b)}$

- a and b real numbers with $a < b$.



Range	$[a, b]$
Mean	$\frac{a+b}{2}$
Variance	$\frac{(b-a)^2}{12}$

Cumulative Distribution Function:

$\boldsymbol{U(a, b)}$

Density $\quad f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$

$$F(x) = \int_a^x \frac{1}{b-a} du = \frac{x-a}{b-a}$$

cumulative distribution function
$P(X \le x)$

Cumulative Distribution Function: $U(a,b)$

$$\text{Density} \qquad f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Distribution} \qquad F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$$

Example1 (1/6):

Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. Both long and short conferences occur quite often. In fact, it can be assumed that the length X of a conference has a uniform distribution on the interval [0, 4].

(a) What is the probability density function?

(b) What is the probability that any given conference lasts at least 3 hours?

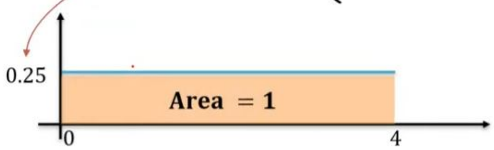
Example1 (2/6):

$\boldsymbol{X \sim U(0, 4)}$

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

a) What is the probability density function?

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \leq x \leq 4, \\ 0, & \text{elsewhere.} \end{cases}$$

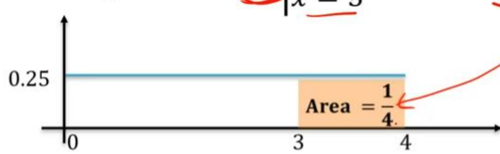


Example1 (4/6):

$\boldsymbol{X \sim U(0, 4)}$

b) What is the probability that any given conference lasts at least 3 hours?

$$P(X \geq 3) = \int_3^4 \frac{1}{4}\,dx = \frac{1}{4}x \Big|_{x=3}^{x=4} = (1) - \left(\frac{3}{4}\right) = \frac{1}{4}$$



Another solution (different way):

Example1 (6/6):

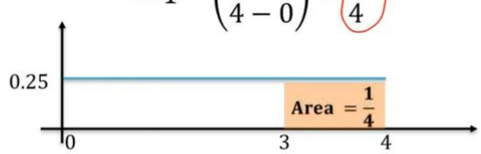
$\boldsymbol{X \sim U(0, 4)}$

$$\text{Distribution} \quad F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } b < x \end{cases}$$

b) What is the probability that any given conference lasts at least 3 hours?

$$P(X \geq 3) = 1 - P(X < 3) = 1 - F(3)$$

$$= 1 - \left(\frac{3-0}{4-0}\right) = \frac{1}{4}$$



Example2 (1/10):

A bus arrives every 10 minutes at a bus stop. It is assumed that the waiting time for a particular individual is a random variable with a continuous uniform distribution.

(a) What is the probability that the individual waits at most 4 minutes?

(b) What is the probability that the individual waits more than 6 minutes (at least 6 minutes)?

(c) What is the probability that the individual waits between 1 and 4 minutes?

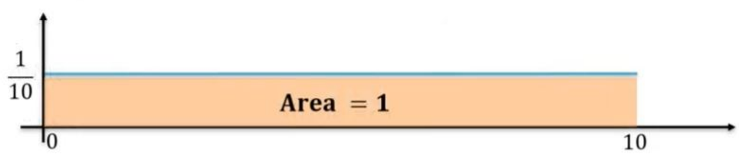
(d) What is the average waiting time?

Example2 (2/10):

$\boldsymbol{X \sim U(0, 10)}$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$f(x) = \begin{cases} \dfrac{1}{10} & \text{if } 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$



Example2 (3/10):

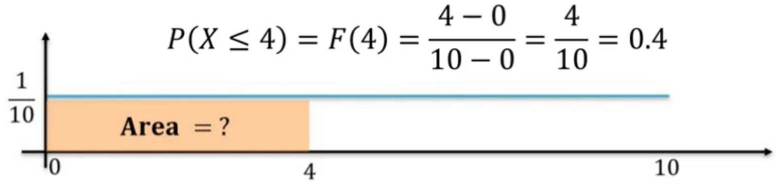
$\boldsymbol{X \sim U(0, 10)}$

Distribution $F(x) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } b < x \end{cases}$

a) What is the probability that the individual waits at most 4 minutes?

$$P(X \le 4) = F(4)$$

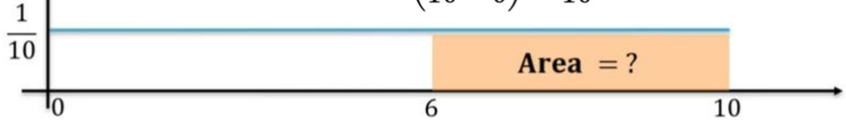
$$P(X \le 4) = F(4) = \frac{4-0}{10-0} = \frac{4}{10} = 0.4$$



b) What is the probability that the individual waits more than 6 minutes (at least 6 minutes)?

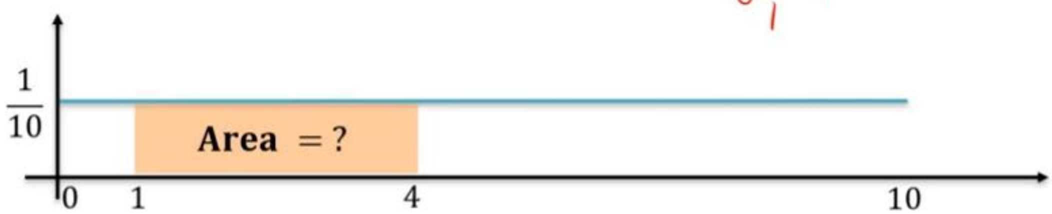
$$P(X \geq 6) = 1 - F(6)$$

$$P(X \geq 6) = 1 - \left(\frac{6-0}{10-0}\right) = \frac{4}{10} = 0.4$$

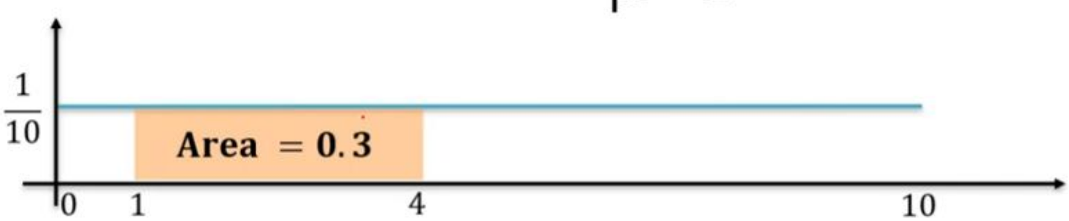


c) What is the probability that the individual waits between 1 and 4 minutes?

$$P(1 \leq X \leq 4) \quad \int_1^4 \frac{1}{10}\, dx$$



$$P(1 \leq X \leq 4) = \int_1^4 \frac{1}{10}\, dx = \frac{1}{10} x \Bigg|_{x=1}^{x=4} = \left(\frac{4}{10}\right) - \left(\frac{1}{10}\right) = \frac{3}{10}$$



Another way using cumulative distribution

$$P(1 \leq X \leq 4) = F(4) - F(1) = \left(\frac{4-0}{10-0}\right) - \left(\frac{1-0}{10-0}\right) = \frac{3}{10}$$

Example2 (10/10):

$X \sim U(0, 10)$

mean $= \boxed{\mu = E(X) = \frac{a+b}{2}}$

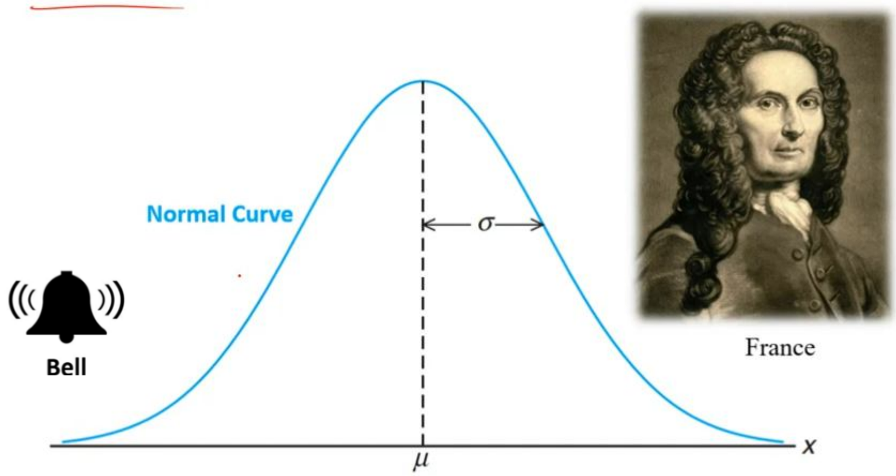
$X \sim U(\underset{a}{0}, \underset{b}{10})$

d) What is the average waiting time?

$$\mu = E(X) = \frac{0+10}{2} = 5 \text{ minutes}$$

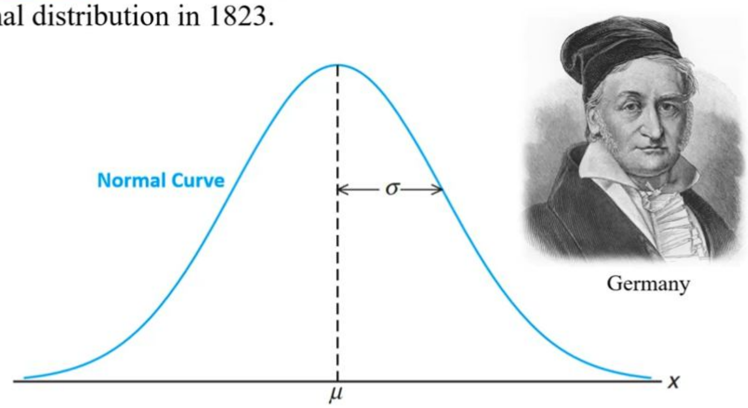
Normal distribution:

In 1733, **Abraham DeMoivre** developed the mathematical equation of the normal curve.



France

The normal distribution is often referred to as the **Gaussian distribution**, in honor of Karl Friedrich Gauss, who discovered the normal distribution in 1823.



Germany

X : continuous random variable

μ : mean

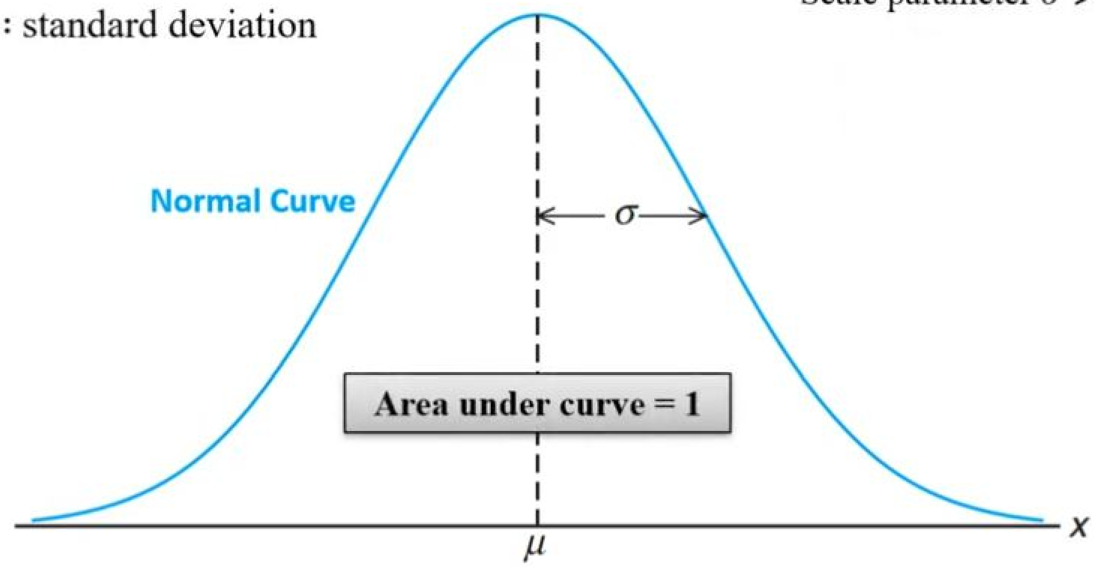
σ^2: variance

σ : standard deviation

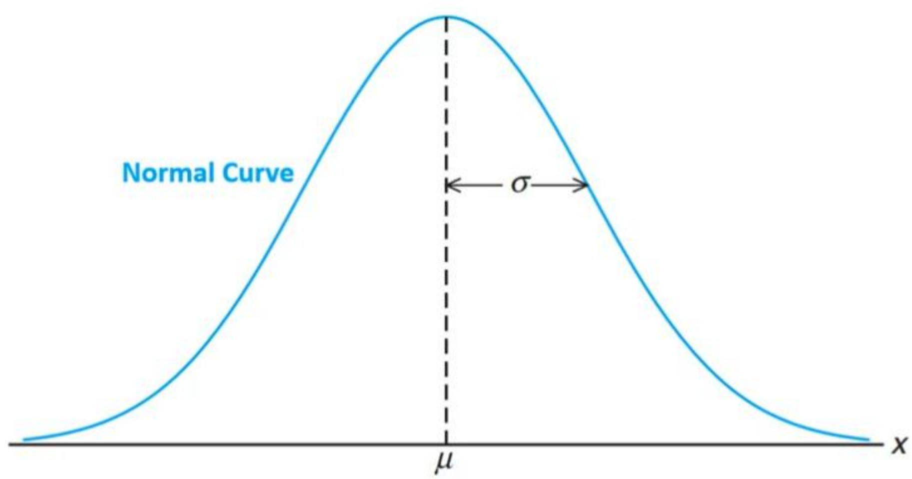
$$\boldsymbol{X \sim N(\mu, \sigma^2)}$$

Location parameter $\mu \in \mathbb{R}$,

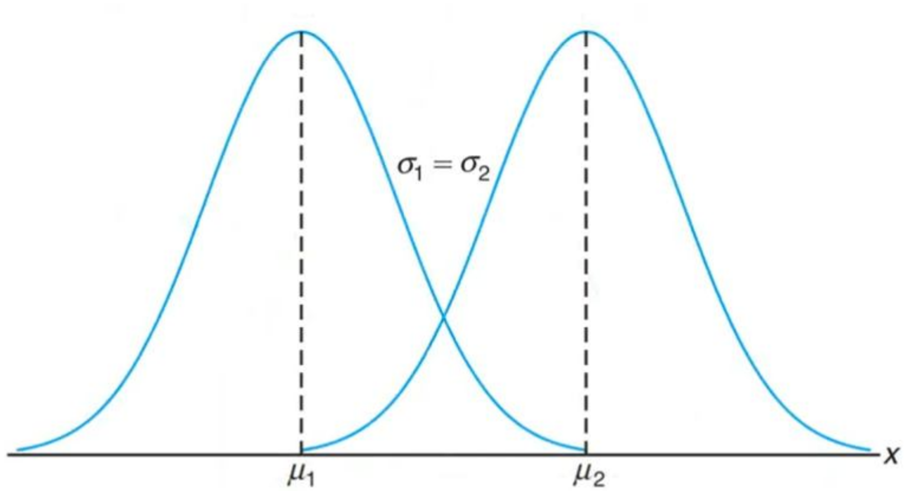
Scale parameter $\sigma > 0$.



