EE2012

in 4dp or in fraction, no need pad 0 if no dp

小赌怡情，大赌伤身

Chapter 1 - Random Experiments and Probability.....5
Mathematical Models.....5
Deterministic Models.....5
Probabilistic Models.....5
Terminology of Probability.....5
Probability experiment.....5
Sample Space (Universal Set).....5
Event.....5
Probability of an event.....5
Types of Sample Space.....6
Finite Discrete.....6
Countable Infinite Discrete.....6
Continous (Uncountable Discrete).....6
Set Notation for Probability.....6
Partition.....7
Probability.....7
Definition.....7
From GEA1000.....8
Why Probability Exists.....8
Axioms.....8
Example.....9
Properties of Probability.....10
Counting Problems.....10
With replacement and with ordering.....11
Example 1:.....11
Example 2: Password.....11
Without replacement and with ordering (Permutation).....11
Example.....12
With replacement and without ordering.....12
Example 1: Allocation Problem (Voting Problem in Disguise).....12
Example 2: Voting Problem.....13
Without replacement and without ordering.....13
Example 1.....14
Example 2.....14
Circular Permutation.....15
Chapter 2: Conditional Probability and Bayes' Rule.....18
Conditional Probability.....18
Properties of Conditional Probability.....18
Example 1.....19

Example 2....19
Independent and Disjoint....20
Monty Hall Problem....20
Total probability....22
Example 1....22
Example 2....23
Bayes’ Rule....24
Example 1....24
Example 2....25
Example 3....25
Example 4....26
Sequential experiment....26
Independent Sequential Experiment....26
Example....27
Dependent Sequential Experiment....27
Example....28
Example 2: Gambler's ruin....29
Chapter 3 - Discrete Random Variables and Distributions....30
Example....31
Probability Mass Functions (PMF)....32
Forms of PMF....33
Expectation of Random Variables....33
Expectation, Median and Mode....34
Properties of Expectation....34
Example 1....35
Example 2....35
Example 3....36
Variance of Discrete RVs....36
Properties of Variance....37
Example 1....37
Example 2....38
Example 3....38
Summary for Common Discrete Random Variables....40
Discrete Uniform Distribution....41
Derivation for Variance....42
Bernoulli Distribution....43
Binomial Distribution....43
Example 1....44
Example 2....45
Geometric Distribution....45
Example....46
Derivation for E[X]....47
Derivation for Var[X]....48
Poisson....49
Derivation for E[X]....50
Derivation for Var[X]....51
Example 1....51
Example 2....52

Cumulative Distribution Functions (CDF).......... 53
Example.......... 55
Gambling Example.......... 55
Chapter 4 Multiple Random Variables and Functions of Random Variables.......... 58
Multiple discrete RVs.......... 58
Mapping.......... 59
Two discrete RVs.......... 59
Example 1.......... 60
Marginal PMF.......... 61
Joint PMF → Marginal PMFs.......... 61
Marginal PMFs → Joint PMF.......... 62
Example.......... 62
Independence of Two RVs.......... 63
Proof.......... 64
Example 1.......... 64
Example 2.......... 64
Covariance and Correlation Coefficient.......... 65
Proof for -1 <= ϱ <= 1.......... 66
Example 1.......... 67
Example 2.......... 68
Functions of one RV.......... 68
Example 1.......... 69
Example 2.......... 70
Example 3.......... 71
General Case for Function of One RV.......... 71
If the function is invertible.......... 71
If the function is not invertible.......... 71
Functions of two RVs.......... 72
Case of Independent RVs.......... 72
Example.......... 73
Discrete Convolution Method.......... 74
Example 2.......... 75
Expectation of RV Sum.......... 75
Proof.......... 76
Variance of RV Sum.......... 77
Proof.......... 78
Sample Average Example.......... 79
Chapter 5: Continuous Random Variables.......... 79
Probability Density Function (PDF).......... 80
Example 1.......... 81
Example 2.......... 81
Example 3.......... 82
Example 4.......... 82
Cumulative Distribution Function (CDF).......... 83
Properties of CDF.......... 84
Example 1.......... 84
PDF and CDF relationship.......... 85
Example.......... 86

Expectation of Continuous RVs....87
Integration Tricks....88
Proof for Tricks....88
Variance of Continuous RVs....89
Summary for Common Continuous RVs....91
Continuous Uniform RV....93
Relationship Between Poisson and Exponential Distribution....94
Gaussian RV....97
Proving Area under Gaussian PDF Sums to 1....99
Standard Gaussian....100
Q-Function Table....101
Converting Arbitrary Gaussian to Standard Gaussian....1
Example....1
Example 2....1
Sum of Independent Gaussian RVs....1
Chapter 6: Multiple Continuous Random Variables....1
Sum of Continuous Independent RVs....1
Sum of Independent Gaussian RVs....1
Sum of Identical and Independent (iid) Gaussian RVs....1
Central Limit Theorem (CLT)....1
Example....1
Continuity Correction....1
Example 1....1
Example 2....1
Summary for Gaussian Approximation....1
Multiple Continuous RVs....1
Finding CDF from PDF for non rectangular domain....1
Expectation....1
Covariance and Correlation Coefficient....1
Example....1
Bivariate Gaussian RVs....1
Proof....1
Bivariate Independent Gaussian RVs....1
Complex Gaussian RVs....1
Circularly Symmetric Complex Gaussian RVs....1
Meaning of Circularly Symmetric....1
Marginal Distribution of Random Variable R and Θ in Polar Version of Circularly Symmetric Complex Gaussian....1
Tutorial....1
Tutorial 1....1
Tutorial 2....1
Tutorial 5....1
Tutorial 6....1
Midterm....1
Quiz 1....1
Quiz 2....1
Quiz 3....1
Quiz 4....1

Quiz 5 1
other stuff 1
Admin Stuff 1

Chapter 1 - Random Experiments and Probability

Mathematical Models

Deterministic Models

- Given an input, can predict the outcomes.
- Example: equation modelling physics like Ohm's law

Probabilistic Models

- Random experiment model with random outcomes
- Example: Probabilistic model can be used to check the probability of radar detection, false alarm

Terminology of Probability

	Example A: 2 coin tosses	Example B: d6 toss
Probability experiment A probability experiment must be repeatable and allows for the exact listing of all the possible outcomes.	The procedure of tossing the coin twice is called a probability experiment.	The procedure of tossing the die is called a probability experiment.
Sample Space (Universal Set) A sample space is the set of all possible outcomes of a probability experiment. S = $\{\zeta_1, \zeta_2, \zeta_3, \ldots\}$ can be finite or infinite set Different outcomes are mutually exclusive or disjoint (ζ_i and ζ_j will not occur concurrently) Consideration of special cases to include into sample space depends on the situation	{HH, HT, TH, TT} eg: coin landing straight up	{1, 2, 3, 4, 5 ,6}
Event A subset of the sample space is called an event. Note that an outcome can also be considered an event but not all events are outcomes. This is clear as there exist subsets of only one element but not all subsets have only one element. for A = {1,2}, when A occurs means 1 or 2 occurs, not both, as they are mutually exclusive	"two in a row" = {HH, TT} "at least one tail" = {HT, TH, TT} "The first toss is heads, second toss is tails." = {HT}	"die shows an even number" = {2, 4, 6}
Probability of an event The probability of an event of the sample space is the total probability that the	Probabilities are numerical values between 0 and 1 (both inclusive), so P(E) takes on a numerical value between 0 and 1	

outcome of the experiment is an element of the event. E = event S = sample space where $E \subseteq S$ $P(E) = P(x \in E)$	and this is the probability assigned to event E.

Types of Sample Space

Finite Discrete

- Coin is tossed once S = {H, T}
- Coin is tossed twice S = {HH, HT, TH, TT}
- Number of heads in three tosses S = {0, 1, 2, 3}
- When the sample space contains only a finite number of outcomes, we only need to assign probabilities to the outcomes so that these probabilities sum up to 1. The probabilities of all other events can then be derived from there.

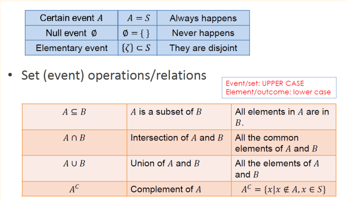
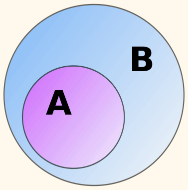
Countable Infinite Discrete

- Number of tosses until the first head appears S = {1, 2, 3, …}
  - Goes to infinity but countable as each can be mapped to integer

Continous (Uncountable Discrete)

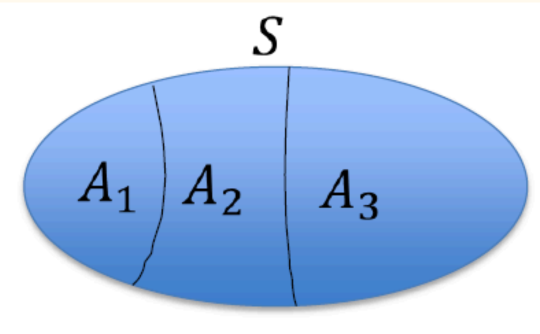
- Randomly pick a real number X between 0 and 1 $S = \{x \mid 0 < x < 1, x \subseteq R\}$
- Randomly pick a real number Y between 0 and X $S = \{y \mid 0 < y < X, y \subseteq R\}$
- infinite amount of number between two numbers, uncountable

Set Notation for Probability

<table>
<tr><th>Terms</th><th>Notations</th><th>Meaning</th></tr>
<tr><td>Certain Event, A
- certain event means event that always happens, not a particular event, the event must be the sample space</td><td>A = S</td><td>Always happens</td></tr>
<tr><td>Null event
- empty set with no outcomes, not {0}</td><td>ø = {}</td><td>Never happens</td></tr>
<tr><td>Elementary event
- set with only one outcome</td><td>$\{\zeta\} \subset S$</td><td>Elementary event are disjoint

<table>
<tr><td>Certain event A</td><td>$A = S$</td><td>Always happens</td></tr>
<tr><td>Null event Ø</td><td>$\emptyset = \{\ \}$</td><td>Never happens</td></tr>
<tr><td>Elementary event</td><td>$\{\zeta\} \subset S$</td><td>They are disjoint</td></tr>
</table>
• Set (event) operations/relations
Event/set: UPPER CASE
Element/outcome: lower case
<table>
<tr><td>$A \subseteq B$</td><td>A is a subset of B</td><td>All elements in A are in B.</td></tr>
<tr><td>$A \cap B$</td><td>Intersection of A and B</td><td>All the common elements of A and B</td></tr>
<tr><td>$A \cup B$</td><td>Union of A and B</td><td>All the elements of A and B</td></tr>
<tr><td>A^C</td><td>Complement of A</td><td>$A^C = \{x \mid x \notin A, x \in S\}$</td></tr>
</table>
</td></tr>
<tr><td>A is a subset of B</td><td>$A \subseteq B$</td><td>All elements in A are in B and A can be equal to B</td></tr>
<tr><td>A is a proper subset of B

</td><td>$A \subset B$</td><td>All elements in A are in B and A cannot be equal to B</td></tr>
</table>

Intersection of A and B - if A ∩ B = ø then A and B are disjoint	A ∩ B = AB	All the common elements of A and B
Union of A and B	A U B = A + B	All the elements of A and B
Complement of A	$A^C = \bar{A}$	

Partition



$$S = \bigcup_{i=1}^{n} A_i \quad for\ disjoint\ set\ A_1,\ A_2,\ ...,\ A_n$$

- If the union of disjoint sets is the sample space then they form a partition of S, if doesnt union up to the whole S it's not a partition.
- MUST BE DISJOINT

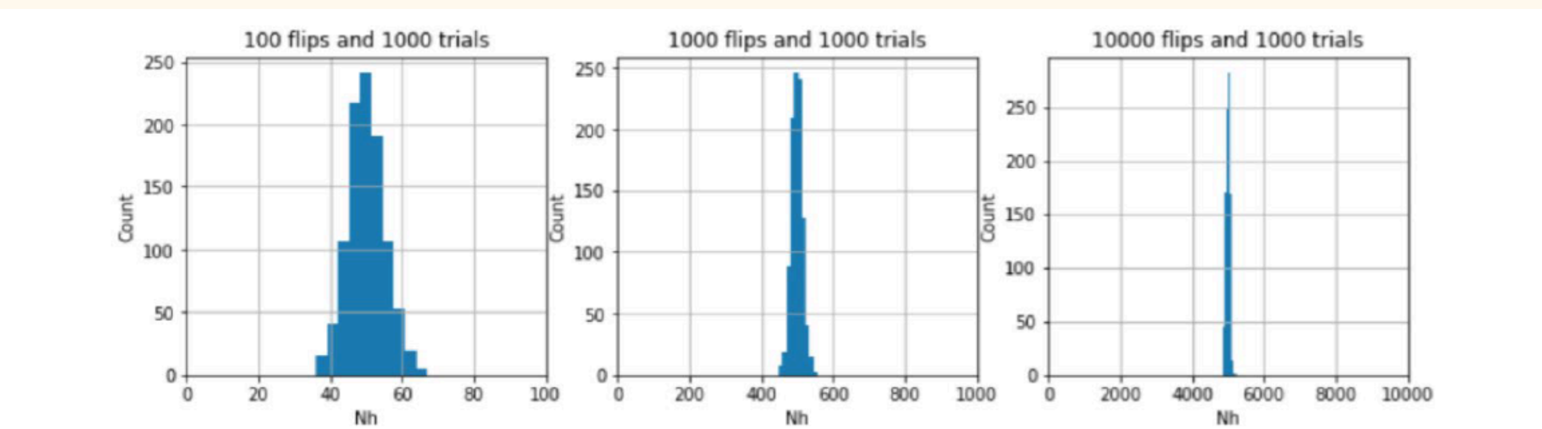
Probability

Definition

$$f_A(n) = \frac{N_A}{n} \quad for\ event\ A$$

- Probability is defined based on relative frequency, the ratio of occurrence of events and total number of experiments conducted.
- Positive real numbers

$$P(A) \lim_{n \to \infty} f_A(n)$$



- With high number of experiment, relative frequency approaches the probability
    - this is impractical for determining probability of certain outcomes like atomic reactions and elections as cannot repeat many times

    - Another way of determining probability is defining event of interest and calculate the probability if sufficient information is available

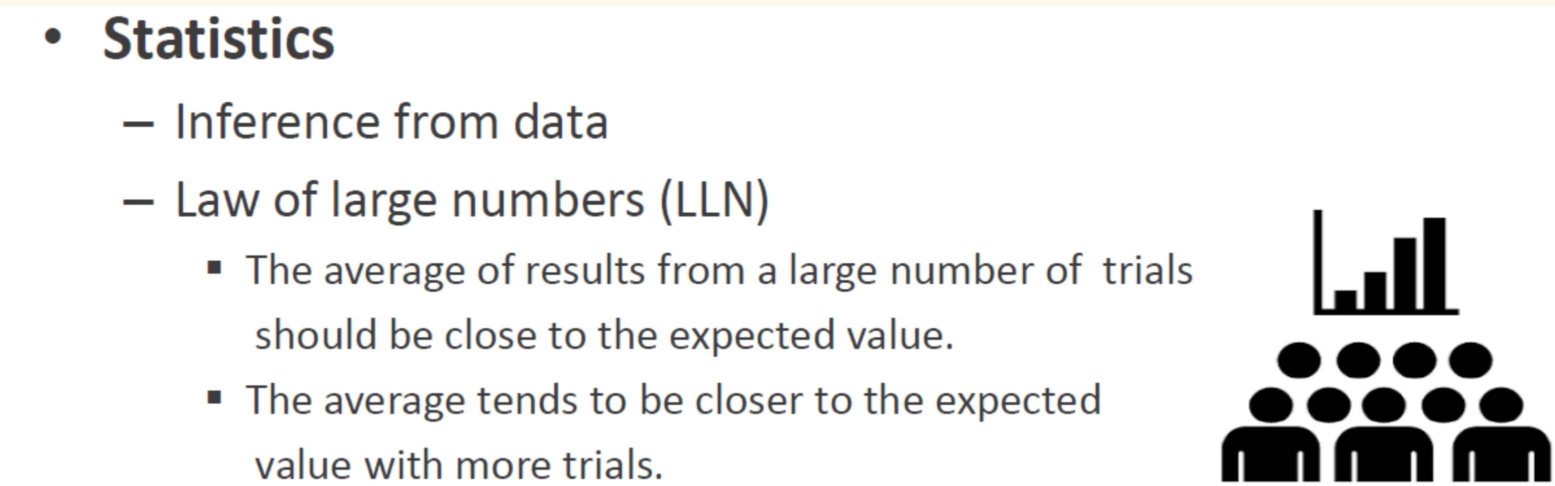
<u>From GEA1000</u>

Mathematically, P(E) is defined as the long run proportion of observing E when a large number of repetitions of the experiment is being performed. P(E) = n(E)/N

- The estimate of P(E) we obtain from these N repetitions of the experiment is likely to be different if the experiment is repeated another N times and to get another estimate.
- Such estimates get more accurate and closer to the true value of P(E) as N approaches infinity.
- Thus it is virtually impossible to verify what is the true probability for an event of a probability experiment. But in the analysis of data, it is sufficient to treat the estimates as if it is the true probability.
- It is more important for the to obey the Rules of Probabilities:
    - The probability of each event E, denoted by P(E) is a number between 0 and 1 (inclusive).
    - If the entire sample space is denoted by S, then the probability of S, P(S) is 1.
    - If E and F are mutually exclusive events (meaning both events cannot occur simultaneously), then the probability of E or F occuring is equal to the sum of the probabilities of E and F. That is,
        - if $P(E \cap F) = 0$ then $P(E \cup F) = P(E) + P(F)$

<u>Why Probability Exists</u>

- Probability is either because insufficient information (if know everything, everything can be calculated and predicted) or nature of the phenomena is random eg: some quantum shits
- Statistics is the inference from data using Law of Large Numbers
    - The average of results from a large number of trials should be close to the expected value.



<u>Axioms</u>

For events $A \subseteq S$ and $B \subseteq S$

1. [Nonnegativity] $P(A) \geq 0$
2. [Normalisation] $P(S) = 1$
3. [Additivity] if A and B are mutually exclusive (disjoint), $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) \quad for\ disjoint\ event\ A_1,\ A_2,\ ...,\ A_n$$

<u>Example</u>

If a fair coin is flipped, the head and tail appear with equal chance. $S = \{H,T\}$. $P(\{H\})$ and $P(\{T\})$ should be assigned the same probability.

- In addition, the additivity implies $P(S) = P(\{H\}) + P(\{T\}) = 1$.
- By the normalization law, $P(\{H\}) = P(\{T\}) = 0.5$.
- The probability assignment is

  $P(\{H,T\}) = 1, P(\{H\}) = 0.5, \quad P(\{T\}) = 0.5, \quad P(\emptyset) = 0.$

- Assume the coin is flipped three times. The event $A = \{2 \text{ heads occur}\} = \{HHT,HTH,THH\}$. Since these outcomes are disjoint,

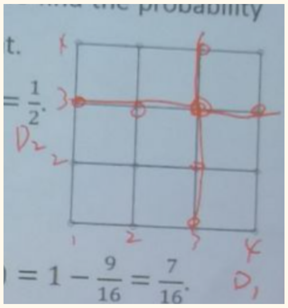
  $P(A) = P(\{HHT\}) + P(\{HTH\}) + P(\{THH\}) = \frac{3}{8}$ by additivity.

P({H}) = P(H) means the same thing but by right should be P({H})

A pair of 4-sided dice, one red and one blue, are rolled. The sides of the dice are marked by different number in $\{1,2,3,4\}$. Since each dice can have 4 sides, there are 16 possible outcomes. Assume both dice are fair, which means all the possible outcomes have the same probability $p = 1/16$. To find the probability of an event, we need to find out the outcomes in the event.

- $P(\{\text{the sum is even}\}) = P(\{\text{the sum is odd}\}) = \frac{8}{16} = \frac{1}{2}$.
- $P(\{\text{the red is equal to the blue}\}) = \frac{4}{16}$.
- $P(\{\text{the red is greater than the blue}\}) = \frac{6}{16} = \frac{3}{8}$.
- $P(\{\text{at least one dice is 4}\}) = 1 - P(\{\text{no 4 for both}\}) = 1 - \frac{9}{16} = \frac{7}{16}$.
- $P(\{\text{the sum is 4}\}) = \frac{3}{16}$.

can use grid to find, a 2d representation of the sample space



P(min(R, B) = 3) = P(3, 3) + P(4, 3) + P(3, 4) = 3/16
P(max(R, B) = 3) = 2(P(1, 3) + P(2, 3)) + P(3, 3) = 5/16

* in exam, simplify fraction

A fair coin is flipped three times. Find the probability of 3 heads.

- M1
  - In a sequential experiment of three independent steps, the coin is flipped once in each step.
  - In each step, head occurs with probability 0.5.
  - All three steps must observe head to have 3 heads in total. The probability of the desired event is $0.5 \times 0.5 \times 0.5 = 0.125$.
- M2
  - Assume three fair coins are flipped together and independently. Observe the number of heads.
  - All the possible cases appear with the same but unknow probability p.
  - There are total $2 \times 2 \times 2$ outcomes and only one leads to the desired event.
  - The desired event probability is $\frac{1 \times p}{2 \times 2 \times 2 \times p} = 0.125$.

There are 3 red marbles and 5 blue marbles in a bag. Find the probability of get one red and one blue marbles in two picks (without putting back).

- M1
  - In a sequential experiment of two steps, one marble is taken out in each step.
  - The desired event $A = \{R_1 B_2, B_1 R_2\}$.
  - $P(A) = P(R_1 B_2) + P(B_1 R_2) = \frac{3}{8} \times \frac{5}{7} + \frac{5}{8} \times \frac{3}{7} = \frac{15}{28}$.
- M2
  - Consider two marbles are picked together. Any two marbles can be picked with equal probability (whatever their colours are).
  - Total number of cases is $8 \times 7 = 56$ and the number of desired cases is $2 \times \binom{3}{1} \times \binom{5}{1} = 30$.
  - The desire event probability is $\frac{30}{56} = \frac{15}{28}$.

implicitly using conditional probability, P(R1B2) = P(B2|R1) P(R1)
2 for ordering, C for choosing

Properties of Probability

Let A and B be two events

- $P(A^C) = 1 - P(A)$
- $0 \le P(A) \le 1$
- $P(\emptyset) = 0$
- if $A \subseteq B$, then $P(A) \le P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- if $A \cap B = \emptyset$ then A and B are disjoint
  - $P(A \cup B) = P(A) + P(B)$
- If event A and B are independent, $P(A|B) = P(A)$, thus $P(A \cap B) = P(A) P(B)$

Counting Problems

- There are 4 types of counting problems:
  - With replacement and with ordering
  - Without replacement and with ordering
  - With replacement and without ordering
  - Without replacement and without ordering
- Order means the order of taking out the items

- When outcomes are equiprobable, the probability can be found by counting the total possible outcomes and the total outcomes that we want.

$$P(A) = \frac{|A|}{|S|} \quad where\ |A| = size\ of\ set\ A,\ |S| = size\ of\ sample\ space$$

*size of set = cardinality

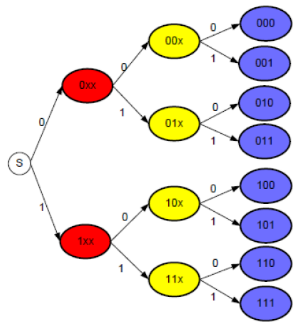
With replacement and with ordering

- After each pick, the information of the item picked is recorded and placed back in to the collection of items
- everytime has n choices

Example 1:

How many binary sequence of length 3 from set {0, 1}? What is the probability for a repetitive sequence?

- $|\{0,1\}| = 2$
- $2 \times 2 \times 2 = 8$ and each sequence is picked with probability $\frac{1}{8} = 0.125$.
- Only two repetitive sequences 000 and 111. The probability to pick a repetitive sequence is $P = \frac{2}{8} = 0.25$.



Tree diagram

- Each choice has 2 outcomes and three choices are made, thus |S| = 2*2*2 = 8
- Since there is only 2 possible repetitive sequences, 000 and 111, probably of a repetitive sequence = 2/8 = 1/4

Example 2: Password

Alice needs to pick a 4-digit password. Assuming any 4-digit sequence will be picked with equal probability, what is the probability that Alice picks a 4-digit password starting with nonzero digit and no repeat numbers.

- Each choice has 10 outcomes and four choices are made, thus |S| = 10^4

$$S = \{a_1 a_2 a_3 a_4 |\ a_1,\ a_2,\ a_3,\ a_4\ \in \{0, 1,\ ...,\ 9\}\}$$

- First digit cannot be 0 thus it has 9 possible outcomes, no repeating numbers means subsequent choice will have decreasing possible outcomes, but note that the second digit chosen will have 9 outcomes due to 0 being an outcome again.

$$|A| = 9 \times 9 \times 8 \times 7 = 4536 \quad where\ A\ set\ of\ desired\ sequences$$

$$\Rightarrow P(A) = \frac{|A|}{|S|} = 4536/10^4 = 0.4536$$

Without replacement and with ordering (Permutation)

- After each pick, the information of the item picked is recorded and not replaced back into the collection of items
- Each pick reduces the number of choices
- Picking k items from n distinct items without replacement and with ordering is a permutation

$$number\ of\ outcomes\ = n(n-1)(n-2)\ldots(n-(k-1))$$
$$= \frac{n(n-1)(n-2)\ldots 1}{(n-k)(n-(k+1))\ldots 1}$$
$$= \frac{n!}{(n-k)!}$$
$$= nPk \quad where\ k \le n$$

Example

Alice, Bruce, Cindy and Daniel sit in a row of four seats at random. What is the probability Alice and Cindy do not sit next to each other?

$$|S| = 4p4 = 4! = 24$$

- Find the number of outcomes for which AC are together
  - Group AC together
  - AC it self can permute and total will have 3 items permuting

$$|A| = no\ in\ a\ group! \times no\ in\ total! = k! \times (n-k+1)!$$
$$|A| = 2! \times 3!$$
$$\Rightarrow P(\overline{A}) = 1 - P(A) = 1 - \frac{2!\times 3!}{24} = 0.5$$

- if round table, will have duplicates!

With replacement and without ordering

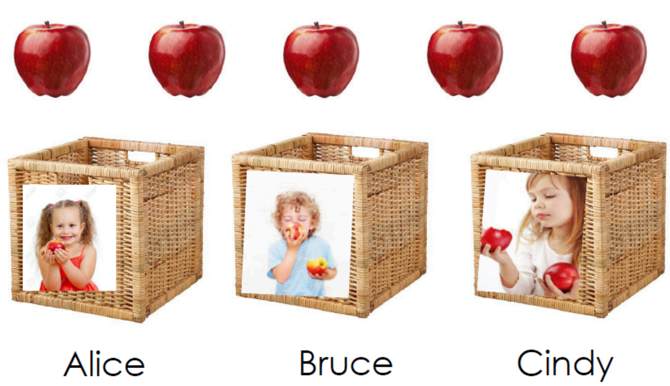
- After each pick, the information of the item picked is recorded without order and replaced back into the collection of items.
- Voting Problem: allocating k votes to n candidates with 0 allocation possible

$$Number\ of\ ways\ = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$$

  - k votes separates with n-1 separators, total position including separator = n+k-1
  - choose n-1 separators or choose k votes

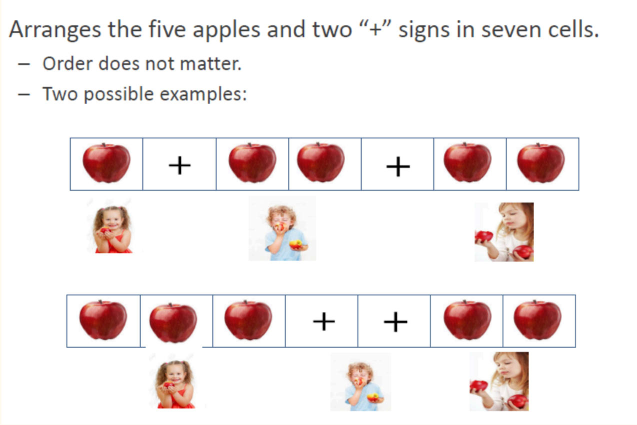
Example 1: Allocation Problem (Voting Problem in Disguise)

How many ways to distribute $k = 5$ apples among $n = 3$ kids Alice, Bruce and Cindy?



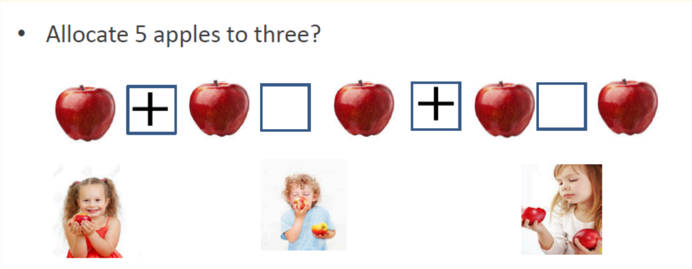
- Item is the person now as they are distinct, the apples are all same
- hard to solve directly need to transform the question into something that can be use combination to solve

- another example is 5 votes for 3 candidates
- k is now more than n (also possible for k <= n)



- If allocation of 0 is allowed, separators can be next to each other.
- Add in 2 separators to separate into the apples into 3 groups
- There is no need to multiply 3! for the permutation of A, B and C, as that is taken account for when separating the apples

$$Number\ of\ ways\ =\ 7C2\ =\ 7C5\ =\ 21$$



- if zero allocation is not allowed, separator cannot be next to each other, must be between apples
- There's four position that the separator can occupy

$$Number\ of\ ways\ =\ 4C2\ =\ 6$$

<u>Example 2: Voting Problem</u>

All the 10 members in a committee will vote for a chairperson among them. There are 3 candidates running for the election. How many possible vote results?

- This questions lacks information, candidates are allowed to vote for themselves
- Assume that all candidates vote for themselves, there are 7 vote left to allocate to 3 candidates

$$Number\ of\ ways\ = \binom{3+7-1}{2} = \binom{9}{2} = 36$$

Without replacement and without ordering

- After each pick, the information of the item picked is recorded without order and not replaced back into the collection of items
- Picking k items from n distinct items without replacement and with ordering is $\frac{n!}{(n-k)!}$ outcomes, so for the same case without ordering, the number of outcomes due to ordering needs to be divided away leaving with $\frac{n!}{k!(n-k)!}$
- In general, picking k times from n distinct items without replacement and without ordering, it is a combination. The number of different outcomes is

$$nCk = \binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!}$$

<u>Example 1</u>

A six-bit binary sequence X has each of its bits being 0 or 1 at random. Its number of nonzero bits is called its Hamming weight. What is the most possible Hamming weight of X?

- Since each bit in X is picked from $\{0,1\}$ at random, each of them is 0 or 1 with probability 0.5. There are 64 possible sequences. Each occurs with the same probability $p = 1/64$.
- Two sequences have the same Hamming weight if they have the same number of 1's. Positions of these 1's do not matter. How many sequences have the same Hamming weight w?

The table below shows the numbers of sequences for different Hamming weight w.

w	0	1	2	3	4	5	6
#	1	6	15	20	15	6	1

Since all sequences appear with the same probability, the most possible Hamming weight is the one with the greatest number of sequences. From the table above, sequences with Hamming weight 3 appear with the highest probability $P = 20p = \frac{5}{16}$.

How about 0 and 1 are not with equal probability?

P(w = 0) = $(1/2)^6$ = 1/64
P(w = 1) = 6C1 $(1/2)^5(1/2)^1$ = 6/64
P(w = 2) = 6C2 $(1/2)^4(1/2)^2$ = 15/64
P(w = 3) = 6C3 $(1/2)^3(1/2)^3$ = 20/64

Binomial distribution with equal probability of success and failure is symmetric
P(w = 4) = 6C4 $(1/2)^2(1/2)^4$ = 15/64
P(w = 5) = 6C5 $(1/2)^1(1/2)^5$ = 6/64
P(w = 6) = $(1/2)^6$ = 1/64

<u>Example 2</u>

(b) Suppose you have an urn of 10 marbles, in which there are 4 blue marbles and 6 red marbles. You randomly pick a sample of 4 marbles <u>without replacement</u>. Let X be the random variable that describes the number of blue marbles found in the random sample. The probability of observing k blue marbles in a sample of 4 marbles is in fact a hypergeometric distribution:

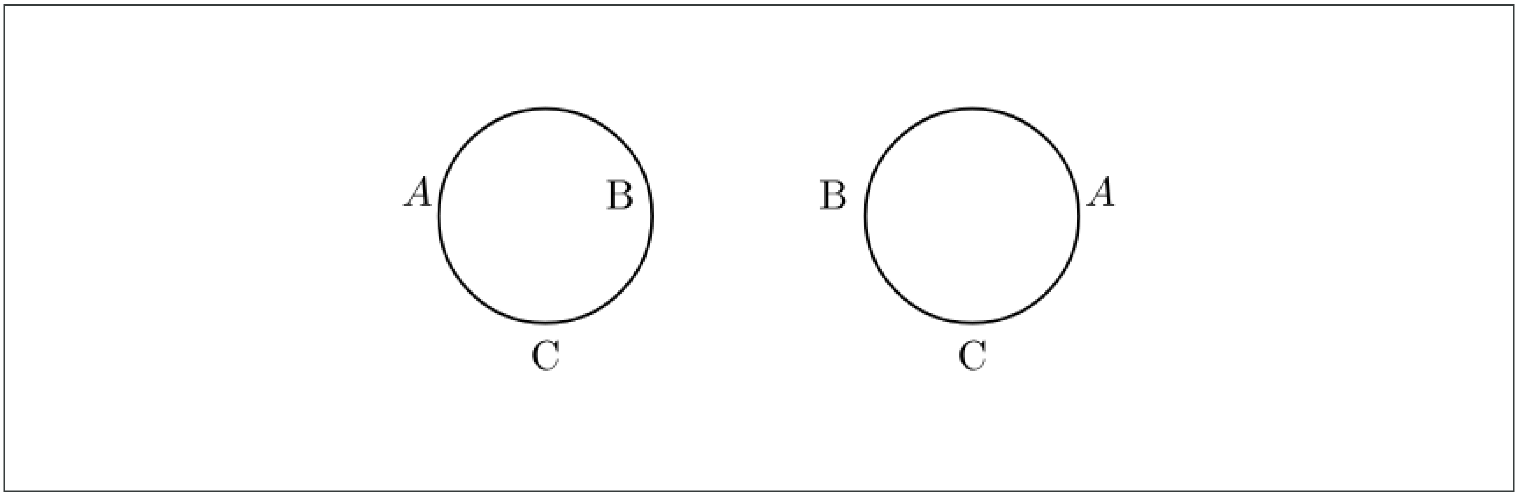
$$P[X = k] = \frac{\binom{4}{k}\binom{6}{4-k}}{\binom{10}{4}}$$

for $k = 0,1,2,3,4$. The mean and standard deviation of X are 1.6 and 0.8, respectively. What is the expectation value of X^2? (5 marks)

Circular Permutation

Circular Permutations

Suppose we have three people named A, B, and C. We have already determined that they can be seated in a straight line in 3! or 6 ways. Our next problem is to see how many ways these people can be seated in a circle. We draw a diagram:



It happens that there are only two ways we can seat three people in a circle. This kind of permutation is called a circular permutation. In such cases, no matter where the first person sits, the permutation is not affected. Each person can shift as many places as they like, and the permutation will not be changed. Imagine the people on a merry-go-round; the rotation of the permutation does not generate a new permutation. So in circular permutations, the first person is considered a place holder, and where he sits does not matter.