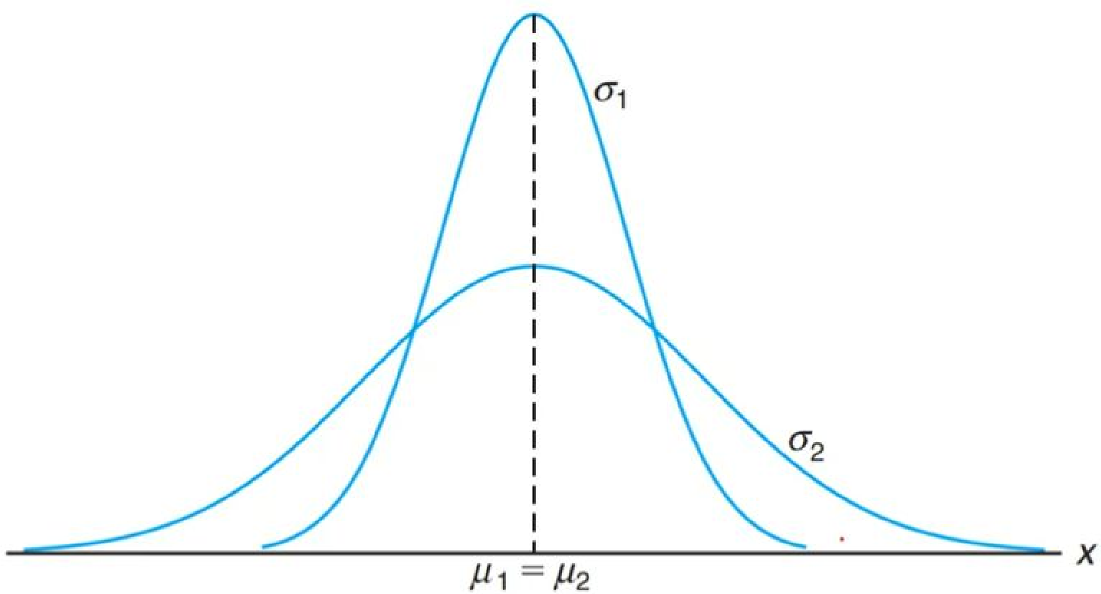
Note that the graph of the normal distribution is ***symmetric*** about its mean μ (i.e., for any $c > 0$, $P(X > \mu + c) = P(X < \mu - c)$, and that σ^2 is a measure of the width of the bell shape.



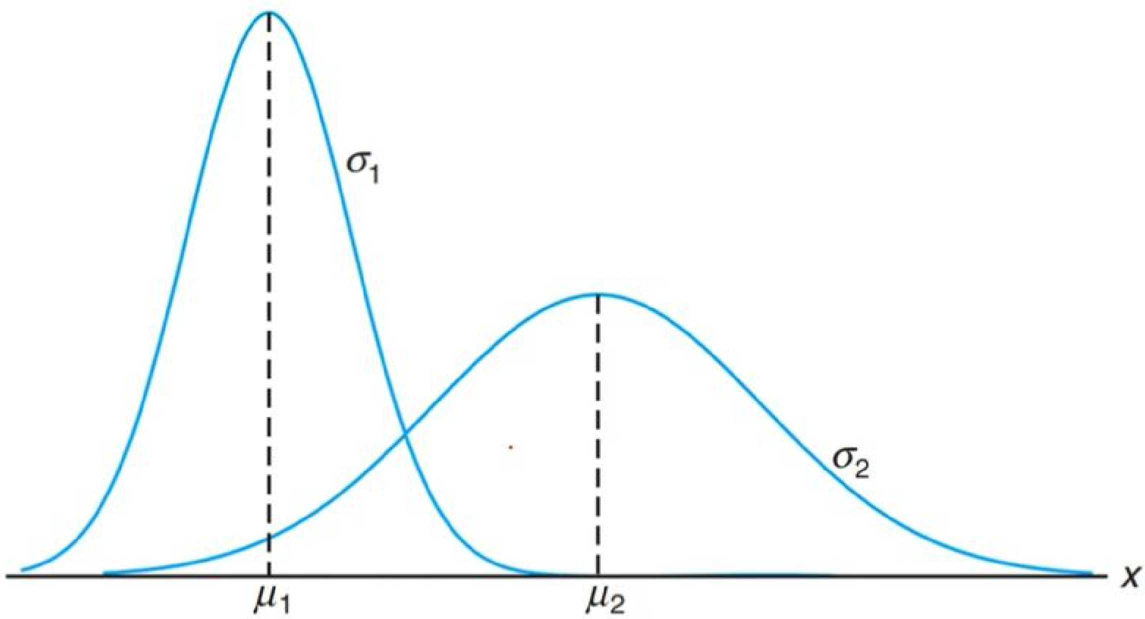
Same standard deviation but different means ($\mu_1 < \mu_2$).



Same mean but different standard deviations ($\sigma_1 < \sigma_2$).

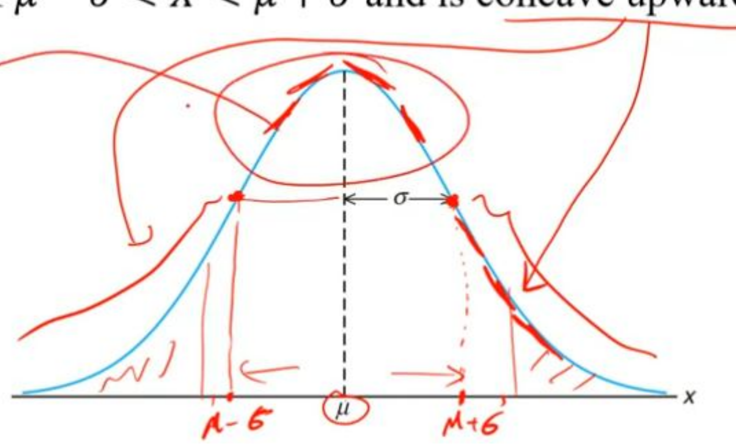


Different means ($\mu_1 < \mu_2$) and standard deviations ($\sigma_1 < \sigma_2$).



Properties of the normal curve (1/2):

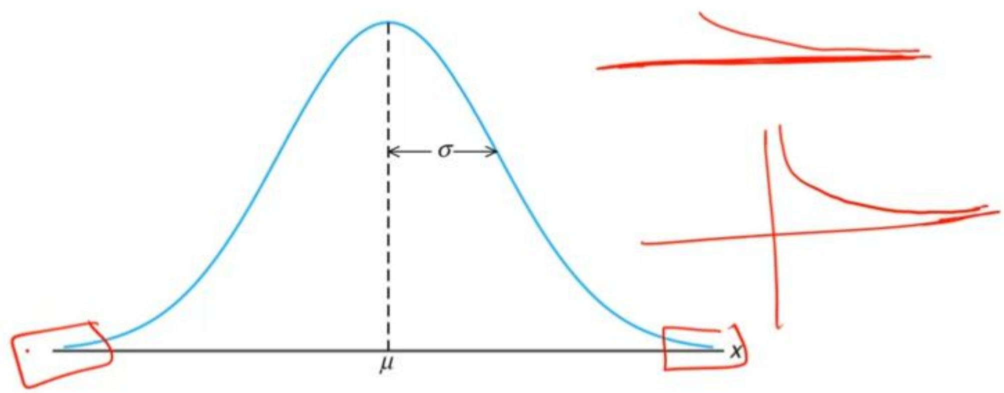
1) Mode = Mean (μ)
2) The curve is symmetric about a vertical axis through the mean (μ).
3) The curve has its points of inflection at $x = \mu \pm \sigma$; it is concave downward if $\mu - \sigma < X < \mu + \sigma$ and is concave upward otherwise.



Concave upward the curve over the tamas line, concave downward: the curve below the tamas line.

Properties of the normal curve (2/2):

4) The normal curve approaches the horizontal axis asymptotically as we proceed in either direction away from the mean.
5) The total area under the curve and above the horizontal axis is equal to 1.



Probability density function of the normal random variable X:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

The mean and variance of X are μ and σ^2, respectively,

where $-\infty < \mu < \infty$ and $\sigma > 0$

Definition:

Probability Density Function PDF of continuous RV of X: (PDF derivative of the CDF)

$$f_X: F_X(x) = \int_{-\infty}^{x} f_x\,(t)dt$$

i.e. $f_X = F'_X$

Note:

- PMF (Probability Mass Function) for discrete RV only.
- PDF (Probability Density Function) for continuous RV only.
- CDF (Cumulative Distribution Function) is always defined

RECAP

- RV is a function: $\Omega \longmapsto \mathbb{R}$
- CDF of a RV. X : $F_X(t) = P(X \leq t)$
- Discrete / Continuous RV
- CDF
- PMF: $f_X(x) = P(X = x)$
- PDF : $f_X = F'_X$
- $X \sim Y$ iff $F_X = F_Y$

Example

Tossing a coin for a head

- Let p = probability of Head in any given coin toss
- X = number of tosses required before getting Head
- We have:

$$P(X = x) = (1 - p)^{x-1} \times p$$

  Tail probability (1-p), Head probability (p) (exactly x-1 tails followed by 1 head)
- We can write the PMF of X as follow:

$$F_X(x) = P(X = x) = \begin{cases} (1-p)^{x-1} \times p, & x = 1,2,.. \\ 0 & ,otherwise \end{cases}$$

  We just described the geometric distribution (describe when repeated a trial number of time till you observe a particular event you are waiting for, p is the probability of observing each single time)

Example

Logistic distribution

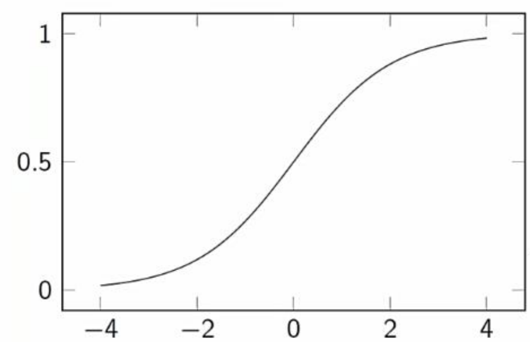
- For continuous RVs, it is sometimes more convenient to specify directly the CDF
- Consider $F_X(x) = \frac{1}{1+e^{-x}}$
- We can derive the pdf:

$$f_X(x) = \frac{d}{dx}F_X(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

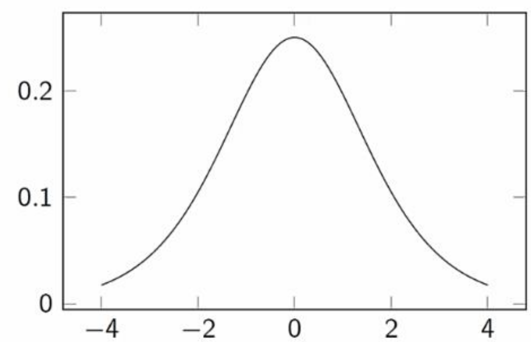
- Both CDF and pdf can be used to compute probabilities over intervals:

$$P(a < X < b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx$$
$$= \int_{a}^{b} f_X(x)dx$$

cdf: $F_X(x) = \frac{1}{1+e^{-x}}$

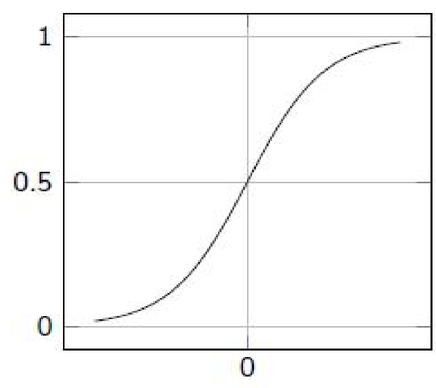


pdf: $f_X(x) = \frac{e^{-x}}{(1+e^{-x})^2}$

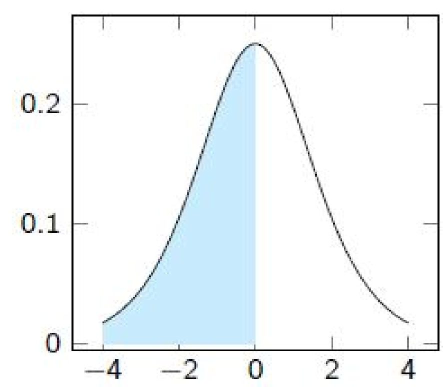


$P(X \leq 0) =?$

$P(X \leq 0) = F_X(0) = \frac{1}{1+e^{-0}} = 1/2$



$P(X \leq 0) = \int_{-\infty}^{0} \frac{e^{-x}}{(1+e^{-x})^2} = 1/2$



You can use F_X to answer the *inverse* question: 'What's the value x such that $P(X \leq x) = 1/2$?'

$x : F_X(x) = 1/2 \implies F_X^{-1}(1/2) = 0$

F_X^{-1} is commonly known as the quantile function

Example: Exam simulation

We throw a regular dice, consider the R.V X= “the score after throw the dice)

Q1: is X a valid R.V? (Y or N or no enough data) (yes valid R.V)

Q2: CDF of X? Q3: PMF of X?

Q1: valid Numbers

Q2 : PMF : $f_X(x) = P(X = x)$

PMF= $f_X(x) = \begin{cases} 1/6 & \text{if } x = 1,2,3,4,5,6 \\ 0 & otherwise \end{cases}$

x	0	1	2	3	4	5	6
P(x)	0	1/6	1/6	1/6	1/6	1/6	1/6
CDF=$F_x(t)$	0	1/6	1/3	1/2	2/3	5/6	1
	t<1	$1 \le t < 2$	$2 \le t < 3$	$3 \le t < 4$	$4 \le t < 5$	$5 \le t < 6$	$T \ge 6$

Example

3 fair coin tosses

X = number of heads

Y = number of tails

Q1: are X and Y identical distributed **(A)Y**, B) N, C) can’t say)?

Q2: what is the probability that (X=Y): A) 1/2, B) 1, **C) 0**, D) none of the above?

Q3: PMF, CDF?

Answer: calculate PMF of X, and PMF of Y

Ω = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Answer Q1:

	HHH	HHT	HTH	HTT	TTT	TTH	THT	THH
P(x)	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
X(wi)	3	2	2	1	0	1	1	2

Px(0) = 1/8 , Px(1) = 1/8+1/8+1/8 = 3/8 , Px(2) = 1/8+1/8+1/8 = 3/8 , Px(3) = 1/8

x	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

	HHH	HHT	HTH	HTT	TTT	TTH	THT	THH
P(y)	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
Y(wi)	0	1	1	2	3	2	2	1

Py(0) = 1/8 , Py(1) = 1/8+1/8+1/8 = 3/8 , Py(2) = 1/8+1/8+1/8 = 3/8 , Py(3) = 1/8 ,

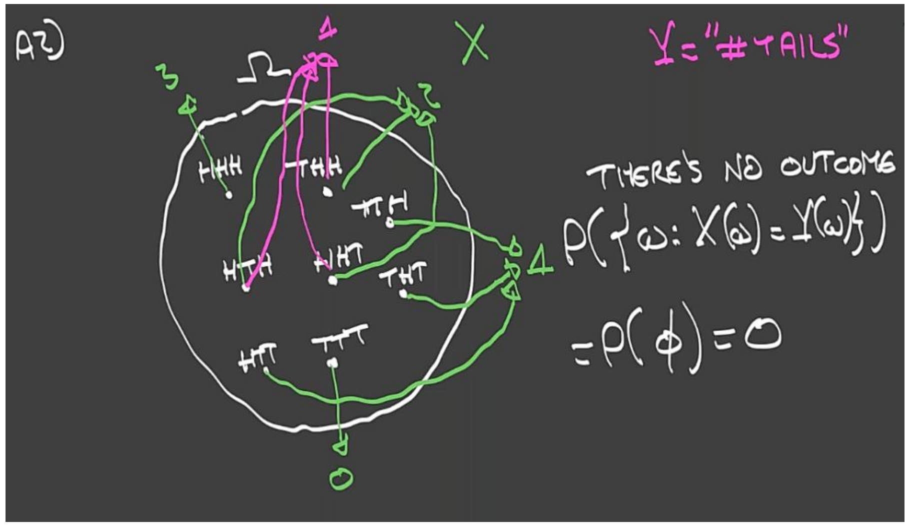
y	0	1	2	3
P(y)	1/8	3/8	3/8	1/8

Answer Q2:

	HHH	HHT	HTH	HTT	TTT	TTH	THT	THH
P(x)	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
X(wi)	3	2	2	1	0	1	1	2
Y(wi)	0	1	1	2	3	2	2	1

That mean $X(wi) \neq Y(wi)$

P.{w : X(w)=Y(w)} = p($\varnothing$) = 0



Example

4-coin tosses

Q1) X ~ Y are X and Y identical distributed?