Circular Permutations: The number of permutations of n elements in a circle is $(n-1)!$

Example 5.3.7

In how many different ways can five people be seated at a circular table?

Solution

We have already determined that the first person is just a place holder. Therefore, there is only one choice for the first spot. We have:

1	4	3	2	1

So the answer is 24.

In how many ways can four couples be seated at a round table if the men and women want to sit alternately?

Solution

We again emphasize that the first person can sit anywhere without affecting the permutation. So there is only one choice for the first spot. Suppose a man sat down first. The chair next to it must belong to a woman, and there are 4 choices. The next chair belongs to a man, so there are three choices and so on. We list the choices below.

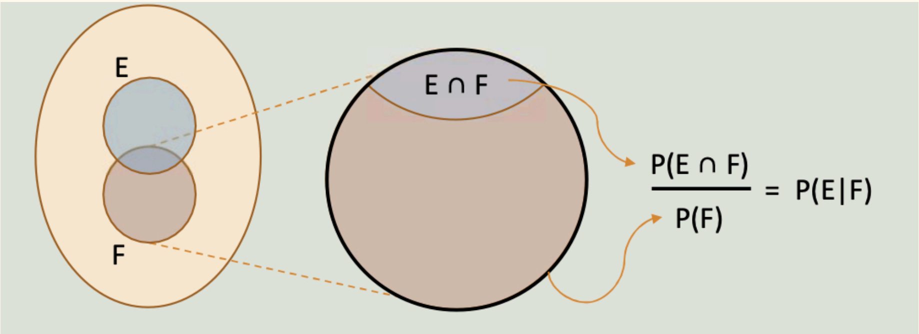
1	4	3	3	2	2	1	1

So the answer is 144.

https://pressbooks.library.torontomu.ca/ohsmath/chapter/5-3-permutations/#:~:text=Circular%20Permutations%3A%20The%20number%20of,circle%20is%20(n%20−%201)!&text=Example%205.3.7-,In%20how%20many%20different%20ways%20can%20five,seated%20at%20a%20circular%20table%3F&text=So%20the%20answer%20is%2024.

Chapter 2: Conditional Probability and Bayes' Rule

Conditional Probability



The probability of E given F measures how likely the outcome of the probability experiment is an element of E, if we already know that it is an element of F. To compute conditional probabilities, the idea of restricting the sample space based on the condition that event F is known to have occurred is used.

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

- If there is no overlap between events E and F, then given that F occurs, E cannot occur.
    - this agrees with $P(E \cap F) = 0$ and thus $P(E|F) = 0$
- If event F itself cannot occur, $P(F) = 0$, then by convention $P(E|F) = 0$
- If event E and F are independent, $P(E|F) = P(E)$, thus $P(E \cap F) = P(E)\,P(F)$
- To avoid dividing by zero, especially with finite resolution in computer, the equation is written as:

$$P(E \cap F) = P(E|F) \times P(F)$$
$$= P(F \cap E) = P(F|E) \times P(E)$$

- If all outcomes are equiprobable:

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

- Conditional probability, $P(A|B)$, can be greater, less, equal to unconditional probability, $P(A)$, conditional probability just means having extra side information.
    - Conditional probabilities are refined estimation of the probability of A based on the side-information available.
    - $P(A|B) = P(A)$ means B gives no extra information, the fact that B occurs does not improve the estimate of A, A and B are independent.

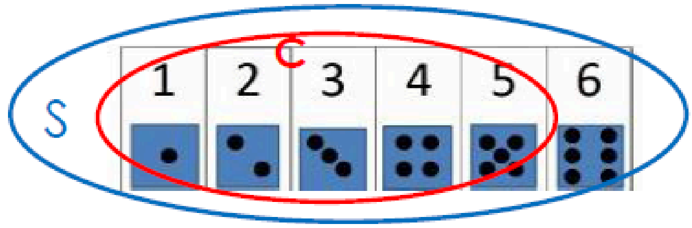
Properties of Conditional Probability

Conditional probability is still probability and thus follows the same properties as unconditional probability. All the properties of probability are valid for conditional probability:

- [Nonnegativity] $P(A|B) \geq 0$
- [Normalisation] $P(S|B) = 1$
- [Additivity] if A_1 and A_2 are mutually exclusive (disjoint) given B, then $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$

Example 1

A fair dice is rolled. Let X is the number on the side facing up. Event $A = \{X$ is even$\}$, $C = \{X \leq 5\}$ and $D = \{X \leq 4\}$. Find the probabilities $P(A)$, $P(A|C)$ and $P(A|D)$.



- Since the dice is fair, $P(A) = \frac{3}{6} = \frac{1}{2}$.
- $P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{2}{5}$.
- $P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{2}{4} = \frac{1}{2}$.

Example 2

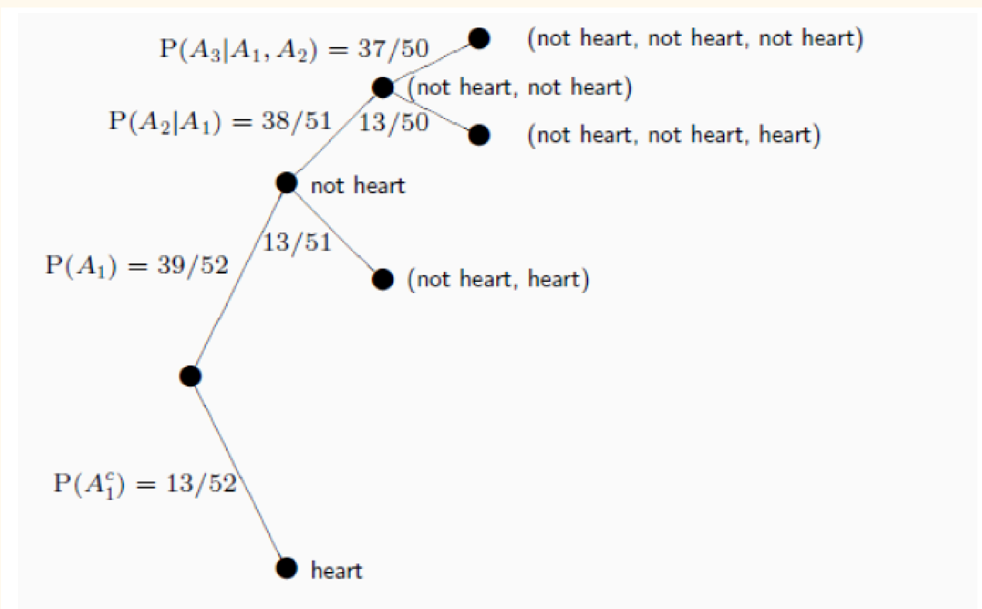
Three cards are drawn randomly without replacement from a deck of 52 cards. What is the probability of not drawing a heart?

- Let A_i, for $i = 1,2,3$, represent the event of **no heart** in the i^{th} draw.
- The sample space can be represented as

  $S = \{(A_1, A_2, A_3), (A_1^c, A_2, A_3), \cdots, (A_1^c, A_2^c, A_3^c)\}$

- To find the probability, we can use a tree diagram with conditional probability alone the tree branches and make use of the chain rule.

|S| = 8



- root nodes and leaves nodes, leaves are the results
- $P(A_1, A_2, A_3)$ = 39/52 * 38/51 * 37/50
  - due to without replacements,

Independent and Disjoint

If P(A|B) = P(A), the side-information of event B does not help refining probability of A, the events A and B are independent:

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

If A ∩ B = ø, they are mutually exclusive (disjoint):

$$P(A \cup B) = P(A) + P(B)$$
$$P(A \cap B) = 0$$
$$P(A|B) = 0$$

- if B occurs, A will never occur, thus a “negative” relationship thus not independent

For P(A) ≠ 0 and P(B) ≠ 0,

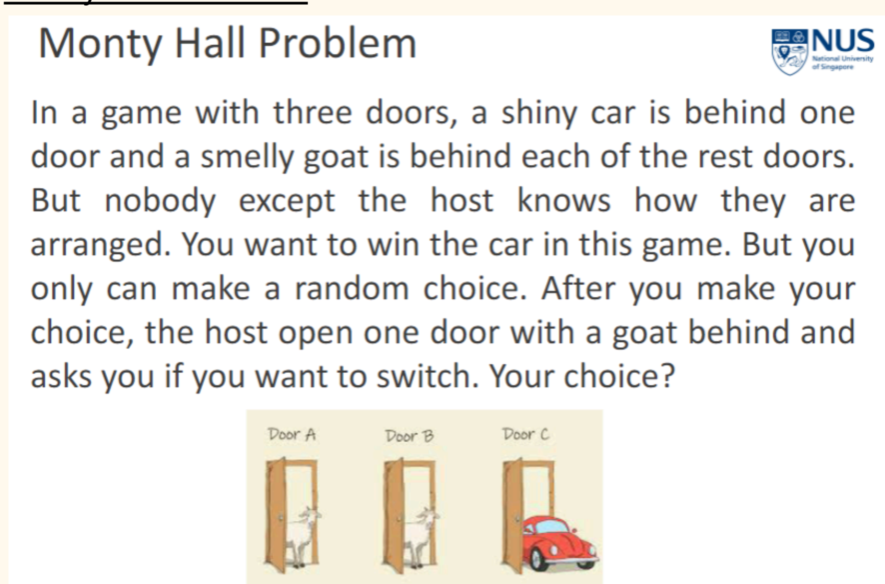
1. If A and B are independent, they are not disjoint

$$A \text{ and } B \text{ are independent} \Rightarrow P(A \cap B) = P(A) \times P(B)$$
$$P(A) \neq 0, P(B) \neq 0 \Rightarrow P(A \cap B) \neq 0$$
$$\Rightarrow A \cap B \neq ø, \text{ not mutually exclusive}$$

2. If A and B are disjoint, they are not independent

$$A \text{ and } B \text{ are disjoint} \Rightarrow P(A \cap B) = 0$$
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$$
$$P(A|B) = 0 \neq P(A), \text{ not independent}$$

Monty Hall Problem



- 2 Goats, 1 Car

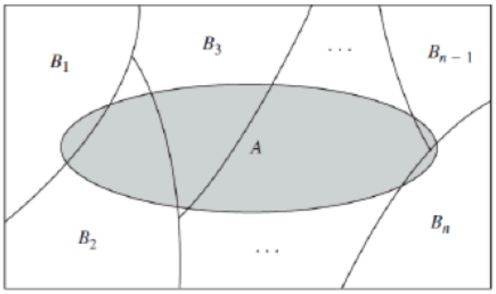
    - if don't switch, 1/3 of getting car
    - if always switch
        - if chose car, 1/3 * 0 to get car
        - if chose goat, 2/3 * 1 to get car
        - total = 2/3 > 1/3
    - It's always better to switch. Even for for any n doors and 1 car
        - What if there are n doors ($n \geq 3$) and only one with a car behind?
            - Winning chance is $1/n$ if no switch.
            - Switch.
                - One may win only when he chooses a door with goat at beginning, which has probability $\frac{n-1}{n}$
                - And he also must switch to the door with the car to win the game, which occurs with probability $\frac{1}{n-2}$ because two doors (with goat behind) are excluded.
                - The winning probability is therefore $\frac{n-1}{n} \times \frac{1}{n-2} = \frac{n-1}{n-2} \times \frac{1}{n}$.

          - The probability gap is $\frac{n-1}{n-2} \times \frac{1}{n} - \frac{1}{n} = \frac{1}{n(n-2)}$. For example, when $n = 3$, $\frac{n-1}{n} \times \frac{1}{n-2} = \frac{2}{3} > \frac{1}{3}$. For $n = 4$, $\frac{n-1}{n} \times \frac{1}{n-2} = \frac{3}{8} > \frac{1}{4}$...
          - Although the probability of winning with switching becomes smaller when n increases, it always strictly greater than the winning probability without switch because $\frac{n-1}{n-2} \times \frac{1}{n} > \frac{1}{n}$.
        - (n-1)/(n-2) > 1
- 3 Goats, 2 Cars
    - if don't switch, 2/5 of getting car
    - if always switch
        - if chose car, 2/5 * 1/3 to get car
        - if chose goat, 3/5 * 2/3 to get car
        - total = 2/5 * 1/3 + 3/5 * 2/3 = 0.5333333333 > 2/5

Total probability

Assume $\{B_1, B_2, \dots, B_n\}$ is a partition of the sample space S. An event A can be represented as union of disjoint sets.

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_n) = \cup_{i=1}^{n} A \cap B_i$$



Events $(A \cap B_i)$, $1 \leq i \leq n$, are disjoint to each other.

Theorem of total probability

$$P(A) = P(\cup_{i=1}^{n} (A \cap B_i))$$

$$= \sum_{i=1}^{n} P(A \cap B_i)$$

$$= \sum_{i=1}^{n} P(A|B_i)\, P(B_i)$$

- For a partition of sample space S, {B1, B2, …, Bn}, the probability of A, P(A), can be expressed as the sum of P(A|B1)P(Bi)
- This is called the projection of A to this partition. By splitting sample space into partitions, it is easier to find P(A) as it is easier to find A n B. (Divide and conquer)
- Since Bi are disjoint, union can be turned into sum
- Since P(A ∩ Bi) is symmetrical, the sum can be evaluated by sum of P(Bi|A)P(A) instead. Dividing both side by P(A) as it is a constant of the sum, shows that sum of P(Bi|A) = 1, meaning that Bi are also a partition of the restricted sample space A. Make sense considering A is a subset of S.

Example 1

Two robot arms are used in two assembly lines B_1 and B_2. They have different defective rates p and q. The numbers of products from the assemble lines are N_1 and N_2, respectively. What is the probability that a randomly picked product is defective from this factory?

– Probability of product picked from assembly lines:

$$\mathrm{P}(B_1) = \frac{N_1}{N_1+N_2},\; P(B_2) = \frac{N_2}{N_1+N_2}.$$

– Let A denote the randomly picked product is defective.

$$P(A) = P\big((AB_1) \cup (AB_2)\big) = P(AB_1) + P(AB_2)$$

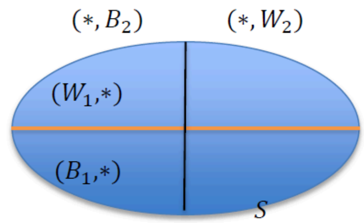
– $P(AB_1) = \mathrm{P}(A|B_1)P(B_1) = p \times P(B_1)$,

$P(AB_2) = q \times P(B_2)$.

$$P(A) = \frac{pN_1 + qN_2}{N_1 + N_2}$$

<u>Example 2</u>

An urn contains two black balls and three white balls. Two balls are selected at random from the urn without putting back and the sequence of colors is noted. Find the probability that both balls are black, and the probability of the second ball is white.



Denote the step i outcome is a black ball by B_i.

$$P(B_1 \cap B_2) = P(B_2|B_1) \times P(B_1) = \frac{1}{4} \times \frac{2}{5} = 0.1.$$

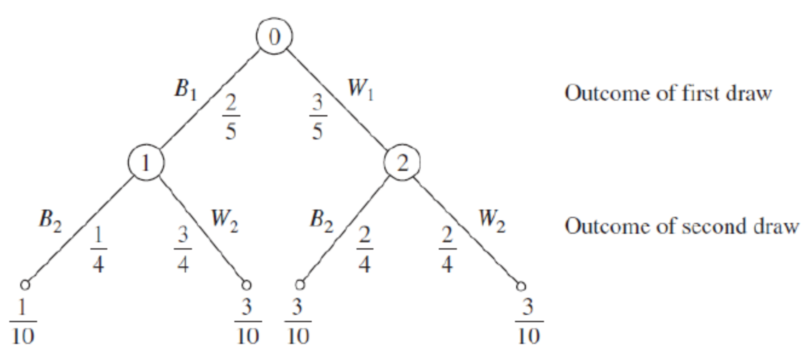
The events $B_1 = \{(B_1, B_2), (B_1, W_2)\}$ and $W_1 = \{(W_1, B_2), (W_1, W_2)\}$ form a partition of the sample space. By the Theorem of total probability, we have

$$P(W_2) = P(W_2 B_1) + P(W_2 W_1) = P(W_2|B_1)P(B_1) + P(W_2|W_1)P(W_1)$$
$$= \frac{3}{4} \times \frac{2}{5} + \frac{2}{4} \times \frac{3}{5} = 0.6.$$

If one picks a white ball will win a prize, any difference when picking first or second? (It is easy to see $P(W_1) = 0.6$ as well.)

- P(W|go first) = 0.6
- P(W|go sec) = P(W| first is W) + P(W| first is B) = 0.6*0.5 + 0.4*0.75 = 0.6
- Probability of getting white is the same if go first or go second
- A tree diagram can be used to visualise the problem

The random experiment can be viewed as a sequential experiment in two steps. A more convenient tool to analyse it is the tree diagram.