Q2) P(X=Y) ? **A) 1/2**, B) 1, C) 0, D) None of the above

Transformations:

X ~ F or X ~ f: X= "count of Heads over 3-coin tosses"

What is the distribution of 3-X ~? Or 1/x ~

If we have distribution X, Y what is (X+Y) ~

Functions of RV:

X is a RV ~ F_X

Any function g(X) is also a RV, say Y = g(X)

For any event A:

$$P(Y \in A) = P(g(X) \in A)$$

Can we express the distribution function of Y in terms of P_X?

Example Bernoulli Distribution : PMF (discreet)

X ~ Ber(p), P $\in$ [0,1]

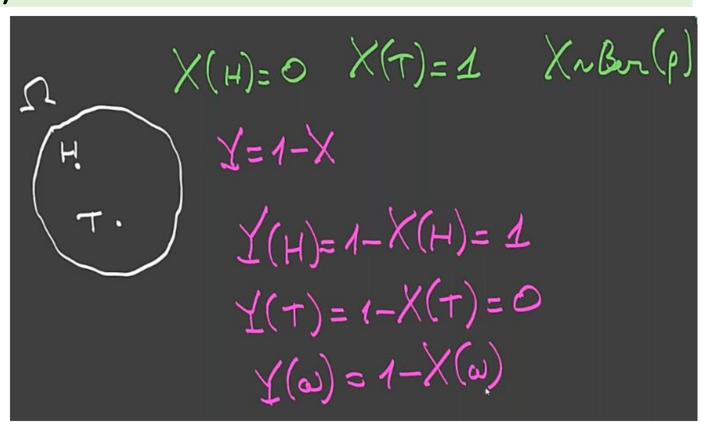
Y = 1 – X ~?

P (X=x) = $\begin{cases} 1-p & x=0 \\ p & x=1 \end{cases}$

Y = 1 – X and Y = y then 1-X = y >> X = 1-y

at y = 0 >> X = 1 which is p

at y = 1 >> X = 0 which is 1-p



P (Y = y) = P (1-X = y) = P (X= 1-y) = $\begin{cases} P(X=1) = p & y=0 \\ p(X=0) = 1-p & y=1 \end{cases}$

Y ~ Ber(1-p)

Example : PMF (discreet)

$X \sim Ber(p)$, $P \in [0,1]$ – what is the distribution of X^2?

$Y = X^2 \sim$

$$P(X=x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \end{cases}$$

$Y = X^2 >> X = \sqrt{y}$

when y = 0 >> X=0 and when y=1 >> x=1

$$P(Y=y) = P(X=\sqrt{y}) = \begin{cases} 1-p & y = 0 \\ p & y = 1 \end{cases}$$

$Y \sim Ber(p)$ >> we can say $X \sim Y$ (X has the same distribution as $Y=X^2$) identically distributed.

Example: PMF (discreet)

$$\text{The random variable } X \sim F_X(x) = \begin{cases} \frac{1}{3}, & x = -1 \\ \frac{1}{3}, & x = 0 \\ \frac{1}{3}, & x = 1 \end{cases}$$

What is the distribution of X^2?

$P(X^2=0) = P(X=0) = 1/3$

$P(X^2=1) = P(X=-1 \cup X=1) = P(X=1) + P(X=-1) = 1/3 + 1/3 = 2/3$

$$Y = X^2 \sim f_Y(y) = \begin{cases} 1/3 & y = 0 \\ 2/3 & y = 1 \end{cases}$$

$Y = X^2 \sim Ber(2/3)$

Transformation of a continuous random variable RV

$X^2 \sim F_X$, $Y=g(X)$, what F_Y? what is the CDF

$$F_Y(t) = P(Y \leq t) = P(g(X) \leq t)$$

If g is monotone **increasing** function, then $P(g(X) \leq t) = P(X \leq g^{-1}(t)) = F_X(g^{-1}(t))$

As example $g(x)=t$ >> the inverse >> $x= g^{-1}(t)$

$F_Y(t) = F_X(g^{-1}(t))$, so density >> the derivative of F_Y >> $F_Y(t)=d/dt\, F_X(g^{-1}(t))$

$F_X(g^{-1}(t))$,

If g is monotone **decreasing** function, then

Because decreasing need to inverse the sine too.

$P(g(X) \leq t) = P(X \geq g^{-1}(t)) = 1-p(X < g^{-1}(t))$

Because X is continuous

$= 1-p(X \leq g^{-1}(t)) = 1 - F_X(g^{-1}(t))$

How do we handle the transformation?

- X is discrete: use PMF table
- X is continuous, and g monotone:
    - use CDF then
    - take the derivative if needed
- You can always use simulations if you're lost.

Functions of RV (cont.)

- $g : \mathcal{X} \mapsto \mathcal{Y}$
- associated with g is the *inverse mapping*:

$$g^{-1}(A) = \{x : g(x) \in A\}$$

- NOTE: in general, $g^{-1}(A)$ maps sets into sets
- if A is a single point $\{y\}$, we can abbreviate to $g^{-1}(y)$
- the result of $g^{-1}(y)$, though, in general, is still a set:

$$g^{-1}(y) = \{x : g(x) = y\}$$

- when $g^{-1}(y)$ is a single point x, we can write:

$$g^{-1}(y) = x$$

Distribution of g(X), X discrete

- if X is discrete, $\mathcal{X}$ is countable
- $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$ is also countable
- thus, Y is discrete, with pmf:

$$\begin{aligned} P(Y = y) &= P(g(X) = y) \\ &= P(X \in g^{-1}(y)) \\ &= \sum_{x \in g^{-1}(y)} P(X = x) \end{aligned}$$

- finally:

$$f_Y(y) = \sum_{x \in g^{-1}(y)} f_X(x)$$

Distribution of Y = g(X), X continuous

In general:

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(g(X) \leq y) \\ &= P(\{x : g(x) \leq y\}) \\ &= \int_{\{g(x) \leq y\}} f_X(x) dx \end{aligned}$$

Distribution of Y = g(X), X continuous, g monotone

$$\begin{aligned} u > v &\implies g(u) > g(v) \quad \textit{(increasing)} \\ &\implies g(u) < g(v) \quad \textit{(decreasing)} \end{aligned}$$

Than g is a 1 to 1 mapping from $\mathcal{X}$ to $\mathcal{Y}$, and $g^{-1}(y)$ is single valued: $g^{-1}(y) = x$
If g is monotone increasing:

$$\begin{aligned} \{x : g(x) \leq y\} &= \{x : g^{-1}(g(x)) \leq g^{-1}(y)\} \\ &= \{x : x \leq g^{-1}(y)\} \end{aligned}$$

Thus:

$$\begin{aligned} F_Y(y) &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)) \end{aligned}$$

A similar line of reasoning applies for a g which is monotone *decreasing*

Theorem

$$\begin{aligned} F_Y(y) &= F_X(g^{-1}(y)), \quad g \textit{ monotone increasing} \\ &= 1 - F_X(g^{-1}(y)), \quad g \textit{ monotone decreasing} \end{aligned}$$

Example: of exam

14:55

EX. $X \sim Ber(p)$, $Y = 1 - X$

Q1) IS Y A VALID R.V.? $\boxed{\text{YES}}$

Q2) IS $X \sim Y$?

-2 A) YES, FOR ANY p

-2 B) NO, FOR ANY p

+6 $\boxed{\text{C) CAN'T SAY, IT DEPENDS ON 'p'}}$

-2 D) NONE OF THE ABOVE

$Y \sim Ber(1-p)$

$Ber(p) \neq Ber(1-p)$

ONLY IF $p = 1-p$

Example

Uniform-Exponential

$X \sim \text{Unif}(0,1)$, $Y = -\ln(X)$, $Y \sim ?$

$f_X(x) = 1, \quad x \in (0,1)$

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

$g = -\ln$ is *monotone decreasing* in $(0,1)$. Thus, g defines a 1 to 1 mapping between $\mathcal{X} = (0,1)$ and $\mathcal{Y} = (0, +\infty)$

For $y > 0$, $y = -\ln(x) \implies x = e^{-y}$

In other words, $g^{-1}(y) = e^{-y}$

We can now 'rediscover' the previously seen theorem, to find:

$$\begin{aligned} F_Y(y) &= P(-\ln(X) \leq y) = P(\ln(X) \geq -y) = P(X \geq e^{-y}) \\ &= 1 - P(X \leq e^{-y}) = 1 - F_X(e^{-y}) \\ &= 1 - e^{-y} \end{aligned}$$

for $y > 0$

Theorem

- *$X \sim f_X$, $\quad Y = g(X)$, g monotone*
- *f_X continuous in $\mathcal{X}$, and $g^{-1}(y)$ has continuous derivatives in $\mathcal{Y}$*
- *then:*

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

for $y \in \mathcal{Y}$, $f_Y(y) = 0$ otherwise

Example:

- $X \sim \text{Gamma}(\beta, n)$, $Y = 1/X$
- $Y = X^2$, X continuous
- $X \sim N(0,1)$, $Y = X^2$ cfr. Statistical Inference, Casella and Berger, 2nd ed., Example 2.1.9

Expected value:

Definition

If X is continuous:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

If X is discrete:

$$E[g(X)] = \sum_{\mathcal{X}} g(x) f_X(x)$$

Example:

$X = \#$ heads in 3 coin tosses

x	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

$$E[X] = 1 \times 3/8 + 2 \times 3/8 + 3 \times 1/8$$
$$= 12/8 = 3/2 = 1.5$$

The expected value of the random variable X, this case it is discreet

If it is continuous will work with density

Example of a Uniform distribution (doing the integral)

$X \sim \text{Unif}(0,1)$

$$f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \int_0^1 x\,dx = \left[\frac{x^2}{2}\right]_0^1$$
$$= 1/2 - 0$$
$$= 1/2$$

RECAP:

Expected Value of R.V

- X is discrete: E[X] = $\sum_{xi \in X} xi.P(X = xi)$, (P is probability mass function)

- X is continuous: E[X] = $\int_{xi \in X} xi.f_x(x)dx$

All the possible values that random value can take

Example: (discrete)

X is the number of heads in 3 coins tosses.

What is the expected value of x > E[X]=? And E[X^2]

x	0	1	2	3	Sum
P(X=x)	1/8	3/8	3/8	1/8	
x.P(X=x)	0	3/8	6/8	3/8	12/8
X^2.P(X=x)	0	3/8	12/8	9/8	24/8

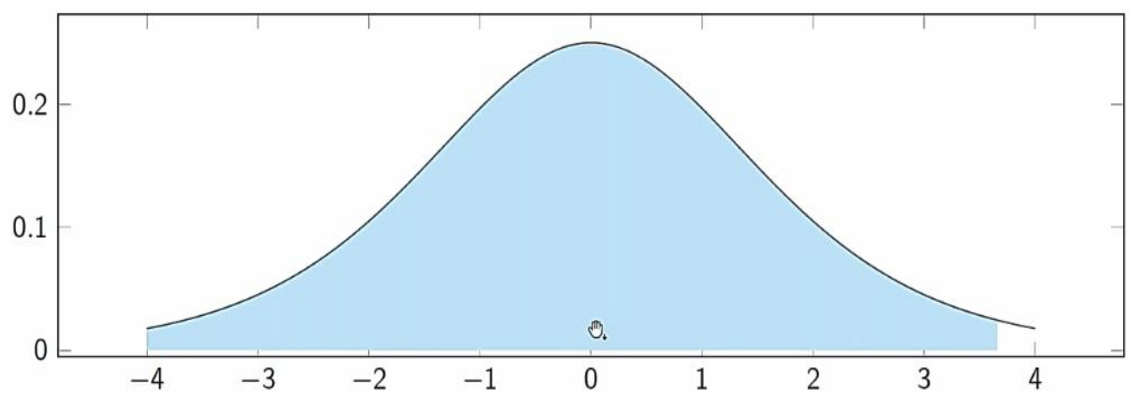
E[X] = 12/8 = 3/2 = 1.5

E[X^2] = 24/8 = 3

Interpretation:

- A First approximation of a RV random variable to A constant.
- Center of gravity of F_x (PDF or CDF)

what's x : $P(X \leq x) = 0.975$? $\quad F_X^{-1}(p) = \ln(p/(1-p)) \implies F_X^{-1}(0.975) \simeq 3.66$



0 is the center to balance the distribution.

Expected Value (E.V) General definition:

Discrete: E[g(x)] = $\sum_{xi \in X} g(xi).P(X = xi)$

Continuous: E[g(x)]] = $\int_{xi \in X} g(xi) f_x(x) dx$

Expected value is **a Linear operator >>**

e.g., E[2X] = 2E[X] E[3X-1] = 3E[X] – 1 E[X+Y] = E[X] + E[Y]

Example: Bernoulli distribution (discrete RV)

$X \sim \text{Ber}(p)$

$$f_X(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \end{cases}$$

$$\begin{aligned} E[X] &= p \times 1 + (1 - p) \times 0 \\ &= p \end{aligned}$$

$E[X^2] = p \times 1^2 + (1-p) \times 0^2 = p$

Example: Uniform distribution (continuous RV)

$X \sim \text{Unif}(0,1)$

$$f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[X] &= \int_0^1 x\,dx = \left[\frac{x^2}{2}\right]_0^1 \\ &= 1/2 - 0 \\ &= 1/2 \end{aligned}$$

Expected Value: some properties

- E[aX + b] = aE[X] + b
- E[ag1(X) + bg2(X) + c] = aE[g1(X)] + bE[g2(X)] + c
- g(x) ≥ 0 ∀x → Eg(X) ≥ 0 (if transformation g(x) non negative, then the expected value will be non negative)
- g1(x) g2 (x) ∀x → Eg1(X) ≥ Eg2(X) (if we have one transformation always greater than or equal the second one then the expected value of the first one will be greater than or equal the expected value for the second one)
- a ≤ g(x) ≤ b ∀x → a ≤ Eg(X) ≤ b (if transformation between 2 values then the expected value of the transformation will be also between the same 2 values.

$$E[X^2] \neq (EX)^2$$
$$E[\log(X)] \neq \log(EX)$$

Example:

X = "number of network outages in a given day"

For each outage 500 Euro loss given the distribution PMF as it is discrete

X (number of outages)	0	1	2
P(X=x) probability	0.7	0.2	0.1

Q1: what is the distribution of the daily losses? (Simple transformation of a random variable)

Q2: what is the expected daily loss? (Calculating the expected value)

Answer 1: Y = 500 X (total daily loss)

X (number of outages)	0	1	2	sum
Y=500 X	0	500	1000	
P(Y=y) probability	0.7	0.2	0.1	
E[X]=x.P(X=x)	0	0.2	0.2	0.4
E[y]y.P(Y=y)	0	100	100	200

We can calculate Expected value of x then multiply by 500 or directly by y.p(Y=y)

Answer 2: E[Y] = E[500X] = 500E[X] = 500(0.4) = 200 EURO

Variance: (is the excepted value of transformation of X squared)

$V[X] = E[(X - EX)^2]$

Interpretation of variance (the name of volatility, risk, uncertainty), the expected value tills us the center of gravity of the distribution, the variance tills us how much it is dancing around the distribution.)