Along the leftmost branch, we have $P(B_1 \cap B_2) = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10}$.

Considering two branches ending with W_2, we have $P(W_2) = \frac{3}{5}$.

Bayes' Rule

Let $\{A_1, A_2, \ldots, A_n\}$ be a partition of the sample space S.

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{P\left(B \cap \left(\cup_{i=1}^{n} A_i\right)\right)}$$

$$= \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B \cap A_i)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

- Given that an observation B occurred, what is the probability it is from the Aj category.
    - line 1 to line 2 use conditional probability
    - line 2 to line 3 use conditional probability for numerator and total probability for denominator
- The denominator, P(B), is calculated using the theorem of total probability
- The numerator is one of the terms in the sum of the denominator.
- Naïve Bayes Classifier

  NUS National University of Singapore

  - Learning from data by conditional probability $P(F|C_i)$ and $P(C_i)$.
    - C_i: class/category of the object
    - F: observed features of the object
    - Obtain statistics for $P(C_i)$ and $P(F|C_i)$ from training data.

Example 1

In the assembly line example, if a defective product is found in the products of this factory (event A), what is the probability that it is from line B_1?

$$P(B1|A) = \frac{P(A|B1)P(B1)}{P(A)}$$

from previous example, P(A) is calculated to be:

$$P(A) = \frac{pN_1 + qN_2}{N_1 + N_2}$$

thus

$$P(B1|A) = \frac{P(A|B1)P(B1)}{\frac{pN_1 + qN_2}{N_1 + N_2}}$$

$$P(B1|A) = \frac{p\frac{N_1}{N_1 + N_2}}{\frac{pN_1 + qN_2}{N_1 + N_2}}$$

$$P(B1|A) = \frac{pN_1}{pN_1 + qN_2}$$

Example 2

- An urn contains two black balls and three white balls. Two balls are selected at random from the urn without putting back and the sequence of colors is noted. Find which color of the first ball is more probable given that the second ball is black.

- $P(B_2) = P((B_2B_1) \cup (B_2W_1)) = P(B_2B_1) + P(B_2W_1)$
$$= P(B_2|B_1)P(B_1) + P(B_2|W_1)P(W_1) = \frac{1}{4} \times \frac{2}{5} + \frac{2}{4} \times \frac{3}{5} = \frac{2}{5}.$$

- $P(W_1|B_2) = \frac{P(B_2W_1)}{P(B_2)} = \frac{P(B_2|W_1)\,P(W_1)}{P(B_2)} = \frac{3}{4}.$
$$P(B_1|B_2) = \frac{P(B_2B_1)}{P(B_2)} = \frac{P(B_2|B_1)\,P(B_1)}{P(B_2)} = \frac{1}{4}.$$

- What is your guess for the first ball colour? What if the second ball is white?

P(W1|B2) = P(W1 n B2)/P(B2) = P(B2|W1)P(W1)/0.4 = (0.5)(0.6)/0.4 = 0.75
P(B1|B2) = P(B1 n B2)/P(B2) = P(B2|B1)P(B1)/0.4 = (0.25)(0.4)/0.4 = 0.25
P(B2) = P(WB, BB) = 0.6*0.5 + 0.4*0.25 =0.4

Example 3

Let A be the event of a plane flying above, and B be the event that the radar detects it.
P(A) = 0.05
P(B|A) = 0.99
P(B|A’) = 0.1

Find the probability that a plane flew above given that the radar detected it.

- Taking A and A’ as disjoint partition of sample space S, bayes’ rule can be applied
- And project B into this partition

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{P\left(B \cap \left(\cup_{i=1}^{n} A_i\right)\right)}$$
$$= \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B \cap A_i)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

-
- P(A|B) = P(B|A)P(A) / [P(B|A)P(A) + P(B|A’)P(A’)]
- P(A|B) = 0.99*0.05 / [0.99*0.05 + 0.1*0.95]
- P(A|B) = 0.34256055363

Not a good radar because roughly ⅔ is false alarm, and there is a need to reduce false alarm rate.

Example 4
check slides
P(rich|happy) = 0.1
P(not rich|happy) = 0.9
Assume P(rich) and P(happy)
then find P(happy|rich) and P(happy|not rich)

Establishing association	
Positive association between A and B: (any of the following)	Negative association between A and B: (any of the following)
rate(A | B) > rate(A | NB) rate(B | A) > rate(B | NA) rate(NA | NB) > rate(NA | B) rate(NB | NA) > rate(NB | A)	rate(A | B) < rate(A | NB) rate(B | A) < rate(B | NA) rate(NA | NB) < rate(NA | B) rate(NB | NA) < rate(NB | A)

Happy because rich or rich because happy need more evidence, only shows correlation

Sequential experiment

$$\begin{array}{llll} A = & (A_1, & A_2, \cdots & A_n) \leftarrow \text{Result of the experiment} \\ \uparrow & \uparrow & \uparrow & \cdots \uparrow \\ S & S_1 & S_2 \cdots & S_n \quad \leftarrow \text{Sample space} \end{array}$$

- Random experiment is conducted in steps, where each step can be s sub-experiment and produce random output
- Overall outcome sithe collection of all the sub experiments out

Independent Sequential Experiment

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2)\cdots P(A_n) = \prod_{k=1}^{n} P(A_k)$$

- Sub-experiments are conducted independently, each step does not depend on the outcome of others.
- eg: Pick balls from sack with replacement.

Example

When a coin is flipped, its head appears with probability p. If it is flipped 3 times, what is the probability of sequence $H_1T_2H_3$?

- The n-flips can be viewed as repeating a sub-experiment 3 times. Each sub-experiment is flipping the coin once.
- In each sub-experiment $\mathrm{P}(H) = p$ and $P(T) = 1 - p$.
- Since 3 sub-experiments are independent,

$$P(H_1T_2H_3) = P(H_1)P(T_2)P(H_3) = p^2(1-p)$$

- Compare with the probability of 2 heads in 3 flips.
  - not the same as 2 heads in 3 flips as this can take any sequence and uses binomial to solve

Dependent Sequential Experiment

$$P(A_1A_2A_3) = P(A_3|A_1A_2)P(A_2|A_1)P(A_1)$$

– Markov chain $P(A_i|A_{i-1}A_{i-2}\cdots A_1) = P(A_i|A_{i-1})$

$$P(A_1, A_2 \dots, A_n) = P(A_n|A_{n-1}) \times P(A_{n-1}|A_{n-2}) \times \cdots \times P(A_2|A_1)P(A_1).$$

- Each subsequent outcomes depends on the whole history of outcomes
- The Markov chain is a special case where the current outcome only depends on the previous outcome, not all the way to the first outcome.
  - markov chain for convolution something something
  - used for particle movement modelling

Example

Two urns 0 and 1 have marbles (marked by 0 and 1) showed by the table. The marbles are picked following the rule below.

- Each picked marble is put back after taking note its number.
- The number on the picked marble indicates which urn to pick next.
- The urn for the firs pick is determined by flip a fair coin. $H \rightarrow \text{urn } 0$.

Find the probability of the urn sequence $A_1A_2A_3 = 0\,1\,0$.

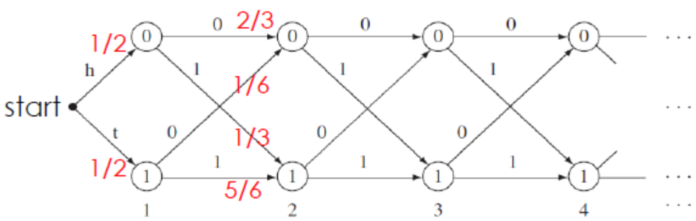
Urn	Marbles with 0	Marbles with 1
0	2	1
1	1	5

- The result of a pick depends on the previous picks. (Dependent sequential)
- $P(A_1A_2A_3) = P(A_3|A_1A_2)P(A_2|A_1)P(A_1) = P(A_3|A_2)P(A_2|A_1)P(A_1)$

$$= \frac{1}{6} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{36}$$

- this is an example of Markov chain

The Markov chain model can be applied to analyze this problem. A graphic tool called trellis diagram is useful for Markov chain model.



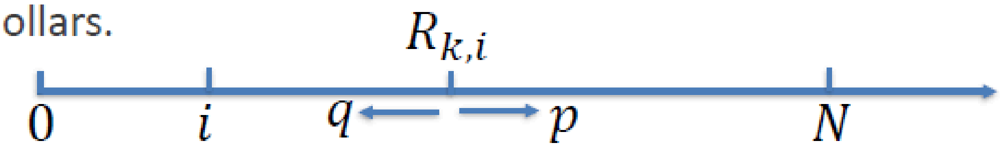
Trellis diagram

The trellis diagram above is for this example. The probability of sequence of urns 010 can be read out from the trellis as $\frac{1}{2} \times \frac{1}{3} \times \frac{1}{6} = \frac{1}{36}$.

-

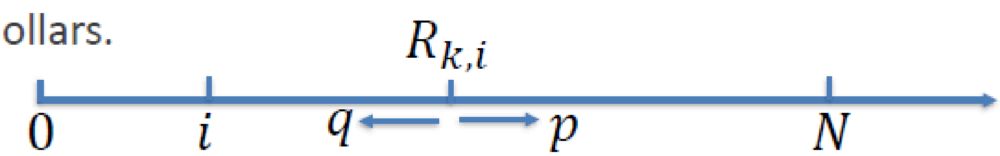
Example 2: Gambler's ruin

[Gambler's ruin] A player starts with i dollar. In each game, he will win 1 dollar with probability p and lose 1 dollar with probability q. He will stop playing when either he loses all his money or have total N dollars ($N > i$). Find the probability he stops playing with N dollars.



- In each game, the player's money varies by 1 . The money the player has is $i + \Delta_1 + \Delta_2 + \cdots + \Delta_k$ after the k^{th} game $\Delta_j \in \{1, -1\}$ for $k \geq j \geq 1$.
- If player starts with amount i, denote his money after k^{th} game as $R_{k,i}$ which is the outcome of the random sequential experiment. Current step outcome depends on only the previous step outcome. – Markov chain
- The player stops playing only when $R_{k,i} = 0$ (lose) or N (win).

[Gambler's ruin] A player starts with i dollar. In each game, he will win 1 dollar with probability p and lose 1 dollar with probability q. He will stop playing when either he loses all his money or have total N dollars ($N > i$). Find the probability he stops playing with N dollars.



- In each game, the player's money varies by 1 . The money the player has is $i + \Delta_1 + \Delta_2 + \cdots + \Delta_k$ after the k^{th} game $\Delta_j \in \{1, -1\}$ for $k \geq j \geq 1$.
- If player starts with amount i, denote his money after k^{th} game as $R_{k,i}$ which is the outcome of the random sequential experiment. Current step outcome depends on only the previous step outcome. – Markov chain
- The player stops playing only when $R_{k,i} = 0$ (lose) or N (win).

– If $r = \frac{q}{p} = 1$, $P_{i+1} - P_1 = iP_1$ and $P_{i+1} = (i+1)P_1$.

$$P_N = NP_1 = 1 \rightarrow P_1 = \frac{1}{N} \rightarrow P_i = \frac{i}{N}.$$

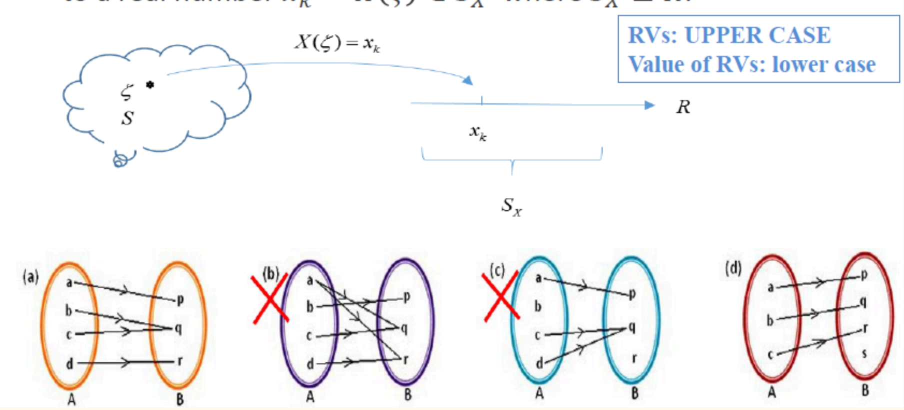
– For a greedy player want to stop at a larger N, the probability for him to win becomes smaller whatever r is.

– The winning probability also depends on the initial money i the player has. In a two-player setting, the chance is in favor of the player with more money. Who is the player in casino with the most money?

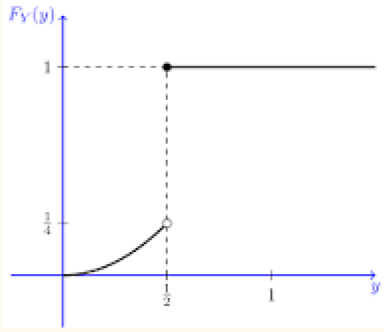
- Example of markov chain as whether to continue or not is based on the amount of money in the previous round
- for a game of win all lose all, P(win) = i/i+j as N = i + j, the player with more money will win with higher chance.

<u>Chapter 3 - Discrete Random Variables and Distributions</u>

- A random variable is defined by the the mapping from $\zeta \in S$ to a real number $x_k = X(\zeta) \in S_X$ where $S_X \subseteq R$.



- mapping elementary outcome in a sample space to real numbers
- Mapping from preimage to image, the mapping must be:
    - one to one or many to one (HHT, HTH, THH = 2)
    - all elementary outcome in sample space must be mapped, otherwise sum of probability is not 1
        - as any outcome in a sample space should have non zero probability
    - can have real value not being mapped to, just mean probability of 0
- All the possible values of X = range of value the outcomes can be mapped to = support of random variable X, S_X
    - General case: for any $S \subseteq S_X$, we need to know $P(X \in S)$ so that we can understand random variable X.
    - Discrete case: for any $x \in S_X$, we need to know $P(X = x)$ so that we can understand random variable X.
    - in other words, probability must be defined for the support of X
    - The support is numbers not the outcomes of the random experiment, do not confuse.
- Types of random variable
    - Discrete random variable, S_X is a countable subset of the real numbers.
    - Continuous random variable, S_X is an uncountable subset of the real numbers.
    - Mixed-type random variable, S_X is an uncountable subset of the real numbers, mixed of the above two types
        - not in syllabus

- deals with delta and unit step function
- 

- Useful as:
  - Easy to represent events of interest as number, mapped to number and with number can use mathematical tools
  - Easy to compute probability of events of interest, as each support are disjoint as they are mapped from elementary outcomes
    - An outcome is only mapped to one value of X.

      – Event $A = \{x_1, x_2\}$. $\{X = x_1\}$ and $\{X = x_2\}$ are disjoint.

      – $P(A) = P(X \in \{x_1, x_2\}) = P(X = x_1) + P(X = x_2)$.

Example

Consider a game of flipping a coin 3 times. Tom wins the game if there are more heads than tails. Otherwise, Jerry wins.

– $\{\text{Tom wins}\} = \{\text{HHH, HHT, HTH, THH}\} = \{2 \text{ or } 3 \text{ heads}\}$

– $\{\text{Jerry wins}\} = \{\text{TTT, TTH, THT, HTT}\} = \{0 \text{ or } 1 \text{ heads}\}$

– Both players only care about the outcome quantitative attribute – the number of heads.

– A random variable X represents the number of heads appearing in three flips.

- $\{X > 1\} = \{\text{Tom wins}\}$
- $\{X \le 1\} = \{\text{Jerry wins}\}$

Easy and meaningful to represent the events of interest.

The head of a coin occurs with probability p. Let X be the number of heads in 3 flips. Find $P(X = 2)$ and $P(X = 2|X \ge 2)$.

ζ	TTT	HTT	THT	TTH	HHT	HTH	THH	HHH
$X(\zeta)$	0	1	1	1	2	2	2	3
P	$(1-p)^3$	$p(1-p)^2$	$p(1-p)^2$	$p(1-p)^2$	$p^2(1-p)$	$p^2(1-p)$	$p^2(1-p)$	p^3
	$\binom{3}{0} = 1$		$\binom{3}{1} = 3$			$\binom{3}{2} = 3$		$\binom{3}{3} = 1$

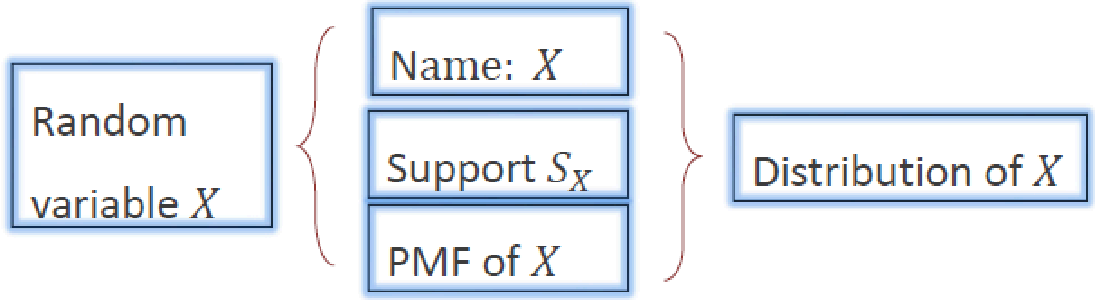
- $P(X = 2) = \binom{3}{2} p^2(1-p) = 3p^2(1-p) = \binom{n}{k} p^k (1-p)^{n-k}$
- $P(X = 2|X \ge 2) = \frac{P(\{X=2\} \cap \{X \ge 2\})}{P(X \ge 2)} = \frac{P(X=2)}{P(X \ge 2)} = \frac{3p^2(1-p)}{3p^2(1-p)+p3} = \frac{3-3p}{3-2p}$.