Applying the same in previous example:
EX. Daily losses >> E[Y] = 200 EURO
Variance $V[Y] = E[(Y-EY)^2]$

Y	0	500	1000	Total (expected value)
Y-EY = Y - 200	-200	300	800	
$(Y - EY)^2 = (Y - 200)^2$	40000	90000	640000	
$P(Y - 200)^2$	0.7	0.2	0.1	
$(Y - 200)$ 2.P $(Y - 200)^2$				110,000

Normally **lower variance** means less volatility, less risk which **is better**

Some properties:

$V(aX + b) = a^2V(X)$, (a and b are constants, can take the a constant out from the variance squaring it)

$V(X) = E[X^2] - E[X]^2$

As example $V[X] = 5$ >> $V[500.X] = 500^2\, V[X] = 500^2 \times 5$

Example

X = number of heads in 3-coin tosses, what is the variance.

Solution:

$V[X] = E[(X-EX)^2] = E[X^2] - E[X]^2$

EX = X.p(X) (for discrete distribution)

	1	2	3	4	5	6	7	8
Ω	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$P(\Omega)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
X #	3	2	2	2	1	1	1	0

X = No of heads	0	1	2	3	E - sum
P(X)	1/8	3/8	3/8	1/8	
E = X. P(X)	0	3/8	6/8	3/8	12/8 = 3/2=1.5
$E[X^2]= X^2.P(X)$	0	3/8	4x3/8=12/8	9x1/8=9/8	24/8 = 3

$V[X] = E[X^2] - E[X]^2 = 3 - (1.5)^2 = 3 - 2.25 = 0.75$

Example Bernoulli distribution

$X \sim \text{Ber}(p)$

$$f_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$$

$$E[X^2] = p \times 1^2 + (1-p) \times 0^2$$
$$= p$$

$$V[X] = E[X^2] - E[X]^2$$
$$= p - p^2$$
$$= p(1-p)$$

Example Uniform distribution (continuous random variable)

$X \sim \text{Unif}(0, 1)$

$$f_X(x) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

$$E[X^2] = \int_0^1 x^2\, dx = \left[\frac{x^3}{3}\right]_0^1$$
$$= 1/3 - 0$$
$$= 1/3$$

$$V[X] = E[X^2] - E[X]^2$$
$$= 1/3 - (1/2)^2$$
$$= 1/12 \simeq 0.0833$$

From previous same example expected value of X calculated was (0.5) so $E[X]^2 = (0.2)^2$

Example Normal distribution (continuous random variable)

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

P.D.F.

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

Probability density function N has 2 parameters , both are the expected value of the distribution and variance

$$E[X] = \mu$$
$$V[X] = \sigma^2$$

Theorem
If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0,1)$

If we have a variable normally distributed then linear transformation of that variable is normally distributed.

Proof.

$$\begin{aligned} P(Z \leq z) &= P\left(\frac{X-\mu}{\sigma} \leq z\right) \\ &= P\left(X \leq z\sigma + \mu\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{z\sigma+\mu} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \end{aligned}$$

substitute: $t = \frac{x-\mu}{\sigma}$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} \exp\left(-\frac{t^2}{2}\right) dt$$

Example Moments of a General Normal distribution

$X \sim N(\mu, \sigma^2)$
We can think of X as $X = Z\sigma + \mu, Z \sim N(0,1)$

$$E[X] = \mu$$
$$V[X] = \sigma^2$$

Chebychev's Inequality

Theorem
Let X be a RV, and $g(x)$ a non-negative function. Then, $\forall r > 0$,

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

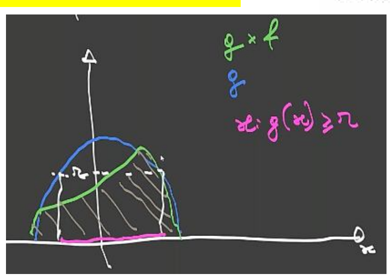
A random variable x and some transformation that is non negative
Then the following inequality: the probability that g(x) is greater than or equal some r a positive threshold is less than or equal to the expected value of x divided by that threshold r

In other word:
The probability that the transformation of x exceeds the threshold is limited by the expected value of this transformation divided by r.

Proof:

By replacing g(x) with r (constant) we can take r outside the integrals

- The product g.f the green curve.
- x: g(x) ≥ r so we stick with the area highlighted under **r**
- so the integral will be smaller than the expected value of g(x)



$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{+\infty} g(x) f_X(x) dx \\ &\geq \int_{x:g(x)\geq r} g(x) f_X(x) dx \\ &\geq r\int_{x:g(x)\geq r} f_X(x) dx \\ &= rP(g(X) \geq r) \end{aligned}$$

Universal Boundary:

$$g(x) = (x-\mu)^2/\sigma^2,\ \mu = E[X], \sigma^2 = V[X],$$
$$r = t^2$$

E[X] : Expected value, V[X] variance, g(x) is non negative.
Per Chebychev's inequality: lets apply the inequality:

$$P\left(\frac{(X-\mu)^2}{\sigma^2} \geq t^2\right) \leq \frac{1}{t^2} E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{1}{t^2}$$

After simplify using algebra:

$$P\left(|X-\mu| \geq t\sigma\right) \leq \frac{1}{t^2}$$

For instance, if we chose t=2, we have:

$$P(|X-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = 0.25$$

Example: Telco company losses

Expected value (daily losses) E[X] = 200 EURO
Variance or risk V[X] = 110,000 EURO^2
Standard deviation (SD(X))= square root of (V[X]) = 333.3 EURO)
So the probability that P(|X-200| ≥ 669 EURO) we can say it will be less than ≤ 0.25

An equality for the normal distribution

$X \sim N(\mu, \sigma)$

$$\begin{aligned} P(|X-\mu| \geq 2\sigma) &= P(|X-\mu|/\sigma \geq 2) \\ &= P(|Z| \geq 2) \\ &= 2\frac{1}{\sqrt{2\pi}} \int_2^{\infty} e^{-\frac{x^2}{2}}\, dx \\ &\simeq 0.0455 \end{aligned}$$

Recap:

- what is a RV
- distribution, cdf, pmf, pdf
- distribution of transformed RV
- expected value, variance
- the Normal distribution
- important inequalities

Definition

A random vector is a function:

$$\Omega \mapsto \mathbb{R}^n$$

Example

Tossing 2 fair dice, $X = \text{sum}$, $Y = |\text{difference}|$

ω	$P(\omega)$	$X(\omega)$	$Y(\omega)$
$(1,1)$	$1/36$	2	0
$(1,2)$	$1/36$	3	1
...			
$(6,6)$	$1/36$	12	0

Joint pmf: discrete RV

Definition

For discrete X, Y, we define the joint pmf as:

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

We have:

$$P((X,Y) \in A) = \sum_{(x,y)\in A} f_{X,Y}(x,y)$$

Two immediate consequences of the definition:

1. $f_{X,Y} \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$
2. $\sum_{(x,y)\in\mathbb{R}^2} f_{X,Y}(x,y) = 1$

Example

continuing on the previous dice example:

$$f(5,3) = P(X=5, Y=3) = P(\{(4,1),(1,4)\}) = 2/36 = 1/18$$

$$f(6,0) = P(X=6, Y=0) = P(\{(3,3)\}) = 1/36$$

$$P(X=7, Y \leq 4) = f(7,1) + f(7,3) = 1/18 + 1/18 = 1/9$$

Example

The complete $f_{X,Y}$ pmf:

Y \ X	2	3	4	5	6	7	8	9	10	11	12
0	1/36		1/36		1/36		1/36		1/36		1/36
1		1/18		1/18		1/18		1/18		1/18	
2			1/18		1/18		1/18		1/18		
3				1/18		1/18		1/18			
4					1/18		1/18				
5						1/18					

Expected Value: discrete RV

$$Eg(X,Y) = \sum_{(x,y)\in R^2} g(x,y) f_{X,Y}(x,y)$$

Example

In the 2 fair dice example, can you compute $\mathrm{E}(X \cdot Y)$?

$$\begin{aligned}
\mathrm{E}[X \cdot Y] &= \sum_{(x,y)\in\mathbb{R}^2} x \cdot y \cdot f_{X,Y}(x,y) \\
&= 2 \cdot 0 \cdot 1/36 + 4 \cdot 0 \cdot 1/36 + \ldots + 8 \cdot 4 \cdot 1/18 + 7 \cdot 5 \cdot 1/18 \\
&= 13 \times 11/18
\end{aligned}$$

Example

Consider the following joint distribution:

$$\begin{aligned}
&f(0,0) = f(0,1) = 1/6 \\
&f(1,0) = f(1,1) = 1/3 \\
&f(x,y) = 0 \quad \text{for any other } (x,y)
\end{aligned}$$

It is easy to verify that f is a valid joint pmf

We can use f to perform computations, like e.g. $P(X=Y) = f(0,0) + f(1,1) = 1/2$

Note: we don't have to explicitly refer to Ω!

Marginal distribution: discrete RV

Theorem

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x,y)$$

Proof.

$$\begin{aligned} f_X(x) = P(X = x) &= P(X = x, Y \in \mathbb{R}) \\ &= \sum_{(x,y) \in A_x} f_{X,Y}(x,y) \\ &= \sum_{y \in \mathbb{R}} f_{X,Y}(x,y) \end{aligned}$$

Example

Recall the following joint pmf:

$$\begin{aligned} f_{X,Y}(0,0) &= f(0,1) = 1/6 \\ f_{X,Y}(1,0) &= f(1,1) = 1/3 \\ f_{X,Y}(x,y) &= 0 \quad \text{for any other } (x,y) \end{aligned}$$

What are $f_X(x)$ and $f_Y(y)$?

$$f_X(x) = \begin{cases} 2/6 & x = 0 \\ 2/3 & x = 1 \end{cases} \qquad X \sim Ber(2/3)$$

$$f_Y(y) = \begin{cases} 3/6 & y = 0 \\ 3/6 & y = 1 \end{cases} \qquad Y \sim Ber(1/2)$$

Note how they're both valid pmfs

Example

Tossing 2 fair dice, X = sum, Y = |difference|

X	2	3	4	5	6	7	8	9	10	11	12	any
Y												
0	1/36		1/36		1/36		1/36		1/36		1/36	3/18
1		1/18		1/18		1/18		1/18		1/18		5/18
2			1/18		1/18		1/18		1/18			4/18
3				1/18		1/18		1/18				3/18
4					1/18		1/18					2/18
5						1/18						1/18
any	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36 3/36	2/36	1/36		

Important Notes. Using the marginal pmfs f_X and f_Y, we can compute $E[X]$ and $E[Y]$, respectively. However, we can only compute $E[XY]$ trough the joint pmf! In general, we need the joint pmf for all questions that involve X and Y *jointly*. e.g., trick question: can you compute $P(X < Y)$ using only f_X and f_Y?
Btw, what's $P(X < Y)$ (using all the available info)?

Without providing the PMF distribution

PROBLEM.

$X \sim Ber(1/3)$

$Y \sim Ber(2/3)$

$P(X < Y) = ?$

-2 A) 1

-2 B) 0

-2 C) 1/2

+6 D) CAN'T ANSWER: NOT ENOUGH INFO

$X \sim Ber(1/3)$

$Y \sim Ber(2/3)$

Q) $E[X+Y]$

$= E[X] + E[Y]$

-2 A) 0

-2 B) 1/2

+6 C) 1

-2 D) NONE OF THE ABOVE

Joint pdf: continuous RV

Definition

$f(x,y) : \mathbb{R}^2 \mapsto \mathbb{R}$ is a joint pdf of (X,Y) if, $\forall A \subset \mathbb{R}^2$,

$$P((X,Y) \in A) = \int \int_A f(x,y)dxdy$$

Note: all $f : \mathbb{R}^2 \mapsto \mathbb{R}$ such that $f \geq 0 \quad \forall (x,y)$, and $\int \int_{\mathbb{R}^2} f = 1$ are joint pdfs for some RV (X,Y)

Expected value: continuous RV

$$E(g(X,Y)) = \int \int_{\mathbb{R}^2} g(x,y) f(x,y)dxdy$$

Marginal distribution: continuous RV

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$X \sim N(0,1)$

$Y \sim N(1,1)$

Q1) $P(X<Y)$? (CAN'T SAY)

Q2) $E[X+Y] = ?$ (1)

Conditional Distribution

$f(x,y) = P(X=x, Y=y)$ $\quad \sum_{x,y} f(x,y) = 1$

$f(y|x) = P(Y=y \mid X=x)$ $\quad \sum_{y} f(y|x) = 1$

Definition

$$f(y|x) = f(x,y)/f_X(x), \quad \forall x : f_X(x) > 0$$

Note: $f(y|x)$ is a valid pmf/pdf!

Independence:

Definition

X, Y are independent iff $f(x,y) = f_X(x) f_Y(y)$

Note: if X, Y are independent, knowledge of X does not give info on Y:

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y)$$

Theorem

If X, Y are independent, then:

1. $A \subset \mathbb{R}$, $B \subset \mathbb{R}$,

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)$$

i.e. $\{X \in A\}$ and $\{Y \in B\}$ are independent

2. *let g be a function of only x, and h be a function of only y. Then:*

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

• INDEPENDENCE IS A STRONG CONDITION

$X \perp Y$

$g(X) \perp h(Y)$

$X \perp Y$

$E[XY] = E[X]E[Y]$

EX. X = "DENIZ'S SCORE AT THE TEST"
Y = "BENOIT'S SCORE"

$X \perp\!\!\!\perp Y$

$\{X > 100\} \perp\!\!\!\perp \{Y > 100\}$

Example

$$f_{X,Y}(0,0) = f(0,1) = 1/6$$
$$f_{X,Y}(1,0) = f(1,1) = 1/3$$
$$f_{X,Y}(x,y) = 0 \quad \text{for any other } (x,y)$$

Are X and Y independent?

Remember: we just need to verify that the definition holds, namely that $f_{X,Y} = f_X \cdot f_Y$
We essentially only need $f_{X,Y}$ to answer this question, as f_X and f_Y can be computed from $f_{X,Y}$ (but not vice versa!)

$$f_X(x) = \begin{cases} 2/6 & x = 0 \\ 2/3 & x = 1 \end{cases} \qquad f_Y(y) = \begin{cases} 3/6 & y = 0 \\ 3/6 & y = 1 \end{cases}$$

- $f_{X,Y}(0,0) = 1/6 = f_X(0)f_Y(0)$
- $f_{X,Y}(0,1) = 1/6 = f_X(0)f_Y(1)$
- $f_{X,Y}(1,0) = 1/3 = f_X(1)f_Y(0)$
- $f_{X,Y}(1,1) = 1/3 = f_X(1)f_Y(1)$

Thus, the answer is **yes**, X and Y are independent

EXAM QUESTION

$f_{X,Y}$:

$Y \backslash X$	0	1	
0	1/6	1/3	1/2
1	1/6	1/3	1/2
	1/3	2/3	

Q1) IS IT VALID? YES
Q2) IS $X \perp\!\!\!\perp Y$? YES

Example

In the 2 dice example, X = sum, Y = |difference|, are X and Y independent?
Take a close look again at the joint pmf:

X / Y	2	3	4	5	6	7	8	9	10	11	12
0	1/36		1/36		1/36		1/36		1/36		1/36
1		1/18		1/18		1/18		1/18		1/18	
2			1/18		1/18		1/18		1/18		
3				1/18		1/18		1/18			
4					1/18		1/18				
5						1/18					

Note how the *domain* of Y changes depending on X, and vice versa
We can thus already exclude that X and Y are independent, without doing any calculation
In general, a necessary (but not sufficient) condition for two RV X, Y to be independent is that their joint domain is a square

Correlation:

Definition

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

Definition

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

- VARIANCE
- MULTIPLE R.V.
- JOINT PMF/PDF
- INDEPENDENCE
- COVARIANCE (CORRELATION)

Theorem:

1.
$$Cov(X, Y) = E(XY) - \mu_X \mu_Y$$

2. *if X, Y are independent, then:*
$$Cov(X, Y) = 0$$
and $\rho_{X,Y} = 0$

3.
$$Var(X + Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

$$V[X+Y] = V[X] + V[Y] + 2\,Cov(X, Y)$$

Proof.

$$\begin{aligned} V[X+Y] &= E[(X+Y)^2] - (E[X+Y])^2 \\ &= E[X^2 + 2XY + Y^2] - (\mu_X^2 + 2\mu_X\mu_Y + \mu_Y^2) \\ &= \mu_{X^2} + 2\mu_{XY} + \mu_{Y^2} - \mu_X^2 - 2\mu_X\mu_Y - \mu_Y^2 \\ &= V[X] + V[Y] + 2Cov(X, Y) \end{aligned}$$

14:15

EX. 3.13 FROM TEXBOOK: PORTFOLIO OPTIMIZATION

INVESTING IN 2 COMPANIES XX, YY

- SHARES OF XX COST 20\$, EXPECTED RETURN 1\$
  S.D. OF 0.5\$

- SHARES OF YY COST 50\$, EXPECTED RETURN 2.5\$
  S.D. OF 1\$

ASSUMPTION: $XX \perp YY$, HAVE 10'000\$ TO INVEST.