- Cumbersome to illustrate HHT or THH, define a discrete random variable X where X = numbers of heads in 3 flips
  - The sample space with the 8 outcomes will be mapped on to the support which is 0, 1, 2, 3.

Probability Mass Functions (PMF)

- Probability mass function (PMF)
  - For any $x_i \in S_X$, the X PMF $p_X(x_i) \triangleq P(X = x_i)$.
  - PMF indicates $P(X = x_i)$ for all $x_i \in S_X$.
  - In previous example of flipping coin 3 times, the PMF of X is

$$p_X(k) = \binom{3}{k} p^k (1-p)^{3-k} \text{ for } k \in S_X = \{0, 1, 2, 3\}.$$



$$PMF\ of\ X \ = \ P(X = x_i) = p_X(x_i) \quad for\ any\ x_i \in S_X$$

- PMF indicates the $P(X = x_i)$ for all $x_i \in S_X$, probability of all the values in the support of X
- Its a function taking x_i and give the probability of $X = x_i$
- To fully define a random variable, it must have:
  - The name, X
  - Support, S_X
  - PMF of X, $p_X(x)$
  - These three property of a random variable is called the distribution of X
- For continuous is called probability density function
- Properties of PMF:

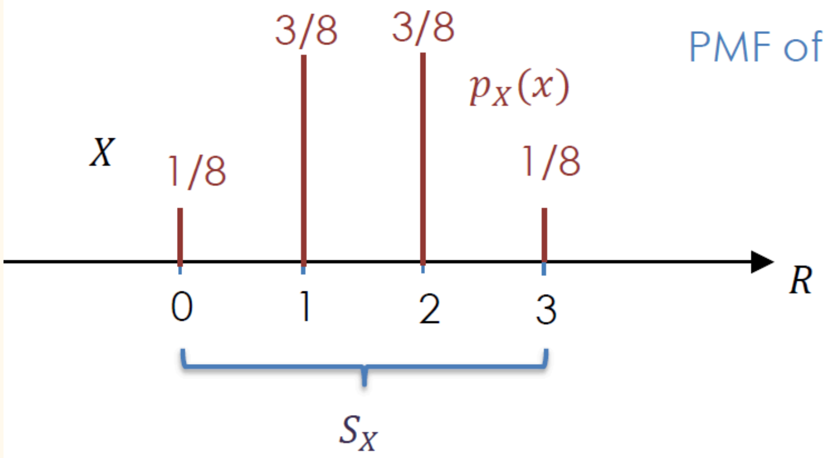
$$p_X(x) \ \geq 0 \quad for\ any\ x \in R$$

$$\sum_{x \in S_X} p_X(x) = 1$$

$$for\ B \subseteq S_X, P(X \in B) = \sum_{x \in B} p_X(x)$$

  - First properties is followed from the properties of probability where it must be non-negative
  - Second follows from the fact that all outcomes are mapped
  - Third is due to the mutual exclusivity of the elementary outcomes and the mapped values

Forms of PMF



PMF of X

X	0	1	2	3
p_X	1/8	3/8	3/8	1/8

- PMF can be visualise as a 2d plot and table,
    - The lines shows the support and the height is probability
- Can also be expressed in closed form which is the formula, but not all distribution has closed form

Expectation of Random Variables

Expectation of a RV X: the probabilistic weighted average of X.

– A discrete RV X with PMF $p_X(x)$ and support S_X. Its expectation is

$$\underbrace{m_X = \mu_X = E[X]}_{\text{Three notations}} \triangleq \underbrace{\textstyle\sum_{x \in S_X} x \times p_X(x)}_{\text{Definition}}.$$

In practise, the expectation of X is estimated by the arithmetic average of samples of X: take n samples x_i $(1 \le i \le n)$ of X. assume the value of x_i appears N_i times within the n samples. Their arithmetic average is

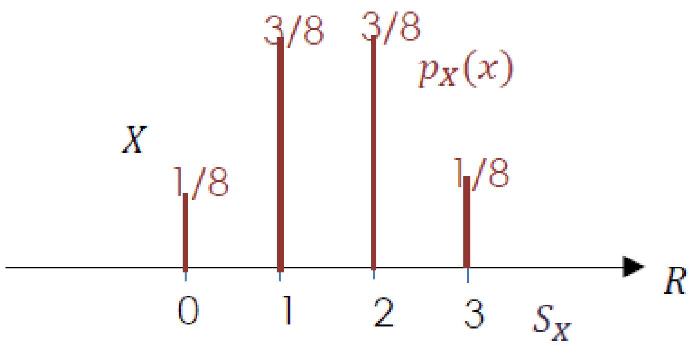
$$\langle X \rangle_n = \frac{\sum_{i=1}^{n} x_i}{\mathrm{n}} = \frac{x_1 N_1 + x_2 N_2 + \cdots x_m N_m}{n} = \sum_{k=1}^{m} x_k \frac{N_k}{n} = \sum_{k=1}^{m} x_k f_k \, .$$

When $n \to \infty$, $f_k \to p_X(x_k)$ and $\langle X \rangle_n \to E[X]$.

- Has two definitions,
    - Probabilistic weighted sum, which requires the knowledge of the distribution.
    - Expectation, E, is the value such that x - E is the closest to 0, minimise the error
- Distributions with the same expectation can be different distributions.
- Probabilistic weighted sum

$$E[X] \;=\; \sum_{x \in S_X} x p_X(x)$$

Expectation, Median and Mode



Expectation: $\sum_{x \in S_X} x p_X(x)$ =1.5 $\notin S_X$

or

Median: $\frac{3+0}{2} = 1.5$ $\in S_X$

Mode: $x = 1$ or 2. $\in S_X$

Expectation is the probabilistic weighted sum of all $x \in S_X \rightarrow E[X] = \sum_{x \in S_X} x p_X(x)$

Median is the middle value of $S_X \rightarrow \frac{MAX_{x \in S_X} x + MIN_{x \in S_X} x}{2}$

Mode is the value x_{mode} in S_X with largest PMF value $\rightarrow \arg_{x \in S_X} MAX\ p_X(x)$

- Median is only has to do with support not with probability
- Median and expectation can have a value that is not in the support
- Value of mode must be in the support
- Not all distributions have expected value and variance, the summation may not converge, thus expected value and variance does not exist.

Properties of Expectation

c and a are real constants

$E[Y] = E[c] = c$	Expected value of a constant is it self, can be treated as a RV with one possible value with P of 1
$E[X + c] = E[X] + c$	Constant can take out
$E[aX] = aE[X]$	Constant can take out
$E[X + Y] = E[X] + E[Y]$	X and Y can be independent and dependent. For variance must be independent.
$E[(X - a)^2] = E[X^2 - 2aX + a^2] = E[X]^2 - 2aE[X] + a^2$	E acts like a linear operations, can expand and shit
$E[XY] \neq E[X]E[Y]$	Unless independant
$E[X^2] \neq E[X]^2$	X is not independent of X

Derivations:

Find the expectations of the RVs below, where c and a are real constants.

1) $Y = c$. 2) $W = X + c$. 3) $U = aX$. 4) $V = (X - a)^2$. 5) $Z = X + Y$.

$E[Y] = c \times 1 = c$

$E[W] = \sum_{w \in S_W} w p_W(w) = \sum_{x \in S_X} (x + c) p_W(x + c)$

$= \sum_{x \in S_X} (x + c) p_X(x) = \sum_{x \in S_X} x p_X(x) + c p_X(x) = E[X] + c$

$E[U] = E[aX] = \sum_{x \in S_X} a x p_X(x) = a \sum_{x \in S_X} x p_X(x) = aE[X]$

$V = X^2 - 2aX + a^2$. $E[V] = E[X^2] - 2aE[X] + a^2$

$E[Z] = E[X] + E[Y]$ -- Proof later in multiple RVs.

<u>Example 1</u>

[Bernoulli RV] A Bernoulli RV X has the support $S_X = \{0,1\}$ and its PMF is

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases}.$$

- Its expectation is $m_X = p \times 1 + (1 - p) \times 0 = p$.

- Used to model bit transmission and bit flips

<u>Example 2</u>

[Discrete uniform RV] A discrete uniform RV has its support $S_X = \{a, a + 1, \cdots, b\}$, where both a and b are integers, and $b - a > 0$. Its PMF is

$$p_X(k) = \begin{cases} \frac{1}{b-a+1}, & k \in S_X \\ 0, & k \notin S_X \end{cases}.$$

- Its expectation $m_x = \sum_{x \in S_X} \frac{x}{b-a+1} = \frac{1}{b-a+1} \sum_{i=a}^{b} i = \frac{a+b}{2}$.

- median = expectation
- expectation can be gotten using series sum
- Note that from a to b there are b-a+1 numbers, so the probability ios 1/(b-a+1)
    - 2 to 5, 4 numbers, 5 - 2 = 3, 3 + 1 = 4

Example 3

Example - Expectation

[Expectation of X^2] Let X be the number of heads in three flips of a fair coin. Find $E[X]$ and $E[X^2]$.

- Based on previous example of flipping 3 fair coins, the PMF of X is listed in the table below.

X	0	1	2	3
p_X	1/8	3/8	3/8	1/8

$$E[X] = \sum_{x\in S_X} x \times p_X(x) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{3}{2}.$$

- X^2 is also a RV. Its support is $S_{X^2} = \{0,1,4,9\}$. Its PMF can be found based on the PMF of X.

X	0	1	2	3
X^2	0	1	4	9
p_{X^2}	1/8	3/8	3/8	1/8

$$E[X^2] = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 4 \times \frac{3}{8} + 9 \times \frac{1}{8} = 3.$$

For random variable X, X^2 must also be a random variable. X^2 is a function of random variable X.
$E(X^2)$ not $E(X)^2$ in general, although there could be cases where it is true by chance.

Y = (X-1)^2

Y =	1	0	1	4
p	⅛	⅜	⅜	⅛

E(Y) = 1*4/8 + 4/8 = 1

Variance of Discrete RVs

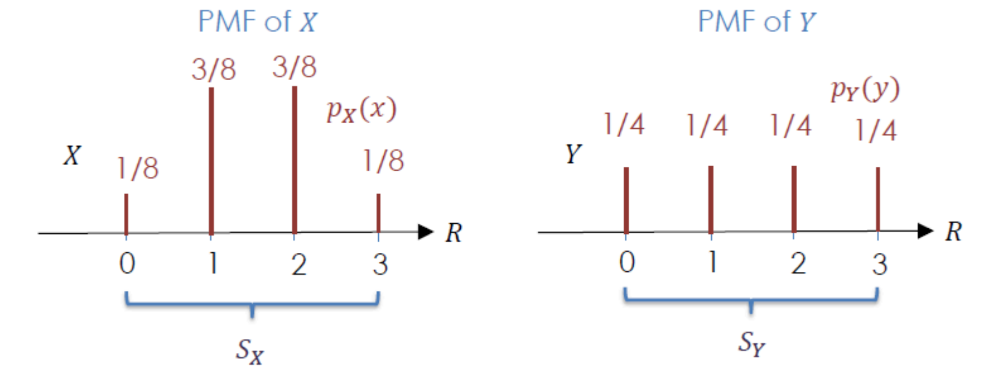
$$\underbrace{\sigma_X^2 = \mathrm{Var}[X]}_{\text{Two notations}} \triangleq \underbrace{E[(X-m_X)^2] = \textstyle\sum_{x\in S_X}(x-m_X)^2 p_X(x)}_{\text{Definition}}$$

- $\sigma_X = \sqrt{\sigma_X^2}$ is called the standard deviation of X.

$$\sigma_X^{\ 2} = Var[X] = E[(X-m_X)^2] = \sum_{x\in S_X}(x-m_X)^2 p_X(x)$$

$$\sigma_X^{\ 2} = Var[X] = E[X^2 - 2m_X X + m_X^{\ 2}] = E[X^2] - 2m_X E[X] + E[m_X^{\ 2}] = E[X^2] - 2m_X^{\ 2} + m_X^{\ 2} = E[X^2] - m_X^{\ 2}$$

$Var[X] = E[X^2] - {m_X}^2 = \sum_{x \in S_X} x^2 p_X(x) - \sum_{x \in S_X} x p_X(x)$

- Measure the scattering of RV around its expected value
- Not all distributions have expected value and variance, the summation may not converge, thus expected value and variance does not exist.
- RV can have the same expectation but different variance
  - 

Properties of Variance

$Var[X] = Var[c] = E[(c - c)^2] = E[0] = 0$	
$Var[X + c] = Var[X]$	Shifting does not change variance
$Var[aX] = a^2 Var[X]$	
$Var[X + Y] = Var[X] + Var[Y]$	X and Y must be independent, can generalised to multiple RVs

Derivations:

$Y = c$, $W = X + c$, $U = aX$ and $H = M + N$, where c and a are real constants, M and N are independent RVs. Find their variances.

$\sigma_Y^2 = E[(y - c)^2] = E[(c - c)^2] = E[0] = 0$

$\sigma_W^2 = E[(X + c - E[X + c])^2] = E[(X - E[X])^2] = \sigma_X^2$

$\sigma_U^2 = E[(aX - E[aX])^2] = E[a^2(X - E[X])^2] = a^2\sigma_X^2$

$\sigma_H^2 = \mathrm{Var}[M + N] = \mathrm{Var}[M] + \mathrm{Var}[N] = \sigma_M^2 + \sigma_N^2$ -- proof later

Example 1

Let X be the number of heads in three flips of a fair coin. Find σ_X^2.

M1: By definition, $\sigma_X^2 = E[(X - m_X)^2] = \sum_{x \in S_X} (x - m_X)^2 p_X(x)$.

$m_X = \frac{3}{2}$ and PMF of X

X	0	1	2	3
p_X	1/8	3/8	3/8	1/8

$\sigma_X^2 = \left(0 - \frac{3}{2}\right)^2 \times \frac{1}{8} + \left(1 - \frac{3}{2}\right)^2 \times \frac{3}{8} + \left(2 - \frac{3}{2}\right)^2 \times \frac{3}{8} + \left(3 - \frac{3}{2}\right)^2 \times \frac{1}{8} = \frac{3}{4}.$

M2: $m_X = \frac{3}{2}$ and $\mathrm{E}[X^2] = 3$ (we found out previously).

$\sigma_X^2 = E[X^2] - m_X^2 = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}.$

Example 2

Find the variance of a Bernoulli RV X.

- $p_X(k) = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \end{cases} \rightarrow p_{X^2}(j) = \begin{cases} p, & j = 1 \\ 1-p, & j = 0 \end{cases}$
- $m_X = p$ and $E[X^2] = p$.
- M1: $\sigma_X^2 = E[(X - m_X)^2] = (1-p)^2 \times p + (0-p)^2 \times (1-p)$

  $= p(1-p)$

  M2: $\sigma_X^2 = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$

Example 3

Let X_i be identical and independent distributed (iid) Bernoulli RVs with $1 \leq i \leq n$.

Find the expectation and variance of the RV $X = \sum_{i=1}^{n} X_i$.

- X is a binomial distributed RV, which can be modeled as the number of heads in n flips if the coin shows head in each flip with probability p. Its support $S_X = \{0,1,2,\cdots,n\}$ and its PMF is

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k \in S_X \\ 0, & k \notin S_X \end{cases}.$$

- The binomial distributed RV X with parameters n (number of flips) and p (probability of head in each flip) has expectation

$$m_X = \mathrm{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = np.$$

- Since $X_i's$ are iid, by the property of variance,

$$\mathrm{Var}\,[X] = \mathrm{Var}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = n\sigma_{X_i}^2 = np(1-p)$$

for the binomial distributed RV X with parameters n and p.

- Intuitively, if P(h) = 0.4, n = 100, get 40 heads, thus E[X] = np
- Identical and independently distributed (iid) RV so variance can sum tgt

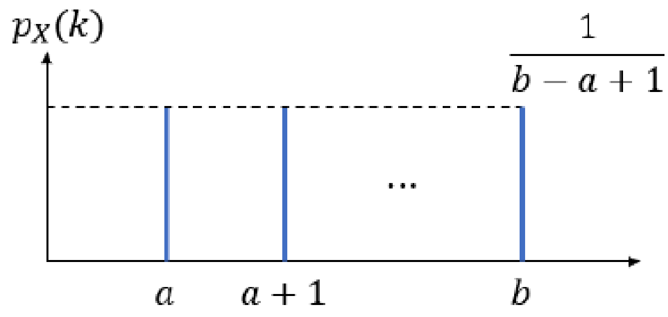
Summary for Common Discrete Random Variables

	PMF	E[X]	Var[X]	Assumptions/Conditions
Discrete Uniform $X \sim U(a,\ b)$	$p_X(x) = \{ \begin{smallmatrix} 1/(b-a+1) & , x \in S_X \\ 0 & , x \notin S_X \end{smallmatrix}$, $S_X = \{a, a+1,\ ...,\ b\}$	$E[X] = \frac{a+b}{2}$	$Var[X] = \frac{(b-a+1)^2-1}{12}$	
Bernoulli $X \sim Bernoulli(p)$	$p_X(x) = \{ \begin{smallmatrix} p & , x = 1 \\ 1-p & , x = 0 \end{smallmatrix}$, $S_X = \{0, 1\}$	$E[X] = p$	$Var[X] = p(1 - p)$	
Binomial $X \sim Bin(n,\ p)$	$p_X(x) = \{ \begin{smallmatrix} nCx\, p^x (1-p)^{n-x} & , x \in S_X \\ 0 & , x \notin S_X \end{smallmatrix}$, $S_X = \{0, 1\ ...,\ n\}$	$E[X] = np$ Highest Prob if in S_X	$Var[X] = np(1 - p)$	n is fixed All trials are independent P(success) = p is fixed
Geometric $X \sim Geo(p)$	$p_X(x) = (1 - p)^{x-1} p$, $S_X = \{1,\ 2,\ ...\}$	$E[X] = \frac{1}{p}$	$Var[X] = \frac{1-p}{p^2}$	All trials are independent P(success) = p is fixed x = no of trials until success, including success
Poisson $X \sim Poisson(\lambda)$	$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$, $S_X = \{0,\ 1,\ ...\}$	$E[X] = \lambda$	$Var[X] = \lambda$	Events are independent Events are disjoint Average of occurance in a given time interval = λ

Discrete Uniform Distribution

- Discrete uniform distributed RV X~U(a,b).

PMF $p_X(k) = \begin{cases} \frac{1}{b-a+1}, & k \in S_X \\ 0, & k \notin S_X \end{cases}$ with $S_X = \{a, a+1, \cdots, b\}$



Discrete uniform PMF

Expectation and variance

- Expectation $m_X = \frac{a+b}{2}$
- Variance $\sigma_X^2 = \frac{(b-a+2)(b-a)}{12} = \frac{n^2-1}{12}$ where $n = b - a + 1$.

- Variance independent on values of b and a only b-a+1, this is same and shifting and variance staying the same

Derivation for Variance

Assume $X \sim \mathrm{U(a,b)}$. Its support is $S_X = \{a, a+1, \cdots, b\}$ and $|S_X| = n = b - a + 1$. Let $Y = X - (a-1)$. Since X and Y only differs by a constant $a - 1$, their variances are the same according to the properties of variance we discussed in lecture. We next will calculate this variance.

Based on the relation between X and Y. The support of $S_Y = \{1,2,\cdots,n\}$. We can calculate the variance of Y by $\sigma_Y^2 = E[Y^2] - (E[Y])^2$. First, Y is uniformly distributed in the range $[1, n]$ and its expectation is $E[Y] = \frac{1+n}{2}$. We also have

$$E[Y^2] = \sum_{i=1}^{n} i^2 \times \frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{1}{6}(n+1)(2n+1),$$

where the sum of squares of natural numbers is $\sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1)$ (refer to this link).

Therefore,

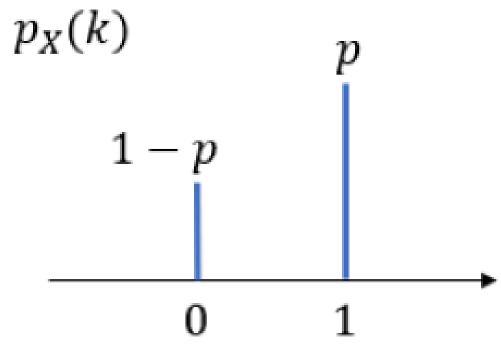
$$\sigma_Y^2 = \frac{1}{6}(n+1)(2n+1) - \left(\frac{1+n}{2}\right)^2 = \frac{1}{12}(\mathrm{n}+1)(4\mathrm{n}+2-3\mathrm{n}-3) = \frac{1}{12}(\mathrm{n}+1)(\mathrm{n}-1) = \frac{\mathrm{n}^2-1}{12}.$$

For $\mathrm{X} = Y + (a-1)$ its support is $S_X = \{a, a+1, \cdots, b\}$ and its variance is the same as the Y where $n = b - a + 1$.

Bernoulli Distribution

- Bernoulli RV $X \sim$ Bernoulli(p).

PMF $p_X(k) = \begin{cases} p, & k = 1 \\ 1-p, & k = 0 \end{cases}$ with $S_X = \{0,1\}$



Bernoulli PMF

Expectation and variance

- Expectation $m_X = p$
- Variance $\sigma_X^2 = p(1-p)$

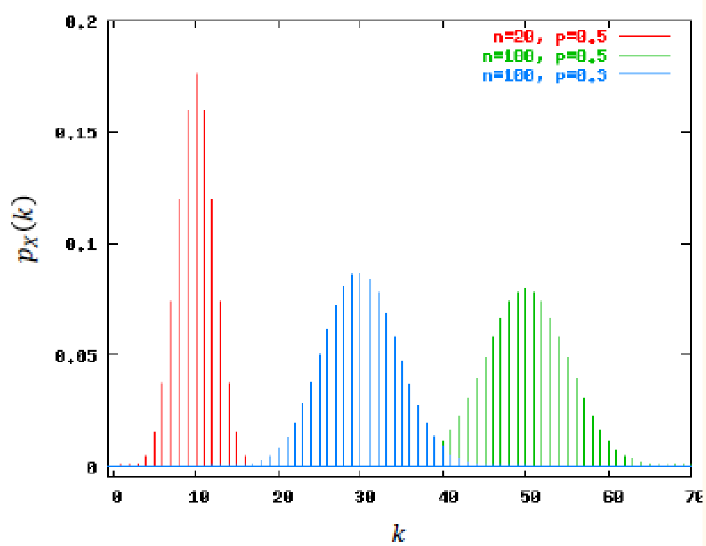
- Only two outcomes, two values in support

Binomial Distribution

- Binomial RV $X \sim \text{Bin}(n,p)$.

PMF $p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k \in S_X \\ 0, & k \notin S_X \end{cases}$ with $S_X = \{0, 1, \cdots, n\}$

- The number of trials n is fixed.
- Any two trials are independent.
- Each trial has the same "success" probability.



Expectation and variance

- Expectation $m_X = np$
- Variance $\sigma_X^2 = np(1-p)$

- binomial = n independent bernoulli
- The sum of probability can be proven to be 1, using binomial expansion (P + (1-P))^n = 1
- Shape of plot
    - Peak is at around exp val

    - Similar shape ⇒ similar variance
- The envelope of plot (the outline) is gaussian ish with higher n ⇒ central limit theorem
- Used for
    - sending package and sending package successful
    - n products with independent defective rate

Example 1

27

Add a reply

A coin with head probability p is flipped 27 times.

1. Find the probability of 1 head.

$$X \sim \text{Bin}(100, p).\ P(X=1) = \binom{27}{1} p^1 (1-p)^{27-1} = 27p(1-p)^{26}.$$

2. Find the probability of at least 2 heads.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X=0) - P(X=1)$$

$$= 1 - (1-p)^{27} - 27p(1-p)^{26}.$$

3. If $50P(X=2) = P(X=4)$, what is the value of p?

$$50\binom{27}{2} p^2 (1-p)^{25} = \binom{27}{4} p^4 (1-p)^{23} \rightarrow p = 0.5.$$

n = 100 p =?

X ~B(100, p)

P(X = 51) = P(X =50)

100C51 (p)^51 (1-p)^49 = 100C50 (P)^50 (1-p)^50

100C51 P = (100C50) (1-P)

P = 100C50/(100C51 + 100C50) = 0.50495049505 = 51/101

Example 2

60% of people who purchase sports cars are men. If 10 sports car owners are randomly selected, find the probability that there are at least 9 men.

- Let X be the number of male sports car owners out of 10 sports car owners. $X \sim \mathrm{Bin}(10, p)$ where $p = 0.6$.
- $P(X \geq 9) = P(X = 9) + P(X = 10)$

  $P(X = 9) = \binom{10}{9} p^9 (1-p)^{10-9} = 0.0403.$

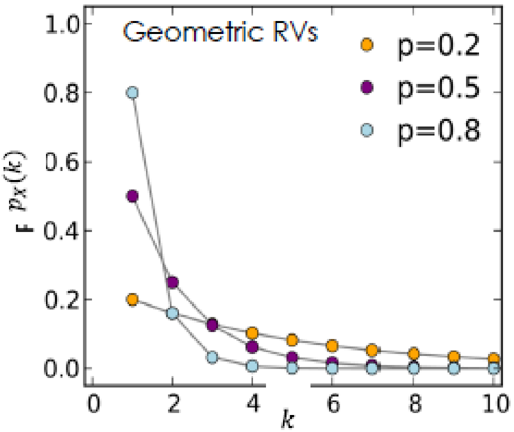
  $P(X = 10) = \binom{10}{10} p^{10} = 0.006$

  $P(X \geq 9) = 0.0463$

Geometric Distribution

- Geometric RV $X \sim \text{Geometric}(p)$.
  - Model: a coin gives head with probability p when it is flipped. The number of flips is counted until the first head appears (including the flip shows the first head). Let's call the final head appearance as "success".
  - PMF $p_X(k) = (1-p)^{k-1} p$ with $S_X = \{1, 2, \cdots\}$
  - Expectation $m_X = p^{-1}$.
  - Variance $\sigma_X^2 = \frac{1-p}{p^2}$.

(Refer to the supplementary material for the derivation.)



- x = number of flips until first success, including success
- Support from 1, because must at least flip once, to infinity
- Shape of plot
  - when k become larger, smaller p has higher probability, as more likely to fail in the first place

Example

A coin with head probability p is flipped in a game. A player can win a prize when the head occurs for the firs time at the fifth flip.

- What is the player's chance to win?

  Let X be a RV of number flips until the first head occurs. $X \sim \text{Geometric}(p)$

  $$P(\text{win}) = P(X = 5) = (1-p)^4 p$$

- What is the probability of at least 10 flips needed to see the first head?

  $$P(X \geq 10) = \sum_{k=10}^{\infty} (1-p)^{k-1} p = p \sum_{k=10}^{\infty} (1-p)^{k-1} = p \frac{(1-p)^9}{1-(1-p)}$$

  $$= (1-p)^9$$

  Geometric series

- Summation to infinity
  - For a geometric series:

    $$u_n = ar^{n-1}, \qquad S_n = \frac{a(1-r^n)}{1-r} \quad (r \neq 1), \qquad S_\infty = \frac{a}{1-r} \quad (|r| < 1)$$
- Equivalent to see 9 tails then the rest is probability of 1

Derivation for E[X]

Its expectation is

$$m_X = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Let $q = 1-p$. Then

$$\sum_{k=1}^{\infty} k(1-p)^{k-1} = \sum_{k=1}^{\infty} kq^{k-1} = \frac{d\left(\sum_{k=1}^{\infty} q^k\right)}{dq}. \quad (1)$$

Since q^k is a is geometric series for $k = 1,2,\cdots$, we have the infinite sum $\sum_{k=1}^{\infty} q^k = \frac{q}{1-q}$ and

$$\frac{d\left(\frac{q}{1-q}\right)}{dq} = \frac{1}{(1-q)^2} = p^{-2}.$$

So $m_X = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p \times p^{-2} = p^{-1}$.

E[X|A] = 1/p + 2 for A = {X>=3}, will add 2 then will be normal

Derivation for Var[X]

Its variance can be found with similar tricks. We can make use of $\sigma_X^2 = \mathrm{E}[X^2] - m_X^2$, where $m_X = p^{-1}$ and

$$\mathrm{E}[X^2] = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k^2(1-p)^{k-1} \tag{2}$$

Let $q = 1 - p$. Then

$$\sum_{k=1}^{\infty} k^2(1-p)^{k-1} = \sum_{k=1}^{\infty} k^2 q^{k-1} = \sum_{k=1}^{\infty} k(k+1)q^{k-1} - kq^{k-1}$$

$$= \sum_{k=1}^{\infty} k(k+1)q^{k-1} - \sum_{k=1}^{\infty} kq^{k-1}. \tag{3}$$

The second term above is the same as the Eqn 1 above, which is p^{-2}. The first term above is

$$\sum_{k=1}^{\infty} k(k+1)q^{k-1} = \frac{d^2\left(\sum_{k=1}^{\infty} q^{k+1}\right)}{dq^2}.$$

Since q^{k+1} is a is geometric series for $k = 1,2,\cdots$, we have the infinite sum $\sum_{k=1}^{\infty} q^{k+1} = \frac{q^2}{1-q}$ and

$$\sum_{k=1}^{\infty} k(k+1)q^{k-1} = \frac{d^2\left(\frac{q^2}{1-q}\right)}{dq^2} = \frac{2}{(1-q)^3} = 2p^{-3}. \tag{4}$$

Bring the result in Eqn 4 back to Eqn 3.

$$\sum_{k=1}^{\infty} k^2(1-p)^{k-1} = 2p^{-3} - p^{-2}. \tag{5}$$

From Eqn 5 and Eqn 2, we have $\mathrm{E}[X^2] = p(2p^{-3} - p^{-2}) = 2p^{-2} - p^{-1}$ and

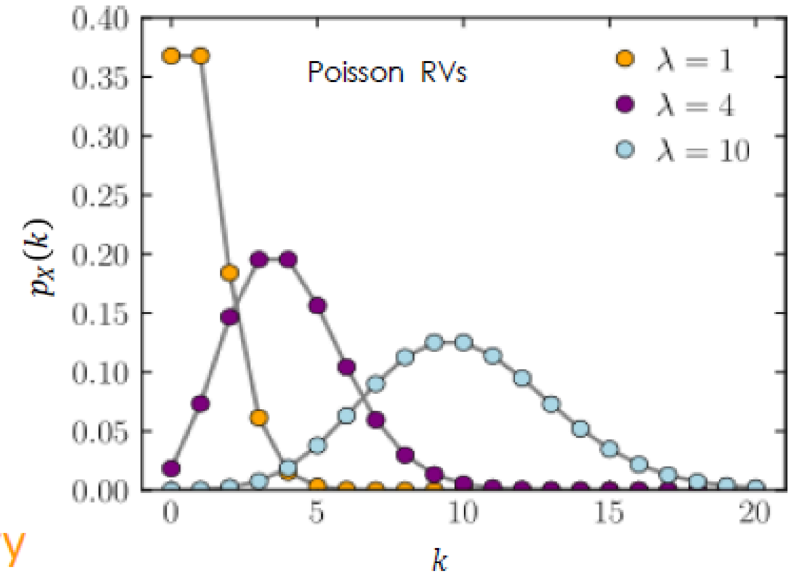
$$\sigma_X^2 = \mathrm{E}[X^2] - m_X^2 = 2p^{-2} - p^{-1} - p^{-2} = p^{-2} - p^{-1} = \frac{1-p}{p^2}.$$

Poisson

- Poisson RV $X \sim \text{Poisson}(\lambda)$.
  - Describe how many times that a random event occurs in a specific time interval. There are a few assumptions.
    - Events are independent.
    - Two events do not occur at the same time.
    - The number of times $k = 0,1,2,\cdots$.
    - The average of k is a constant (denoted as λ) in a time interval.
  - Examples
    - Number of buses arriving at a bus stop in a time interval.
    - Number of data packets arriving at a router in a time interval.
    - Number of calls received by a call centre in a time interval.

- multiple task for mcu, how often can mcu takes cares of the task, what's the processing speed to take care of 90% of the tasks
- support 0 to inf
- Distribution is defined by lambda and the time interval

– PMF $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ with $S_X = \{0,1,2,\cdots\}$



– Expectation and variance
  - Expectation $m_X = \lambda$
  - Variance $\sigma_X^2 = \lambda$

(Refer to the supplementary materials for the derivation.)

- lambda is average appearance in the given time interval
- $P_x(x = 0) = e^{-\lambda}$
- Variance = Expectation

- Shape of plot
    - curve maximum shift with lambda
    - as k → infinity, p → 0
- Sum of P from k = 0 to k = infinity, = 1
    - Assume $X \sim \text{Poisson}(p)$ that its PMF is $p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ and it support $S_X = \{0, 1, 2, \cdots\}$.

      Although by the probability axioms, we know that

      $$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1.$$

      But it is not obvious from the expression. So let's first prove $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$.

      It is easy to prove by using the Taylor series expansion of e^x, which is $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. When $x = \lambda$. The sum of all the Poisson PMF terms is

      $$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = = e^{-\lambda} \times e^{\lambda} = 1. \qquad (1)$$

<u>Derivation for E[X]</u>

By definition, Poisson RV X expectation is

$$m_X = \sum_{k=0}^{\infty} k p_X(k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}.$$

Take out the first term, which is zero. Let $k' = k - 1$,

$$m_X = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} + 0 \times \frac{\lambda^0}{0!} e^{-\lambda} = \sum_{k=1}^{\infty} \lambda \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}$$

$$= \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} e^{-\lambda} = \lambda.$$

Note that in the last step above, $\sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!} e^{-\lambda} = 1$ as proved in Eqn (1).

Derivation for Var[X]

To find the variance of X, we make use of $\sigma_X^2 = \mathrm{E}[X^2] - m_X^2$, where $m_X = \lambda$ and

$$\mathrm{E}[X^2] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} \qquad (2)$$

Since $k^2 = k(k-1) + k$. Eqn 2 can be rewritten as

$$\mathrm{E}[X^2] = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}. \qquad (3)$$

The second term above is the expectation of X, which is λ. The first term above is

$$\sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = 0 \times (0-1) \frac{\lambda^0}{0!} e^{-\lambda} + 1 \times (1-1) \frac{\lambda^1}{1!} e^{-\lambda} + \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda}$$

The first two terms in the equation above are separated from the summation for case $k = 0$ and 1. They both contribute zero to the summation and therefore can be ignored. By changing the variable $k'' = k - 2$, the above third term becomes

$$\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} = \lambda^2 \sum_{k''=0}^{\infty} \frac{\lambda^{k''}}{k''!} e^{-\lambda} = \lambda^2 \qquad (4)$$

In Eqn 4, we make use of the fact that $\sum_{k''=0}^{\infty} \frac{\lambda^{k''}}{k''!} e^{-\lambda} = 1$ as shown in Eqn 1. The variance of X is

$$\sigma_X^2 = \mathrm{E}[X^2] - m_X^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Example 1

for lambda = 5, which k is the largest?