3 PORTFOLIOS:
A) BUY 500 SHARES OF XX
B) BUY 200 SHARES OF YY
C) BUY 250 SHARES OF XX + 100 SHARES OF YY.

Q. Expected returns and variance from the 3 portfolios

	E.R.	SD.	VAR.	COST
XX	1	0.5	0.25	20
YY	2.5	1	1	50

A) BUY 500 SHARES OF XX

A = "RETURNS FROM PORTFOLIO A" = 500 * X

X = "RETURNS FROM 1 SHARE OF XX"

$$E[A] = 500\,E[X] = 500\ \$$$

$$V[A] = V[500X] = 500^2 \cdot 0.25 = 62'500\ \$^2$$

B) $E[B] = 500\ \$$

$V[B] = 40'000\ \$^2$

V[B] < V[A], so B is lower risk

c) $C = 250X + 100Y$

$$E[C] = E[250X + 100Y] = 250E[X] + 100E[Y]$$
$$= 500\$$$

X, Y ARE INDEP.

$$V[C] = V[250X + 100Y] = V[250X] + V[100Y]$$
$$= 250^2 \underbrace{V[X]}_{0.25} + 100^2 \underbrace{V[Y]}_{1} = 25'625$$

C is the lowest risk

Multivariate Distributions: general case

$\mathbf{X} = (X_1, \ldots, X_n)$
If $\mathbf{X}$ is *discrete*, we define the joint pmf f such that:

$$f(\mathbf{x}) = f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

If $\mathbf{X}$ is *continuous*, we define the joint pdf f such that:

$$\forall A \in \mathcal{B}, \quad P(X \in A) = \int \cdots \int_A f(\mathbf{x}) d\mathbf{x}$$

In general, the joint cdf:

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

$X_1, \ldots, X_n$ are mutually independent iff:

$$f(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

Random Samples

Definition
$\{X_1, \ldots, X_n\}$ is a random sample of size n from the population $f(\theta)$, $\theta \in \Theta$ iff:

- $\{X_i\}$ are mutually independent
- $X_i \sim f(\theta) \quad \forall i$

We commonly abbreviate the above conditions with the following statement:

$X_1, \ldots, X_n$ *are i.i.d., with* $X_i \sim f$

From the definition of independence, we'll have:

$$f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Definition
A statistic Y_n is a function of the random sample $\{X_i, i = 1, \ldots, n\}$:

$$Y_n = T(X_1, \ldots, X_n)$$

Example:

RANDOM SAMPLE: EX.

$$X_i = \begin{cases} 0 & \text{IF STUDENT "}i\text{" FAILS} \\ 1 & \text{" PASSES} \end{cases} \qquad n=80$$

$$X_i \sim Ber(p), \text{ iid}$$

$$\bar{X}_{80} = \frac{1}{80}\sum_{i=1}^{80} X_i = \text{"fraction of students passing the test"}$$

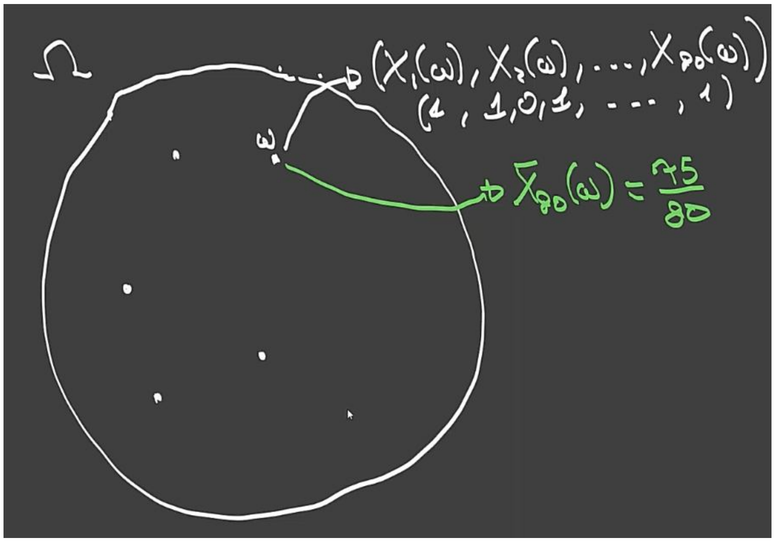
Example:
All have the same distribution X, and same expected value and same variance.

$$X_i \sim X, E[X] = \mu, V[X] = \sigma^2$$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n} n E[X]$$
$$= E[X] = \mu$$

$$V[\bar{X}_n] = \frac{1}{n^2}\sum_{i=1}^{n} V[X_i] = \frac{1}{n^2} n V[X]$$
$$= \frac{1}{n} V[X] = \frac{\sigma^2}{n}$$



Ex.: NOISY MEASUREMENTS

X_i = "measurement trial i"

$E[X_i] = \mu \quad V[X_i] = \sigma^2$

$n = 10 \quad E[\bar{X}_{10}] = \mu \quad V[\bar{X}_{10}] = \frac{\sigma^2}{10}$

Example: sample variance

$X_i \sim X, E[X] = \mu, V[X] = \sigma^2$

$$\bar{S}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$\begin{aligned} E[S_n^2] &= E\left(\frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2)\right) \\ &= \frac{1}{n-1}\left(nE[X^2] - nE[\bar{X}_n^2]\right) \\ &= \frac{1}{n-1}\left(n(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)\right) \\ &= \sigma^2 \end{aligned}$$

1. Random Vector
2. joint pmf/pdf
3. independence, conditional distribution
4. expected value, covariance, correlation
5. random samples

Convergence:

Stochastic Processes: full sequence of random variable never-ending.

Stochastic processes convergence in distribution:

Definition

$$X_n \to^d X$$

iff:

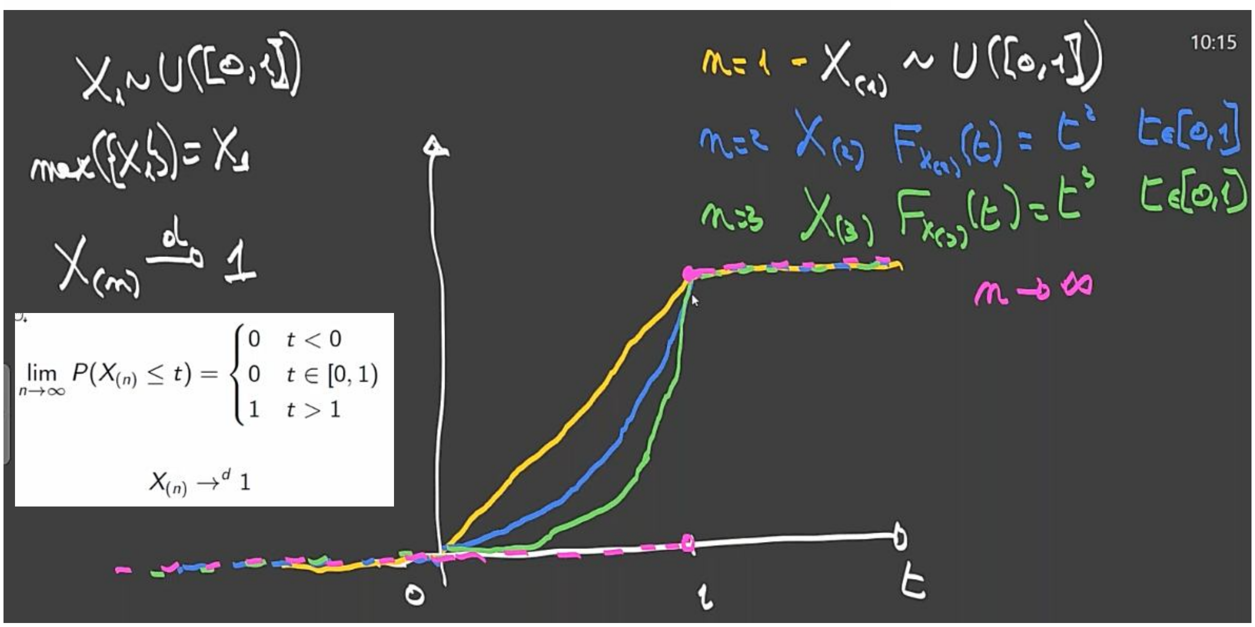
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \forall x \quad \text{where } F_X \text{is continuous}$$

Example

X_i i.i.d. $\sim \text{Unif}(0,1), \quad i = 1, \dots, n$

$X_{(n)} = \max_{1\le i\le n} X_i$

$$P(X_{(n)} \le t) = \begin{cases} 0 & t < 0 \\ t^n & t \in [0,1) \\ 1 & t > 1 \end{cases}$$



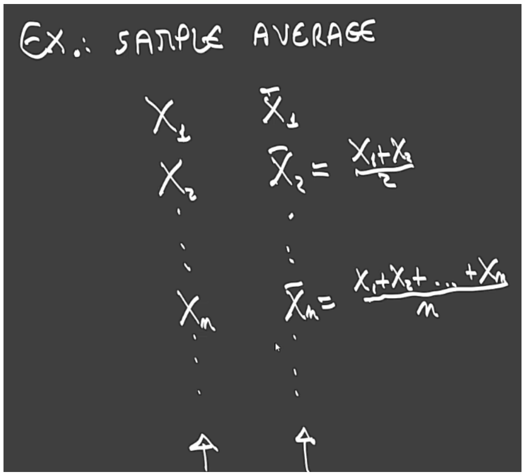
Convergence in Probability: is the limit of probabilities converges to 1

Definition

$$X_n \to^p X$$

iff, $\forall \epsilon > 0$:

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$$



Example

X_i i.i.d. $\sim$ Unif(0, 1)
$X_(n) = \max_{1\le i\le n} X_i$. Does X_n converge in probability ? To what?
$\forall \epsilon > 0$:

$$\begin{aligned} P(|X_{(n)} - 1| \geq \epsilon) &= P(X_{(n)} \geq 1+\epsilon) + P(X_{(n)} \leq 1-\epsilon) \\ &= 0 + P(X_{(n)} \leq 1-\epsilon) \end{aligned}$$

The X_i are i.i.d.:

$$P(X_{(n)} \leq 1-\epsilon) = P(X_i \leq 1-\epsilon, i = 1, \dots, n) = (1-\epsilon)^n \to 0$$

As n goes to infinity the probability that the difference greater than epsilon is goes to zero or conversely that the probability that distance is smaller than epsilon is goes to 1 so the maximum converges in probability to 1

Thus, $X_{(n)} \to^p 1$

Theorem
If $X_n \to^p X$, then $X_n \to^d X$

P: probability, d: distribution

Theorem
$X_n \to^p \mu$, a constant, iff $X_n \to^d \mu$

Almost sure convergence:

Definition

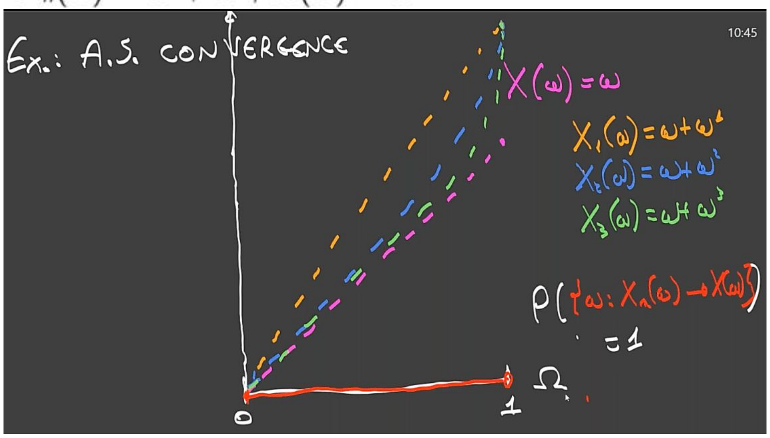
$$X_n \to^{a.s.} X$$

iff, $\forall \epsilon > 0$:

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$$

Example

Let the *sample space* $\Omega = [0,1]$, with the uniform probability distribution
$X_n(\omega) = \omega + \omega^n$, $X(\omega) = \omega$



$\forall \omega \in [0,1)$, $\omega^n \to 0$, and $X_n(\omega) \to X(\omega)$
for $\omega = 1$, $X_n(\omega) = 2 \quad \forall n$, so $X_n(1)$ does not converge to $X(1)$
however, $P([0,1)) = 1$, so $X_n \to X$ on a set of probability 1
that is, $X_n \to^{a.s.} X$

Theorem

Almost sure convergence implies convergence in probability

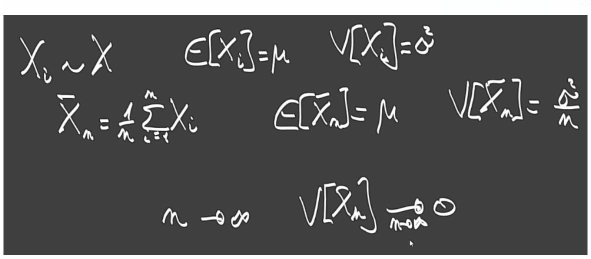
The converse is not always true

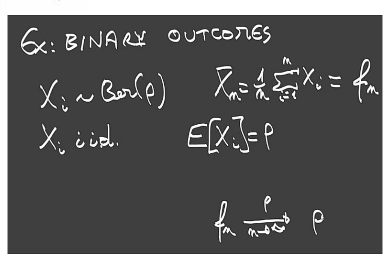
Weak law of large numbers

Theorem

X_i i.i.d., $EX_i = \mu$, $VarX_i = \sigma^2 < \infty$
$\bar{X}_n = 1/n \sum_{i=1}^n X_i$
Then,

$$\bar{X}_n \to^{p} \mu$$





Proof.