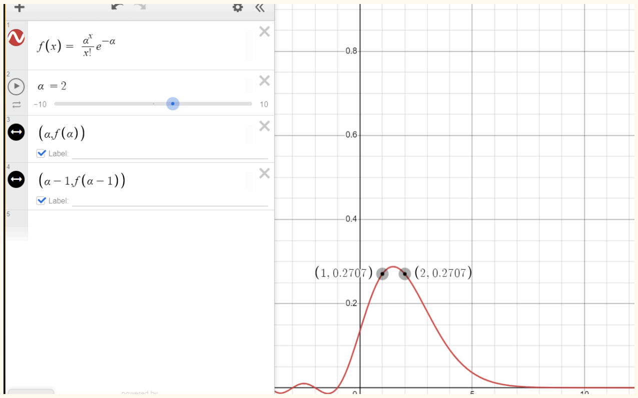
5*5*...5

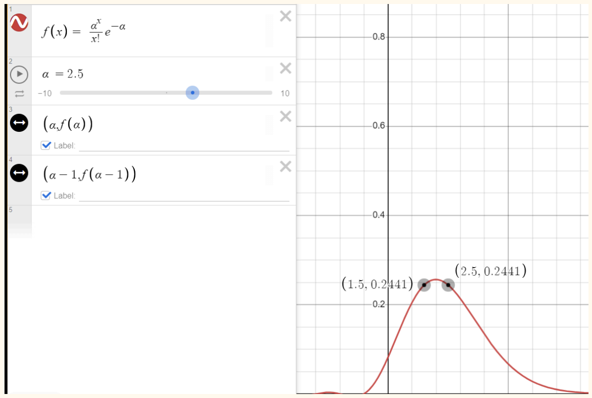
1*2*3…k

will peak at k = 5, after that denominator dominates

at k = 4 to k = 5 is multiplied by 5/5, so they are both values of k for which the p is largest,

- this is true for all integer lambda, in fact P(lambda) = P(lambda - 1) for all lambda but it is only the largest probability if lambda is integer
- for non integer lambda

- 



- since k is only integer, if lambda is not integer, it cannot be the peak of a discrete plot, the peak must be between lambda and lambda - 1
- https://www.desmos.com/calculator/az7o8bpqmt

before that is still smaller that k =5 or k=4

Example 2

[Arrivals at a packet router] The number of packets arrival in t seconds at a router is a Poisson RV with $\lambda = \alpha t$, where α = 4000 packets/second is the average arrival rate. Find the following event probabilities.

- More than 2 packets in 1 milliseconds.

  Let X be the RV representing the number of packets in a period of 1 ms $=$ $10^{-3}s$. $\lambda = 4000 \times 10^{-3} = 4$.

  $$P(X > 2) = \sum_{k=3}^{\infty} p_X(k) = 1 - \sum_{k=0}^{2} p_X(k) = 1 - p_X(0) - p_X(1) - p_X(2)$$

  $$= 1 - e^{-4}\left(1 + 4 + \frac{4^2}{2}\right) = 0.7619$$

- average arrival rate or arrival rate are different from lambda, lambda is the rate*time period

At most 2 packets in 2 milliseconds.

Let Y be the packets in 2 millisecond. $Y \sim \text{Poisson}(\lambda)$ where

$\lambda = 4000 \times 2 \times 10^{-3} = 8.$

$$P(Y \leq 2) = p_Y(0) + p_Y(1) + p_Y(2) = e^{-8}\left(1 + 8 + \frac{8^2}{2}\right) = 0.0138.$$

- time period different, distribution will be different because lambda is different

No packet arrivals in 1 milliseconds.

$P(X = 0) = e^{-4} = 0.0183$. ($\lambda = 4$ since time duration is 1 ms.)

Let packet arrival time interval be T. This result also means

$P(T > 10^{-3}s) = P(X = 0) = e^{-4}$.

It can be generalized to $P(T > t) = e^{-\lambda} = e^{-\alpha t}$ for any $t > 0$.

- If there's no packet in 1ms then the interarrival time is greater than 1ms
- P(interarrival time is greater than 1ms) = P(if there's no packet in 1ms)

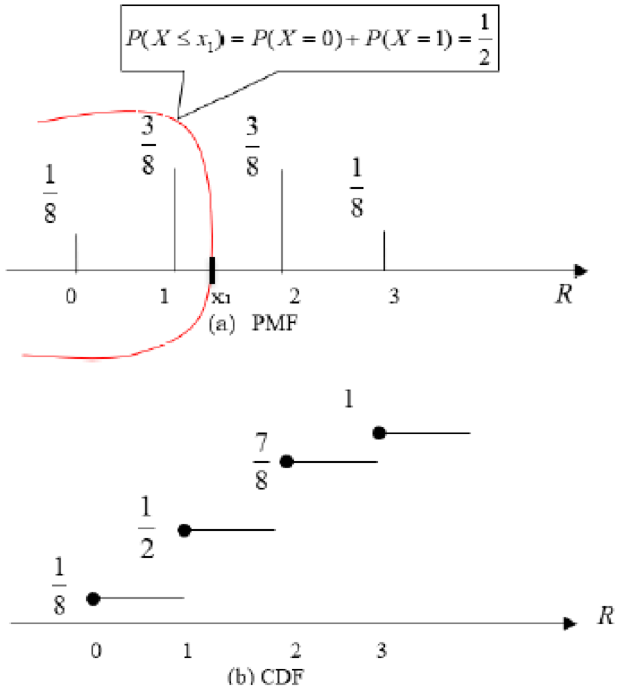
Cumulative Distribution Functions (CDF)

Cumulative distribution functions (CDF)

notation

definition

- CDF definition: $F_X(x) \triangleq P(X \leq x)$ for $x \in R$.
  - Staircase shape for discrete RVs.
  - Compare with PMF.



- Properties of CDF
  - $0 \leq F_X(x) \leq 1$ for all x.
  - $\lim_{x \to -\infty} F_X(x) = 0$
  - $\lim_{x \to \infty} F_X(x) = 1$
  - $F_X(x)$ is non-decreasing.
  - $F_X(x)$ is right-continuous.

- x is any real, not necessarily from support
- for discrete variable, it's staircase shape
- Non decreasing as probability can only add up, but the maximum at 1
- Right continuous = from right to left can approach value but from left to right cannot
  - FX(0.9) = 1/ 8 reach to 1 from the left can never include the p at 1
  - FX(1.1) = 1 / 2 reach to 1 from the right can include the p at 1

CDF

Evaluate probability using CDF.

Assume $a < b$ and $a, b \in R$.

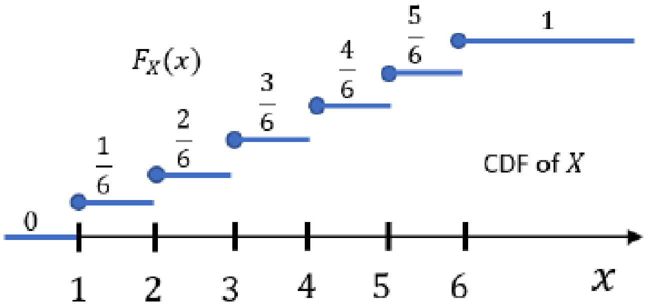
- $P(X \leq a) = F_X(a)$
- $P(X < a) = F_X(a) - p_X(a)$ (Note $p_X(a) = 0$ if $a \notin S_X$ with a bit abuse of notation.)
- $P(X > b) = 1 - P(X \leq b) = 1 - F_X(b)$
- $P(X \geq b) = 1 - P(X < b) = 1 - F_X(b) + p_X(b)$
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$
- $P(a < X < b) = P(X < b) - P(X \leq a) = F_X(b) - p_X(b) - F_X(a)$
  $= F_X(b) - F_X(a) - p_X(b)$
- $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F_X(b) - F_X(a) + p_X(a)$
- $P(a \leq X < b) = P(X < b) - P(X < a)$
  $= [F_X(b) - p_X(b)] - [F_X(a) - p_X(a)]$
  $= F_X(b) - F_X(a) - p_X(b) + p_X(a)$

- abuse because a is only from support, if a not from support , P = 0

Example

RV X is the number obtained when a fair dice is rolled. Find the following probabilities $P(X \leq 3)$, $P(X > 4)$ and $P(2 < X \leq 5.3)$.

- The support of X is $S_X = \{1,2,3,4,5,6\}$ and its PMF is $p_X(x) = 1/6$ for $x \in S_X$. Its CDF $F_X(x)$ is illustrated by the staircase diagram.
  $P(X \leq 3) = F_X(3) = 1/2$
  $P(X > 4) = 1 - F_X(4) = 1/3$
  $P(2 < X \leq 5.3) = F_X(5.3) - F_X(2) = 1/2$



- Using the PMF,
  $P(X \leq 3) = P(X \in \{1,2,3\}) = 1/2$
  $P(X > 4) = P(X \in \{5,6\}) = 1/3$
  $P(2 < X \leq 5.3) = P(X \in \{3,4,5\}) = 1/2$

Gambling Example

Gambling game settings

- Finite amount of money for each player.
  - They can have equal or different initial funds.
  - A player exits the game if the player has zero balance.
- The same game is independently played n rounds.
- In each round, each player put the same amount of bid. Winning player wins bid. Otherwise, the player loses his bid to winner.
- Players are independent.

- Equal chance but different amount of money
  - A player wins or loses one round with probability 0.5.
  - To make it simple, only two players A and B with the same bid d each round.
  - Initially, player A has money p and player B has money q.

Assume n rounds are played. For player A, his gain in each round is an random variable X_i. His overall gain is $X = \sum_{i=1}^{n} X_i$, which has expectation

$$\mathrm{E}[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} 0.5 \times d + 0.5 \times (-d) = \sum_{i=1}^{n} 0 = 0$$

On the other hand, by the time the game stops, player A either loses all his money ($-p$) or takes all player B's money ($+q$). That is his final gain is either $-p$ or q. Assume these two results occurs with probability $P_{\mathrm{A}0}$ and P_{A1}.

$$P_{\mathrm{A}0} + P_{A1} = 1 \qquad \text{---- (1)}$$

Play A's expected gain is zero based on the previous analysis.

$$E[X] = (-p) \times P_{\mathrm{A}0} + q \times P_{A1} = 0 \qquad \text{--- (2)}$$

Based on (1) and (2), we have

$P_{A1} = \frac{p}{p+q}$ → Probability A wins = Probability B loses.

$P_{A0} = \frac{q}{p+q}$ → Probability A loses = Probability B wins.

- If $p > q$, player A has a higher chance to win over player B.
- In general , players with more money wins with higher chance.
- As individual player, do you have more money than casino?
- What if I am the lucky winner?

- chance of winning proportional to initial fund

A player almost sure loses when addicted to gambling.

- Player A has finite money p and bid d each round. He can afford at most $n_A = p/d$ continuous loses (rare but the fastest way to exit the game).
- Common players lose for almost sure if he keeps playing.
  - Assume player A loses in each round with probability p_0.
  - The probability of a sequence of n_A losses is $P_1 = p_0^{n_A}$.
  - The probability P_1 may be small. But such long sequence of loses will happen almost sure because $1-(1-P_1)^k \to 1$, if the player keeping playing (k becomes larger). → Common player will lose all sooner or later if keeping playing.
  - For casino, its n_A is significant larger and P_1 is much smaller. It can survive before its opponent exits and get recovered from loss.

1-P1 is the prob of not losing in nA time, then addicted will go for another nA time, for k time
prob of loosing = 1 - not losing for k times = 1 - (1-p1)^k → 1

- "You can hardly win over me in gambling because I take a commission from your bets" -- Stanley Ho

- Conclusion:
  - Small amount of gambling is entertaining, but excessive gambling harms (小赌怡情，大赌伤身).

Chapter 4 Multiple Random Variables and Functions of Random Variables

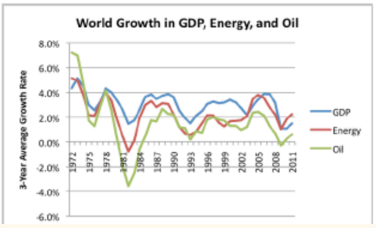
- joint behaviour of data, each data is one variable, for data science and machine learning

Multiple discrete RVs

- Represent different (random) quantities of a random experiment.
- They are often correlated.
  - Height H and weight W of students in a high school. They are related by nutrition information.

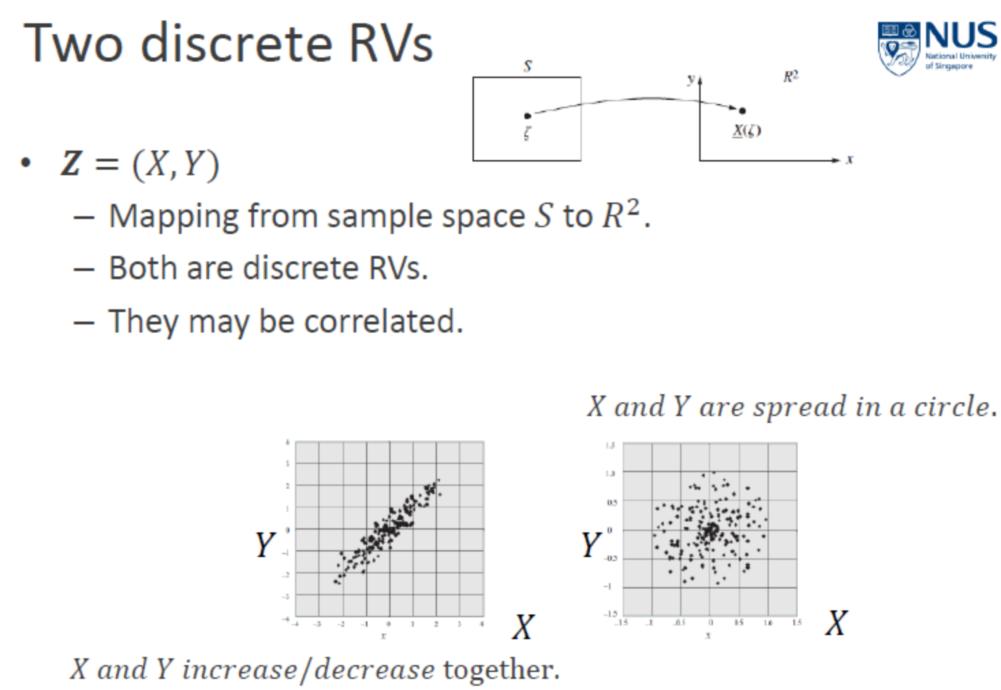
    Event $A = \{H \leq 120 \,\&\, W \leq 40\}$ → undernutrition
  - GDP and energy consumption?

    GDP and petrol price?



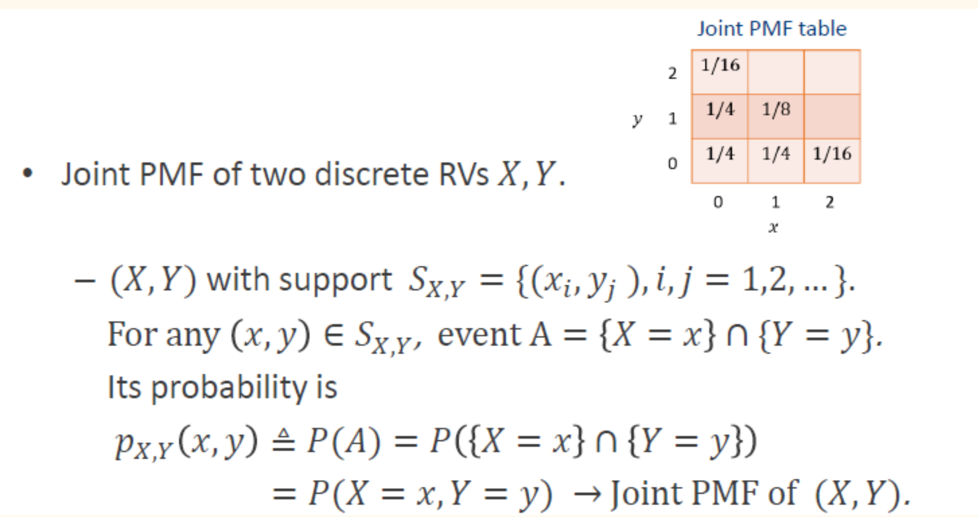
- Joint distribution, and from that the correlation of the variables

Mapping



- Instead of mapping each sample into a number line, it's mapping to a 2 dimensional space.
    - N RVs → R^N
- If has a trend, then may be correlated ⇒ dependent
    - example: circle plot above
        - for a given X, the value Y can take is different, the probability of the value of Y is also different
        - the value Y depends on the value of X
        - X and Y are dependent
- Uncorrelated can be not independent

Two discrete RVs



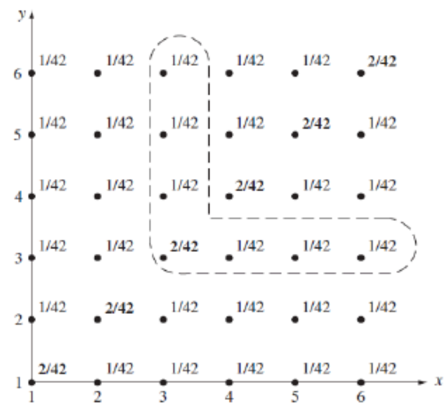
- Joint distribution ⇒ Joint PMF
    - Means need intersect for the joint event
    - Must occur simultaneously
    - Can be in a table or a closed form if exist
    - Can have Expected and Variance
    - Joint behaviour of the two variable

$$Joint\ PMF = p_{X,Y}(x,\ y) = P(\{X = x\} \cap \{Y = y\}) = P(X = x,\ Y = y)$$

- Note that its X = x comma Y = y, not intersect
- Number of support for each variable does not necessarily need to be the same
- Sum of all possibility is also 1

Example 1

A random experiment of rolling two dice and recording the number pair (X, Y). Their joint PMF $p_{X,Y}(x, y)$ is given by the two-dimensional table shown below. Find $P(\min(X, Y) = 3)$.