$\forall \epsilon > 0,$

$$P(|\bar{X}_n - \mu| \geq \epsilon) = P((\bar{X}_n - \mu)^2 \geq \epsilon^2)$$

By Chebychev's inequality:

$$P(|\bar{X}_n - \mu| \geq \epsilon) = P((\bar{X}_n - \mu)^2 \geq \epsilon^2) \leq \frac{E(\bar{X}_n - \mu)^2}{\epsilon^2} = \frac{V[\bar{X}_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Hence,

$$\begin{aligned} P(|\bar{X}_n - \mu| < \epsilon) &= 1 - P(|\bar{X}_n - \mu| \geq \epsilon) \\ &\geq 1 - \sigma^2/(n\epsilon^2) \\ &\rightarrow_{n\rightarrow\infty} 1 \end{aligned}$$

Strong Law of Large Numbers

Theorem

X_i i.i.d., $EX_i = \mu$, $VarX_i = \sigma^2 < \infty$

$\bar{X}_n = 1/n \sum_{i=1}^{n} X_i$

Then,

$$\bar{X}_n \rightarrow^{a.s.} \mu$$

Central Limit Theorem

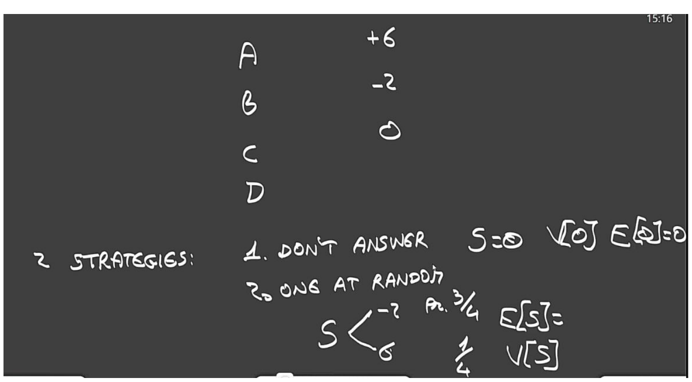
Theorem

$X_i, i = 1, \ldots, n$ i.i.d, $EX_i = \mu$, $0 < VarX_i = \sigma^2 < \infty$

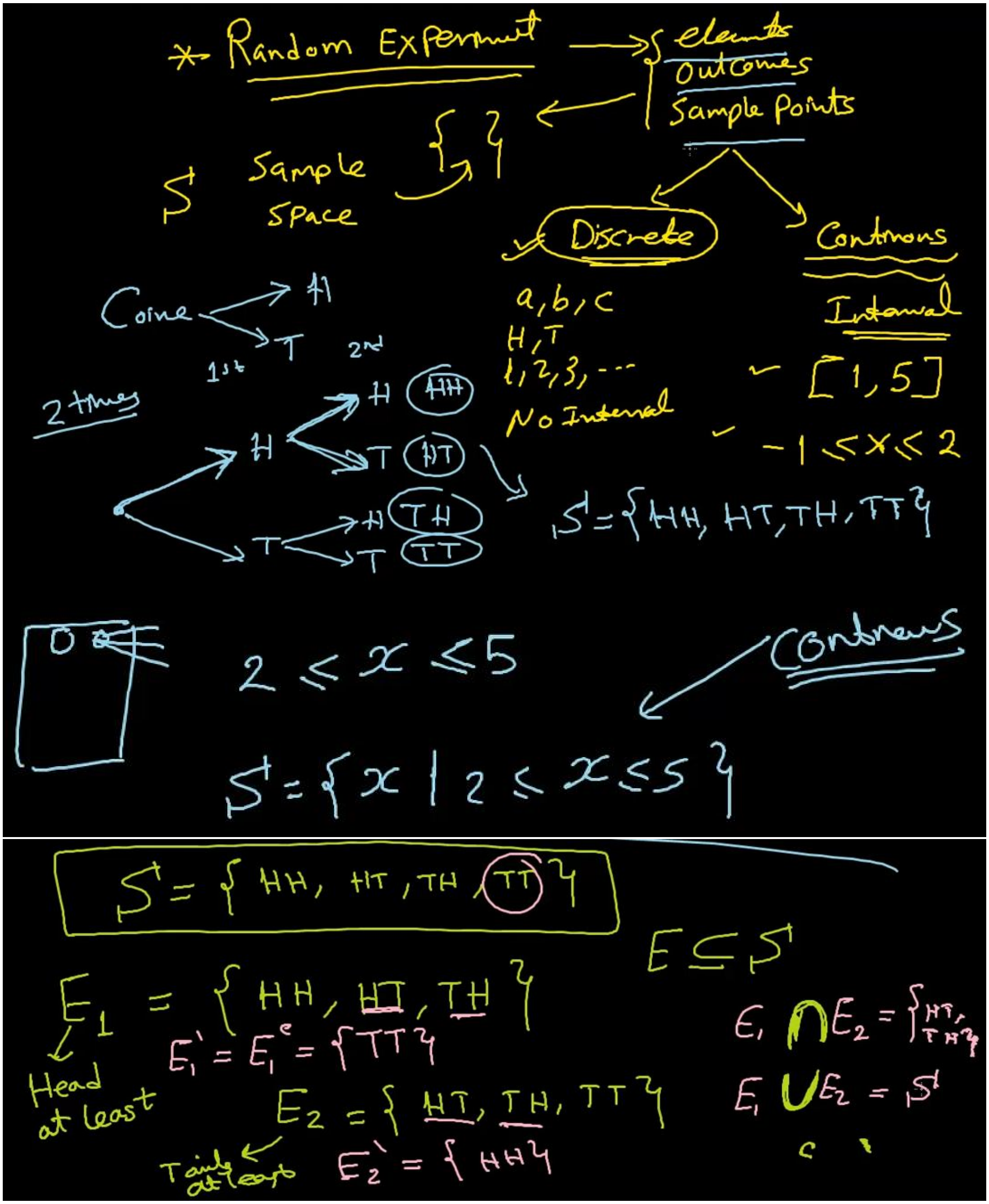
$\bar{X}_n = 1/n \sum_{i=1}^{n} X_i$

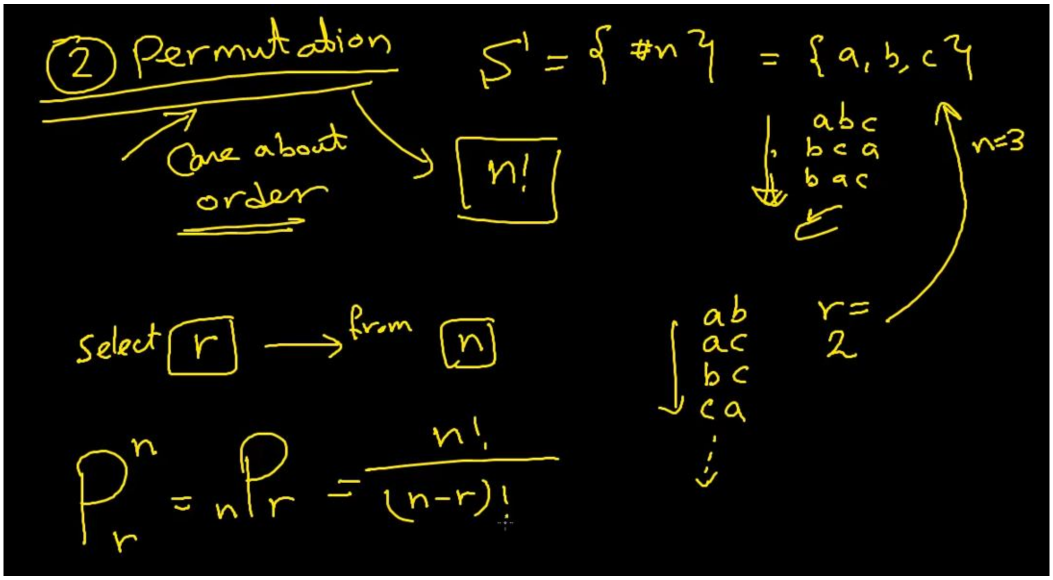
Then:

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \rightarrow^{d} N(0,1)$$

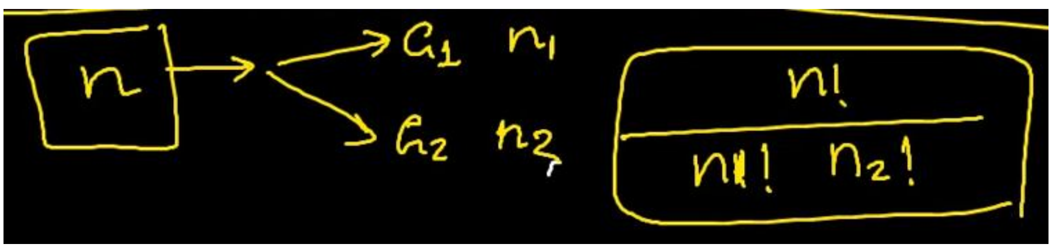


RECAP:

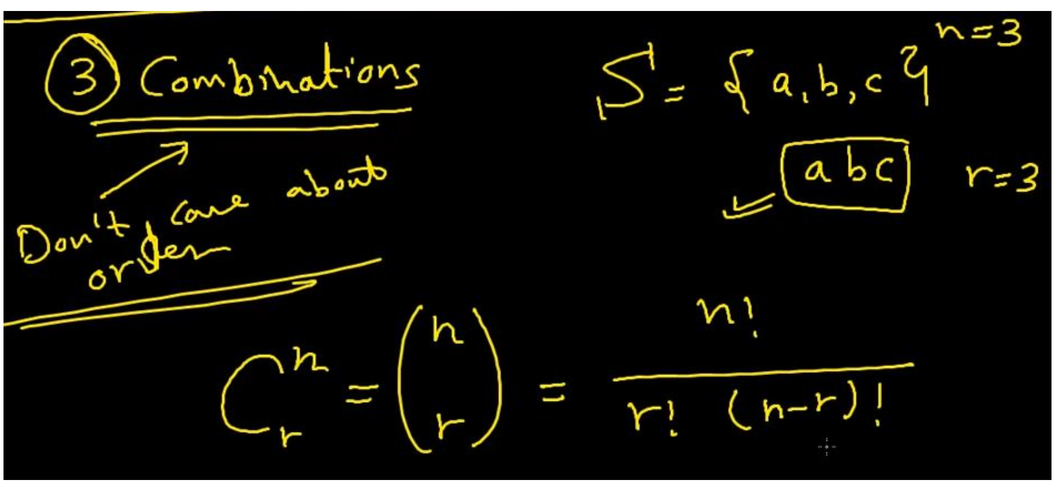

* Random Experiment
elements
outcomes
Sample points
S Sample Space { }
Discrete
Continuous
Interval
a, b, c
H, T
1, 2, 3, ...
No Interval
[1, 5]
-1 ≤ x ≤ 2
Coine
H
T
2 times
1st
2nd
HH
HT
TH
TT
S = {HH, HT, TH, TT}
2 ≤ x ≤ 5
Continuous
S = {x | 2 ≤ x ≤ 5}
S = {HH, HT, TH, TT}
E ⊆ S
E1 = {HH, HT, TH}
Head at least
E1' = E1c = {TT}
E2 = {HT, TH, TT}
Tails at least
E2' = {HH}
E1 ∩ E2 = {HT, TH}
E1 ∪ E2 = S



Permutation of similar groups:



If order not important (not care about order) (for example abc only one choice)



Probability

$$1 \geq P(E) \geq 0$$

$$E \subseteq S$$

$$P(S) = 1 \qquad P(\emptyset) = 0$$

$$P(E') = 1 - P(E)$$

$$P(E) = \frac{n(E)}{n(S)} \quad \text{Equally Likely}$$

Conditional Probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad P(B) > 0$$

given

1. A, B Disjoint

$$A \cap B = \emptyset$$

$$P(A \cap B) = 0$$

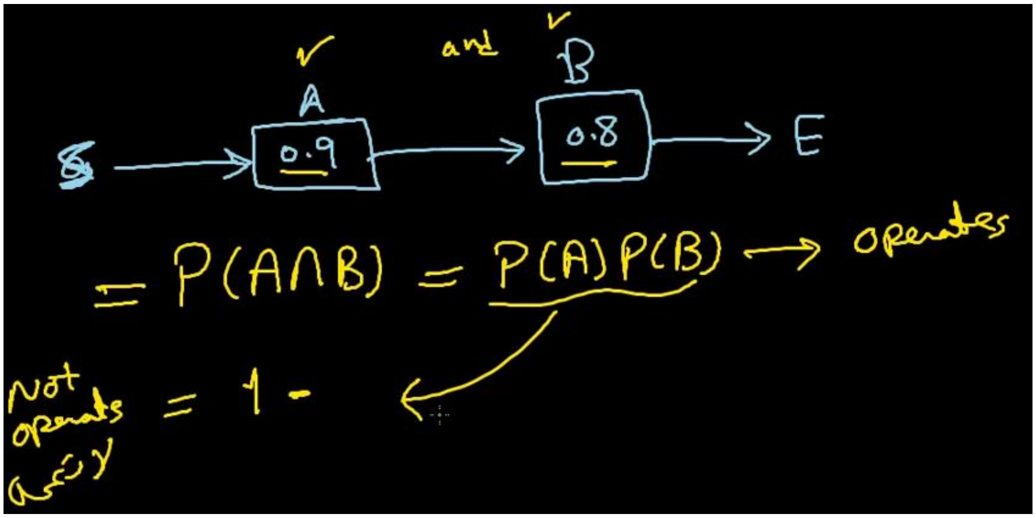
$$P(A \mid B) = 0$$

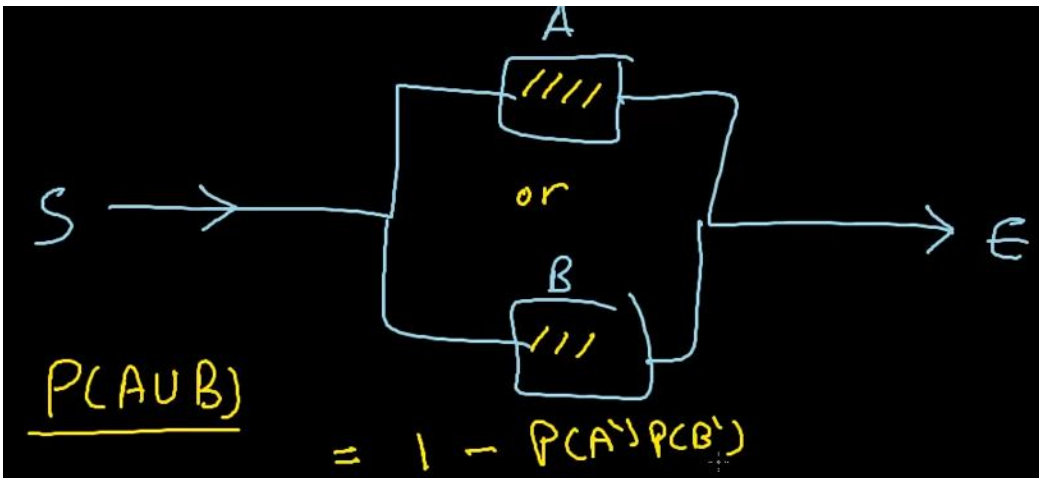
2. A, B independent

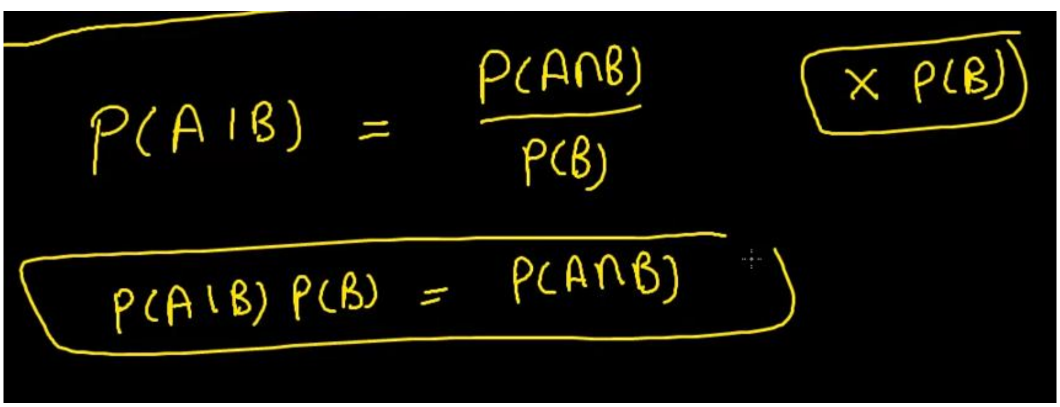
$$P(A \cap B) > 0$$

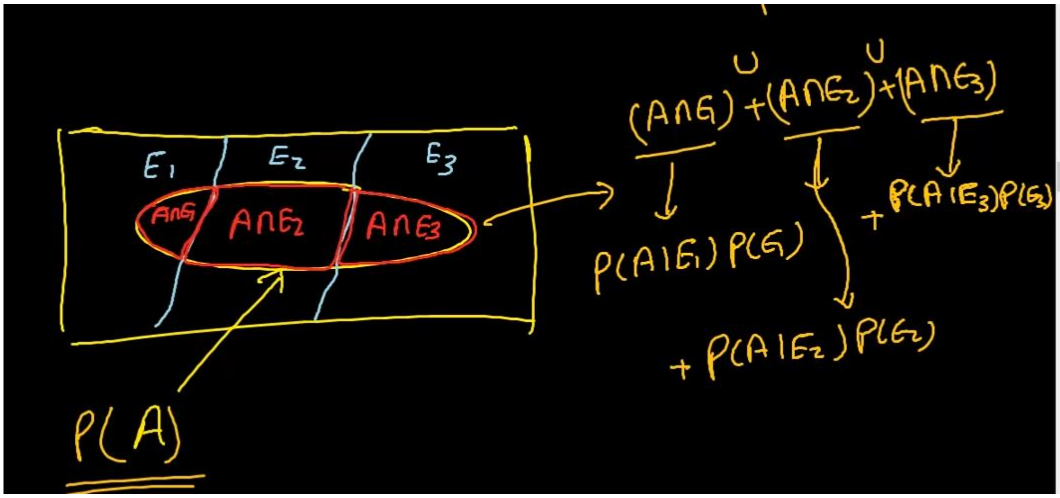
$$P(A \mid B) = P(A) \quad , \quad P(B \mid A) = P(B)$$

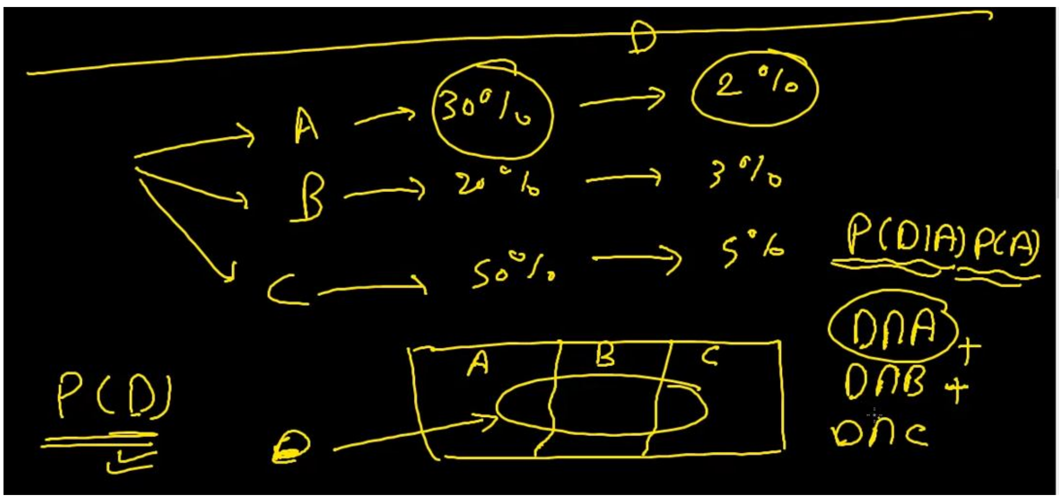
$$P(A \cap B) = P(A)\,P(B)$$

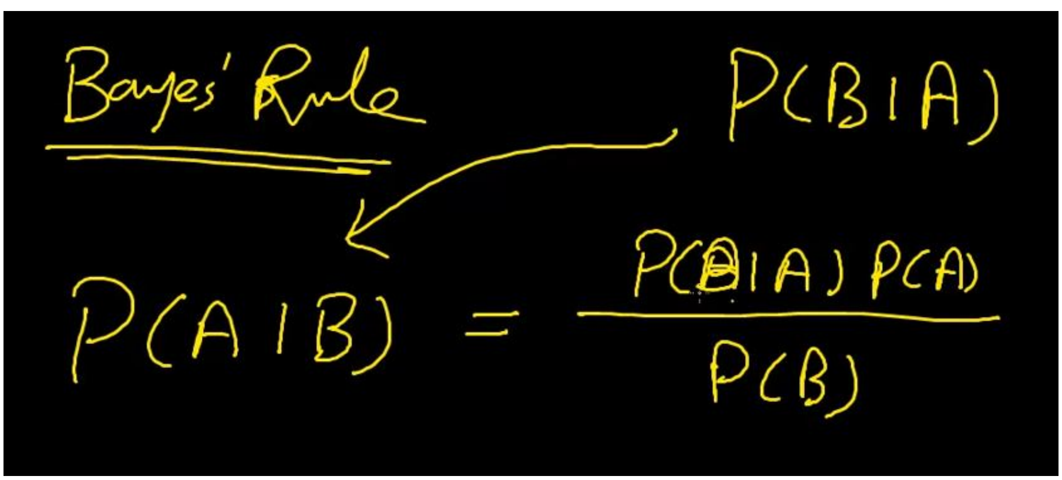

and
A
B
S
0.9
0.8
E
= P(A∩B) = P(A)P(B) → operates
Not operates = 1 -

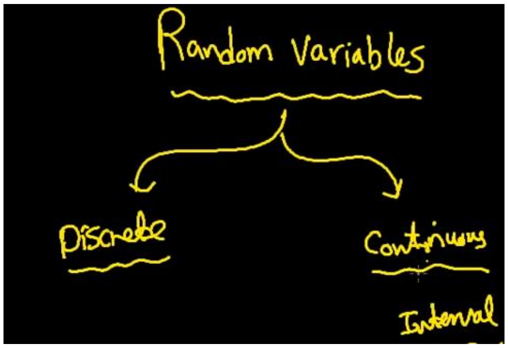

A
S
or
E
B
P(A∪B)
= 1 − P(A')P(B')


P(A|B) = P(A∩B) / P(B)
× P(B)
P(A|B) P(B) = P(A∩B)


E1
E2
E3
A∩E1
A∩E2
A∩E3
P(A)
(A∩E1) + (A∩E2) + (A∩E3)
P(A|E1) P(E1)
+ P(A|E3)P(E3)
+ P(A|E2) P(E2)

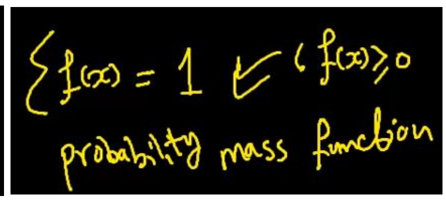

D
A
30%
2%
B
20%
3%
C
50%
5%
P(D|A) P(A)
D∩A +
D∩B +
D∩C
P(D)

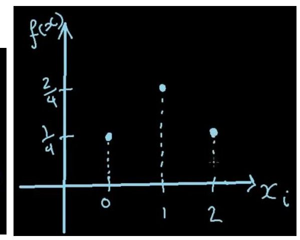

Bayes' Rule
P(B|A)
P(A|B) = P(B|A) P(A) / P(B)



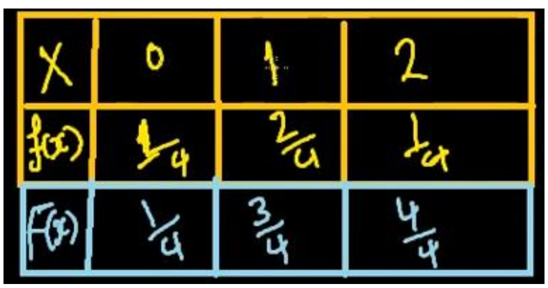
Discrete:

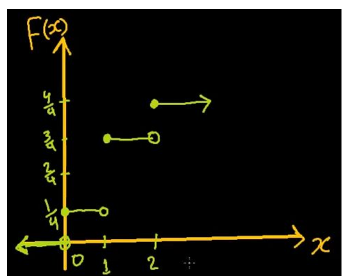
Probability mass function:





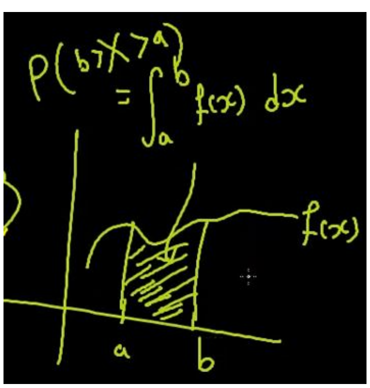
Cumulative function





Continuous: (intervals: area under curve)

Probability density function (pdf)



R.V. Mean & Variance & Standard deviation

$E(X)$ $V(X)$ $\sigma = \sqrt{V(X)}$

μ σ^2

$\bar{x}$

Discrete

$$E(X) = \sum x f(x)$$

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

$$E(X^2) = \sum x^2 f(x)$$

$$\mu = E(X) = (0)\left(\tfrac{1}{4}\right) + (1)\left(\tfrac{2}{4}\right) + (2)\left(\tfrac{1}{4}\right) = 1$$

$$E(X^2) = (0)^2\left(\tfrac{1}{4}\right) + (1)^2\left(\tfrac{2}{4}\right) + (2)^2\left(\tfrac{1}{4}\right) = 1.5$$

$$V(X) = 1.5 - (1)^2 = 0.5$$

$$\sigma = \sqrt{0.5}$$

Continuous

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

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

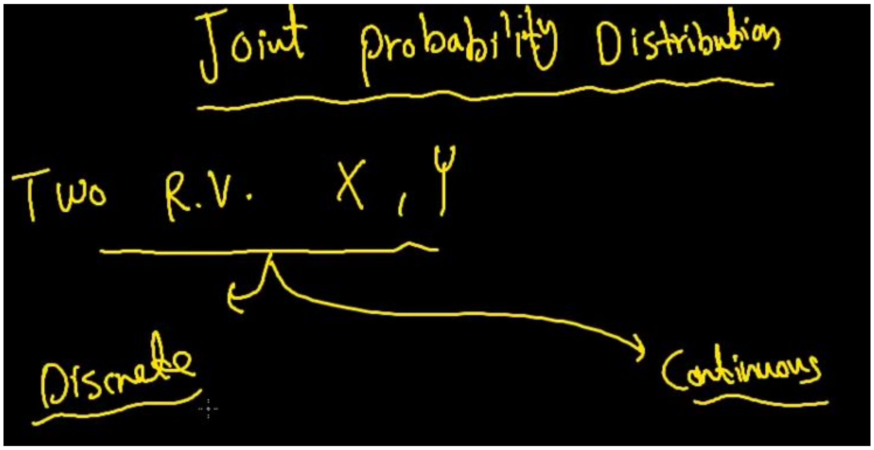
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

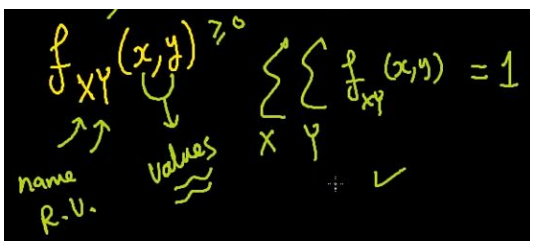
$\frac{x^2}{3}$ $\boxed{2 > x > -1}$ $E(X) = \int_{-1}^{2} x\left(\frac{x^2}{3}\right) dx$

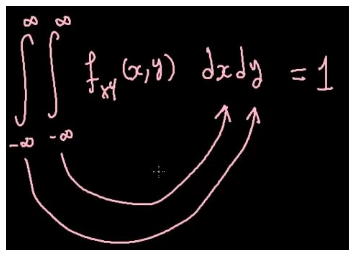
$$E(X^2) = \int_{-1}^{2} x^2\left(\frac{x^2}{3}\right) dx = \int_{-1}^{2} \frac{x^4}{3}\, dx$$

$$V(X) = \frac{33}{15} - \left(\frac{15}{12}\right)^2 = \checkmark$$

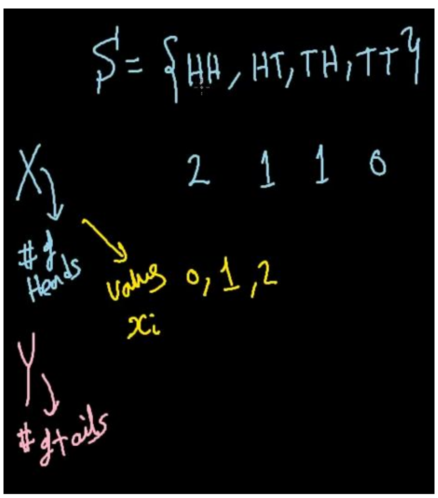
$$\sigma = \sqrt{V(X)}$$

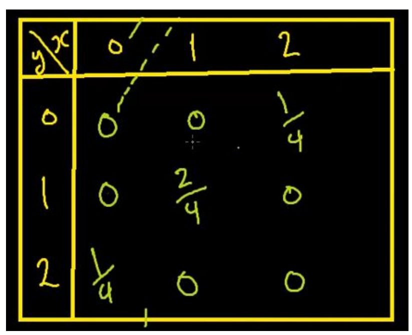


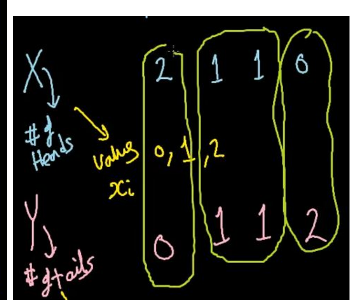


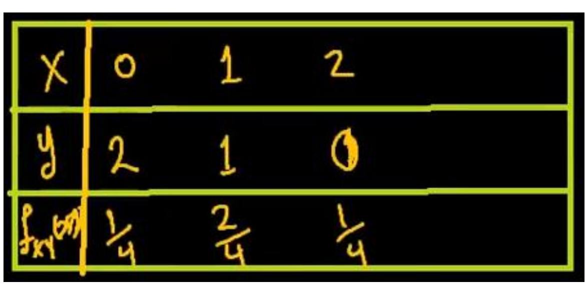


Discrete example









Continuous example

$$f_{XY}(x,y) = \frac{2}{5}(2x+3y) \qquad 0 \le x \le 1, \quad 0 \le y \le 1$$

$$\int_0^1 \int_0^1 \frac{2}{5}(2x+3y)\,dx\,dy$$

$$\int_0^1 \frac{2}{5}\left(x^2+3yx\right)\Big|_{x=0}^{x=1} dy$$

$$\int_0^1 \frac{2}{5}(1+3y)\,dy$$

$$= \frac{2}{5}\left(y+\frac{3y^2}{2}\right)\Big|_{y=0}^{y=1}$$

=1 >> it is joint probability density function

If x and y is constants not interval

$x=1$ $y=2$

$$f_{XY}(1,2) = 0$$

$$=0 \qquad \int_0^0\int_0^0 dx\,dy$$

$$P(x_1 > X > x_2,\ y_1 > y > y_2$$

Marginal prob. distribution

Discrete

$$f_X(x) = \sum_y f_{XY}(x,y)$$

$$f_Y(y) = \sum_x f_{XY}(x,y)$$

Continuous

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx$$

$y \backslash x$	1	1.5	2.5	3
1	1/4	0	0	0
2	0	1/8	0	0
3	0	1/4	0	0
4	0	0	1/4	0
5	0	0	0	1/8
$f_X(x)$	1/4	3/8	1/4	1/8

$f_X(x)$ $\sum_y$

x	1	1.5	2.5	3
$f_X(x)$	1/4	3/8	1/4	1/8

$$f_{XY}(x,y) = \frac{2}{5}(2x+3y)$$

$$1 \geq x \geq 0$$

$$1 \geq y \geq 0$$

$$f_X(x) = \int_0^1 \frac{2}{5}(2x+3y)\, dy$$

$$= \frac{2}{5}\left(2xy + \frac{3y^2}{2}\right)_{y=0}^{y=1}$$

$$= \frac{2}{5}\left[\left(2x + \frac{3}{2}\right) - 0\right]$$

$$\boxed{f_X(x) = \frac{4}{5}x + \frac{3}{5}}$$

Mean & Variance

$$E(X) = \sum_{x} x \, f_X(x)$$

$$E(Y) = \sum_{y} y \, f_Y(y)$$

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

$$E(Y) = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy$$

x	1	1.5	2.5	3
$f_X(x)$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{8}$

$$E(X) = (1)\left(\tfrac{1}{4}\right) + (1.5)\left(\tfrac{3}{8}\right) + (2.5)\left(\tfrac{1}{4}\right) + (3)\left(\tfrac{1}{8}\right)$$

$$E(X) = \int_{0}^{1} x\left(\tfrac{4}{5}x + \tfrac{3}{5}\right) dx$$

$$V(X) = E(X^2) - \left(E(X)\right)^2$$

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

2. There are 3 arrangements of the word DAD, namely DAD, ADD, and DDA. How many arrangements are there of the word PROBABILITY?

3 arrangements of the word DAD (DAD, ADD, DDA), How many arrangements are there of the word PROBABILITY?

11 character, B and I doubled >> 11!/2!x2!

4. There are six men and seven women in a ballroom dancing class. If four men and four women are chosen and paired off, how many pairings are possible?

6 men and 7 women, if 4 men and 4 women are chosen and paired off, how many pairing are possible?

4- 6M and 7W

we choose 4M and 4M

pair them

$$\binom{6}{4} \times \binom{7}{4} \times 4!$$

6. Suppose that there are ten students in a classroom. What is the probability that no two of them have a birthday in the same month?

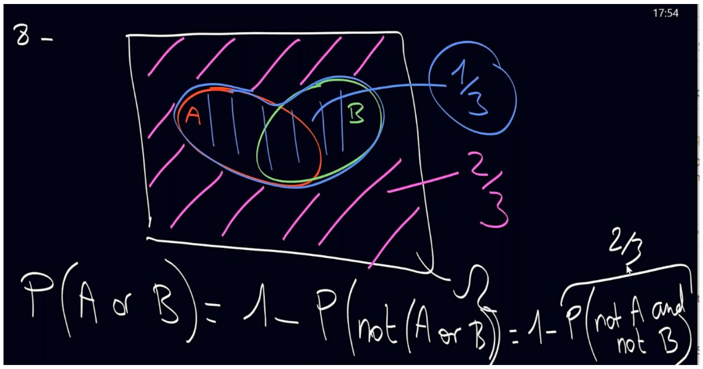
6- 10 students

P(no 2 of them have birthday the same month)

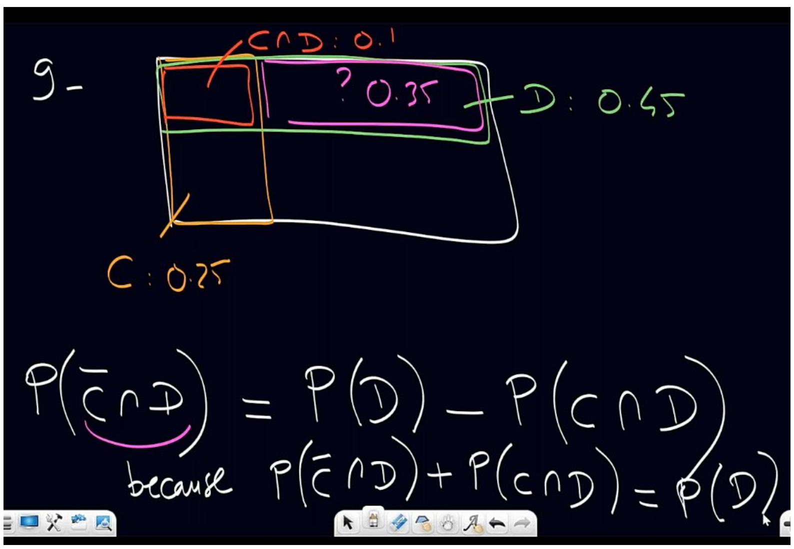
$$\frac{12}{12}\, 1 \times \frac{11}{12} \times \frac{10}{12} \times \frac{9}{12} \times \cdots \times \frac{3}{12} = \frac{\frac{12!}{2!}}{12^{10}}$$

guy 1 2 3 4 10

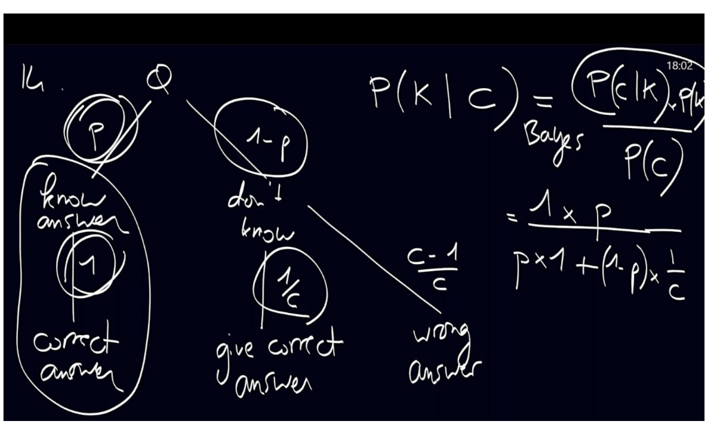
8. Let A and B be two events. Suppose the probability that neither A or B occurs is $2/3$. What is the probability that one or both occur?



9. Let C and D be two events with $P(C) = 0.25$, $P(D) = 0.45$, and $P(C \cap D) = 0.1$. What is $P(C^c \cap D)$?



14. Suppose you are taking a multiple-choice test with c choices for each question. In answering a question on this test, the probability that you know the answer is p. If you don't know the answer, you choose one at random. What is the probability that you knew the answer to a question, given that you answered it correctly?

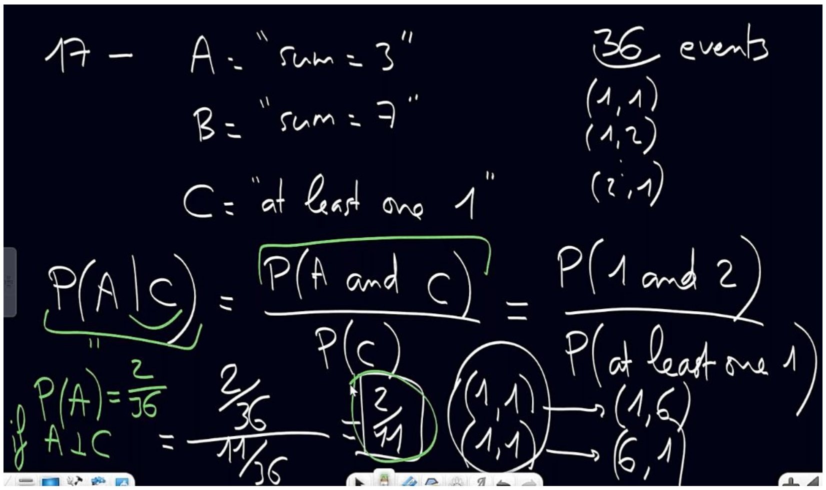


17. Two dice are rolled.
A = 'sum of two dice equals 3'
B = 'sum of two dice equals 7'
C = 'at least one of the dice shows a 1'

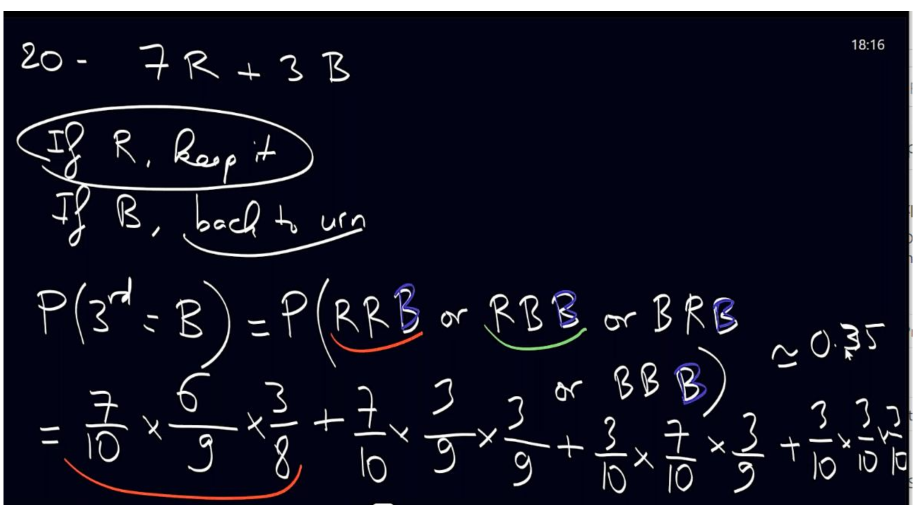
(a) What is $P(A|C)$?

(b) What is $P(B|C)$?

(c) Are A and C independent? What about B and C?



20. Suppose you have an urn containing 7 red and 3 blue balls. You draw three balls at random. On each draw, if the ball is red you set it aside and if the ball is blue you put it back in the urn. What is the probability that the third draw is blue?
(If you get a blue ball it counts as a draw even though you put it back in the urn.)



21. Some games, like tennis or ping pong, reach a state called *deuce*. This means that the score is tied and a player wins the game when they get *two* points ahead of the other player. Suppose the probability that you win a point is p and this is true independently for all points. If the game is at deuce what is the probability you win the game?

18:19

21 tennis, ping-pong, ... → deuce

you win if you win 2 consecutive points

$P(\text{win a point}) = p$

$P(\text{win match}) = ?$

WW → win match
LL → lose match
WL, LW → back to deuce

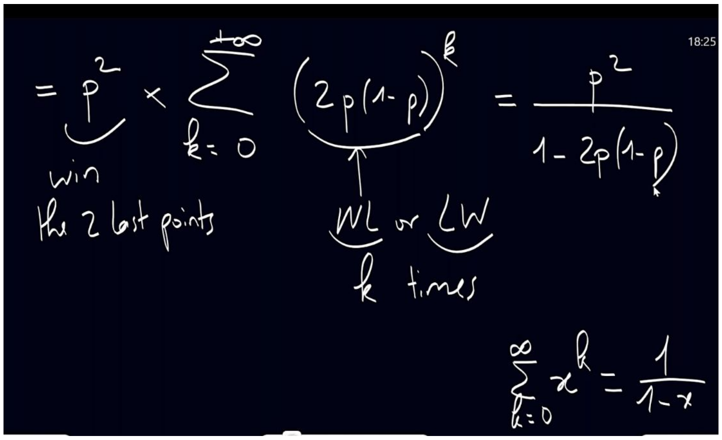
18:22

$$P(\text{win match}) = P\big(WW \text{ or } (LW \text{ and win}) \text{ or } (WL \text{ and } \underline{\text{win}})\big)$$

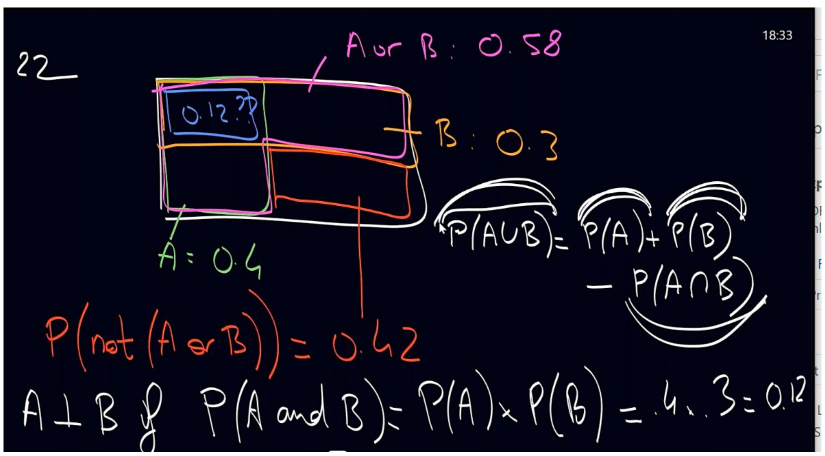
$$= p\times p + 2(1-p)\times p\times(\underbrace{\text{win}}_{=\,p^2+2(1-p)p\times \text{win}})$$

$$= p^2 + 2p(1-p)\Big(p^2 + 2p(1-p)\big(p^2 + 2p(1-p)(p^2 + \cdots)\big)\Big)$$

$$= p^2 + p^2\,2p(1-p) + p^2\big(2p(1-p)\big)^2 + p^2\big(2p(1-p)\big)^3 + \cdots$$



22. Suppose that $P(A) = 0.4, P(B) = 0.3$ and $P((A \cup B)^C) = 0.42$. Are A and B independent?

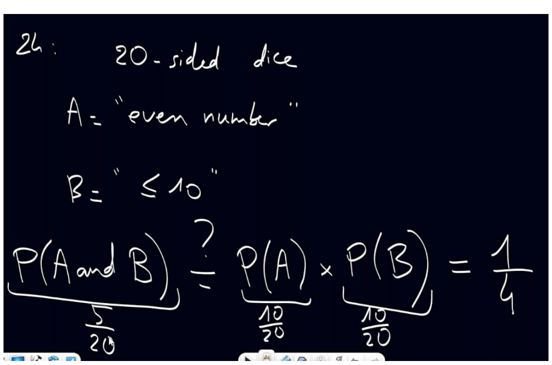


24. You roll a twenty-sided die. Determine whether the following pairs of events are independent.

(a) 'You roll an even number' and 'You roll a number less than or equal to 10'.

(b) 'You roll an even number' and 'You roll a prime number'.

Prime numbers = {2,3,5,7,11,13,17,19}



25. Suppose A and B are events with $0 < P(A) < 1$ and $0 < P(B) < 1$.

(a) If A and B are disjoint can they be independent?

(b) If A and B are independent can they be disjoint?

(c) If $A \subset B$ can they be independent?

If A and B disjoint

$A \cap B = \emptyset \rightarrow$ are they independent?

$$0 = P(A \text{ and } B) = \underbrace{P(A)}_{>0} \times \underbrace{P(B)}_{>0}$$

The cross not equal to the product so they cannot be independent.

If A and B independent

Then the cross is greater than 0, so they cannot be disjoint.

$$0 = \underbrace{P(A \text{ and } B)}_{>0}$$

$A \subset B$

$A \perp B$? $\quad P(\underbrace{A \cap B}_{A}) = P(A) \times P(B)$

cannot be true because $P(B) < 1$

26. Directly from the definitions of expected value and variance, compute $E(X)$ and $\mathrm{Var}(X)$ when X has probability mass function given by the following table:

X	-2	-1	0	1	2
p(X)	1/15	2/15	3/15	4/15	5/15