- $x, y \in \{1,2,3,4,5,6\}$ and $|S_{X,Y}| = 36$.
- $\mathrm{A} = \min(X, Y) = 3$ can be $(3, \geq 3)$ and $(\geq 3, 3)$.
- $P(A)$ is the sum of these probabilities indicated by the dotted-line loop.

$$P(A) = p_{X,Y}(6,3) + p_{X,Y}(5,3) + p_{X,Y}(4,3) + p_{X,Y}(3,3) + p_{X,Y}(3,4) + p_{X,Y}(3,5) + p_{X,Y}(3,6) = \frac{4}{21}.$$

- Are these two dice loaded?

- 6 * 6 possible combinations, but does not mean all combination are of possible
- Dice are not biassed ⇒ fair dice
  - Can see from the marginal probability
  - P(X = x) = 7/42 = 1/6 for all x
  - P(Y = y) = 7/42 = 1/6 for all y
- The distribution look “weird” is because the dice are not independent
  - The probability of getting same value is higher
- Fair but loaded.
  - loaded can means not fair or dependent
  - Fair and unloaded is no the same

Marginal PMF

- If the joint PMF $p_{X,Y}(x,y)$ of (X,Y) is known, how to find the PMF $p_X(x)$ of X?
  - $X = x_i$, Y can be any possible value.

    $\{X = x_i\} = \cup_{(x_i,\, y_j) \in S_{X,Y}} \{X = x_i, Y = y_j\}$

    $P(X = x_i) = \sum_{(x_i,\, y_j) \in S_{X,Y}} p_{X,Y}(x_i, y_j) = \sum_{\text{all } y_j} p_{X,Y}(x_i, y_j).$

    Similarly,

    $P(Y = y_j) = \sum_{(x_i,\, y_j) \in S_{X,Y}} p_{X,Y}(x_i, y_j) = \sum_{\text{all } x_i} p_{X,Y}(x_i, y_j).$

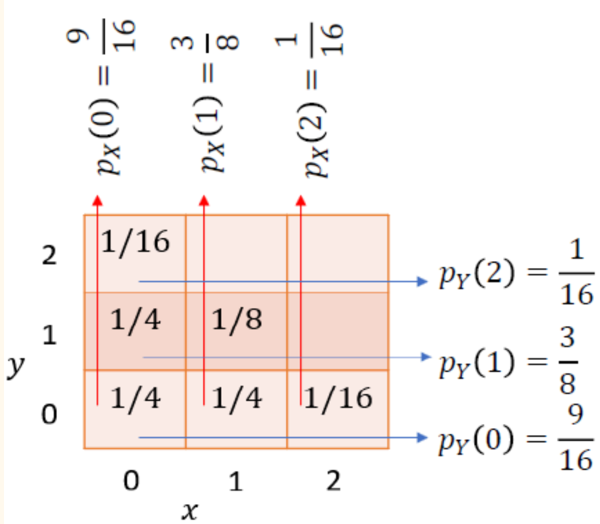
- Fix one given value Y (or X) and sum all the probability X (or Y)

$$p_X(x) = P(X = x) = P\Big(\bigcup_{y_j \in Y} \{X = x,\ Y = y_j\}\Big) = \sum_{y_j \in Y} P(X = x,\ Y = y_j)$$

Joint PMF → Marginal PMFs

- The table below illustrates how the marginal probabilities are computed.

joint PMF → marginal PMFs



Sum of the columns or rows in the joint PMF table.

Marginal PMFs are also probabilities. Probability Axioms applies.

marginal PMFs → joint PMF ?

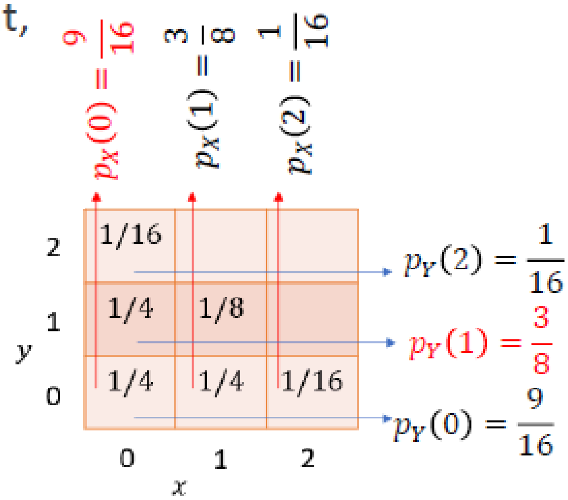
- Graphically equivalent to summing along the column and rows
  - because they are disjoint
- Same as single variable PMF

Marginal PMFs → Joint PMF

– Marginal PMFs → joint PMF ?

In some cases, $p_{X,Y}(x,y) = p_X(x) \times p_Y(y)$. But it is not working in most of the time. For example, in the figure on the right,

$$p_{X,Y}(0,1) = \frac{1}{4} \neq p_X(0) \times p_Y(1) = \frac{9}{16} \times \frac{3}{8}$$



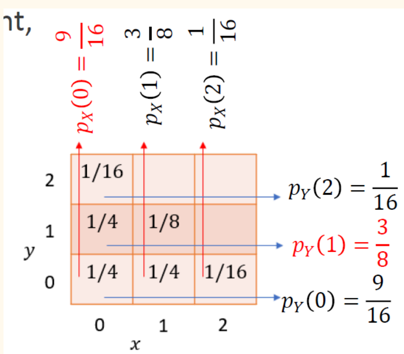
– How are X and Y associated?

The question asked by data scientists.

- Independent
- Correlated/Non-correlated

- Not possible for all cases as information is lost when Joint PMF → Marginal PMFs
  - knowing the PMF of two variables separately, not possible to find the relationship/joint behaviour
- If all the joint probability is the product of the marginals probabilities means the two variables are independent
  - if true for some cases, means not independent and coincidence (2+2 = 2*2)
  - if false for 1 case, means not independent
  - if true for all case, then independent

Example



P(X=1|Y=1) = 1/8 / (1/4 + 1/8) = 1/3
P(X=1 intersect Y=1) = joint probability
P(Y=1) = marginal probability

Independence of Two RVs

- If $P(AB) = P(A)P(B)$, events A and B are independent.
  - All events defined only by A are independent with events defined only by B.
  - RVs X and Y can represent many (infinite) events, one needs to check if all of them are independent? ← tedious!

- Using the old definition, have to check all events, not possible with distribution of infinite support, like geometric and Poisson

- The **necessary** and **sufficient** condition that two discrete RVs X and Y are independent is

$$p_{X,Y}(x, y) = p_X(x) \times p_Y(y)$$ (Refer to the supplementary materials for the proof.)

- If X and Y are independent, $g(X)$ and $h(Y)$ are also independent. ($g(\cdot)$ and $h(\cdot)$ are deterministic functions.)

- If it has a close form, then it's easy to check using the PMF.
  - if the probability of z has x inside the close form ⇒ z is dependent on x
- Functions of independent variable are also independent, deterministic means the function depends on the input variable
  - A deterministic function always gives the same result, for the same input
  - This is as opposed to a probabilistic function.
  - If we denote the function mapping by $\mathbf{F}(\cdot)$, two observations can be generalized to this type of problem.

    - All $x_i \in A_k$, where $A_k \subseteq S_X$, may be mapped to the same $y_k \in S_Y$. That is $y_k = \mathbf{F}(x_i)$ for all $x_i \in A_k$.
    - The PMF of Y can be calculated by $p_Y(y_k) = P(Y = y_k) = \sum_{x_i \in A_k} P(X = x_i) = \sum_{x_i \in A_k} p_X(x_i)$. Therefore, $y_k \times p_Y(y_k) = \sum_{x_i \in A_k} \mathbf{F}(x_i) \times p_X(x_i)$.

    By the expectation definition and the above observations, we have a general result on the expectation for the function of an RV as below.

$$\mathrm{E}[Y] = \sum_{y_k \in S_Y} y_k \times p_Y(y_k) = \sum_{y_k \in S_Y} \sum_{x_i \in A_k} \mathbf{F}(x_i) \times p_X(x_i) = \sum_{x_j \in S_X} \mathbf{F}(x_j) \times p_X(x_j).$$

    Applying this result to this problem, we have

$$\mathrm{E}[Y] = \sum_{x_j \in S_X} x_j^2 p_X(x_j) = (-1)^2 \times p_X(-1) + 0^2 \times p_X(0) + 1^2 \times p_X(1) = 2/3.$$

Proof

Let X and Y be a pair of independent discrete RVs. Assume event A_1 only involves RV X and event A_2 only involves RV Y. So A_1 and A_2 are independent because X and Y are independent.

Specifically, let $A_1 = \{X = x_i\}$ and $A_2 = \{Y = y_j\}$ where (x_i, y_j) is any possible pair from $S_{X,Y}$. We have the probability of the joint event $A_1 \cap A_2$ is

$P(A_1 \cap A_2) = P(X = x_i, Y = y_j) = p_{X,Y}(x_i, y_j)$.

Also, since A_1 and A_2 are independent, the probability of the joint event $A_1 \cap A_2$ can also be

$P(A_1 \cap A_2) = P(A_1)P(A_2) = P(X = x_i)P(Y = y_j) = p_X(x_i)p_Y(y_j)$.

Therefore, if two discrete RVs X and Y are independent, for any $(x_i, y_j) \in S_{X,Y}$, it is true that

$$p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j).$$

We next show if $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$ for any $(x_i, y_j) \in S_{X,Y}$, then

$$P(A_1 \cap A_2) = P(A_1)P(A_2).$$

$P(A_1 \cap A_2) = \sum_{x_i \in \{x:A_1\}} \sum_{y_j \in \{y:A_2\}} p_{X,Y}(x_i, y_j)$. Since $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$,

$$P(A_1 \cap A_2) = \sum_{x_i \in \{x:A_1\}} \sum_{y_j \in \{y:A_2\}} p_X(x_i)p_Y(y_j) = \sum_{x_i \in \{x:A_1\}} p_X(x_i) \sum_{y_j \in \{y:A_2\}} p_Y(y_j) = P(A_1)P(A_2).$$

Since A_1 and A_2 are any events involves X and Y, respectively, we conclude that X and Y are independent.

Example 1

The joint PMF of two RVs X and Y are shown in the figure on the right. Are they independent?

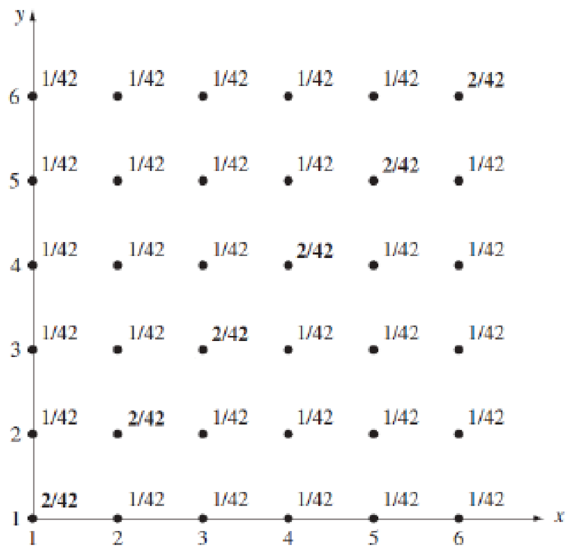
- From the joint PMF, the marginal PMFs

$$p_X(x) = \frac{1}{6} \text{ for } x \in \{1,2,3,4,5,6\}.$$

$$p_Y(y) = \frac{1}{6} \text{ for } y \in \{1,2,3,4,5,6\}.$$

- It can be verified that at least one (x_i, y_j) does not satisfy $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$.
- Therefore, they are dependent.
- Actually, $p_{X,Y}(x_i, y_j) \neq p_X(x_i)p_Y(y_j)$ for all possible (x_i, y_j) in the table. But verifying one is sufficient to illustrate their dependency.

If you want to verify for independent, you need to check $p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j)$ for all possible (x_i, y_j).



- Given the joint pmf, sum columns and rows get marginal pmf
- then check all case whether satisfy independence, if one fails the check means dependent
- Dice are fair but loaded
    - individually is fair, when together has dependent

Example 2

A coin is flipped m times in the morning, the number of heads is x. It's flipped n times in the afternoon, the total number of heads is z. Are x and z independent?

- Let y = the number of heads in the afternoon
- X ~ B(m, p)
- Y ~ B(n, p)
- Z = X + Y
- Z ~ B(m+n, p)
- P(X=x, Y=y) = P(X=x | Y=y) P(Y=y) = P(X=x) P(Y=y)
    - X and Y are independent

    - By intuition, if between two flips is independent, between two batches of flips is also independent
- P(X=x, Z=z) = P(Z=z|X=x) P(X=x)
    - = [m C z-x (p)^(z-x) (1-p)^(m - (z-x))] [m C x (p)^(x) (1-p)^(m - x)]
    - P(Z=z|X=x) = [m C z-x (p)^(z-x) (1-p)^(m - (z-x))] != (m+n) C z (p)^(z) (1-p)^(m+n - z) = P(Z=z)
    - given a value of x, P(Z=z|X=x) depends on x, thus Z and X not independent

<u>Covariance and Correlation Coefficient</u>

Covariance between X and Y

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

Correlation coefficient between X and Y

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

where $\sigma_X = \sqrt{\mathrm{Var}[X]}$ and $\sigma_Y = \sqrt{\mathrm{Var}[Y]}$ are the standard deviations of X and Y.

- $-1 \leq \rho_{X,Y} \leq 1$ (Refer to the supplementary materials for the proof.)
- If $Y = aX + b$ with a and b being real constants, then $\rho_{X,Y} = 1$ when $a > 0$ and $\rho_{X,Y} = -1$ when $a < 0$.
- If $\rho_{X,Y} = 0$, then X and Y are uncorrelated.

- If two variable are dependent, want to know more about the relationship thus correlation and covariance
- Correlation coefficient
    - uncorrelated does not mean independent but independent means uncorrelated
        - correlation and independent are not the same
        - intuition is that dependent can have many relationships not just linear, just because not linearly associated doesn't mean not associated.
    - correlation only for linear association, doesnt tell about other relationship
        - but correlation good because can linearize or when range is small can assume linear

$$Cov[X,\ Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

- Derivation

$$\begin{aligned}
&Cov[X,\ Y] \\
&= E[(X - E[X])(Y - E[Y])] \\
&= E[XY - E[X]Y - E[Y]X + E[X]E[Y]] \\
&= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y]] \\
&= E[XY] - E[X]E[Y]
\end{aligned}$$

- E[XY] = E[X]E[Y] if and only if x and y are independent ⇒ covariance is 0 ⇒ uncorrelated
    - independent ⇒ uncorrelated
    - not the other way around because covariance = 0, E[XY] = E[X]E[Y] can have coincident

- If two RVs X and Y are independent, their joint PMF satisfies

  $p_{X,Y}(x,y) = p_X(x)p_Y(y)$.

  $E[XY] = \sum_{(x,y)\in S_{X,Y}} xy \times p_{X,Y}(x,y) = \sum_{(x,y)\in S_{X,Y}} xy \times p_X(x) \times p_Y(y)$

  $= \sum_{x\in S_X} xp_X(x) \times \sum_{y\in S_Y} yp_Y(y) = E[X]E[Y]$

  Their covariance is $\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y] = 0$.

- Covariance of X and X is the variance of X, Cov[X, X] = $E[X^2]$ - $E[X]^2$ = Var[X]

$$\rho_{X,Y} = \frac{Cov[X,Y]}{\sigma_X \sigma_Y}$$

- Correlation coefficient is normalised covariance by their standard deviation. Because if the variance/dynamic range of one is big, covariance will get affected but correlation coefficient wont.
- Linear relationship is symmetric, if y = ax + b ⇒ x = cx + d
- Analysing ϱ
  - ϱ = 1, can fit perfectly in a positive gradient line
  - Cov > 0, ϱ > 0, positively linearly correlated
  - Cov = 0, ϱ = 0, not linearly correlated, not necessarily independent
  - Cov < 0, ϱ < 0, negatively linearly correlated
  - ϱ = -1, can fit perfectly in a negative gradient line
- Covariance of X, X is 1 as Cov[X, X] = Var[X]

if Y is a constant ⇒ not related to X??

Proof for -1 <= ϱ <= 1

- Proof of $-1 \le \rho_{X,Y} \le 1$.

  Z^2 is always positive or zero and $E[Z^2] \ge 0$ for any RV Z.

  Let $Z = \frac{X-E[X]}{\sigma_X} \pm \frac{Y-E[Y]}{\sigma_Y}$.

$$Z^2 = \left(\frac{X-E[X]}{\sigma_X}\right)^2 \pm 2\frac{X-E[X]}{\sigma_X} \times \frac{Y-E[Y]}{\sigma_Y} + \left(\frac{Y-E[Y]}{\sigma_Y}\right)^2$$

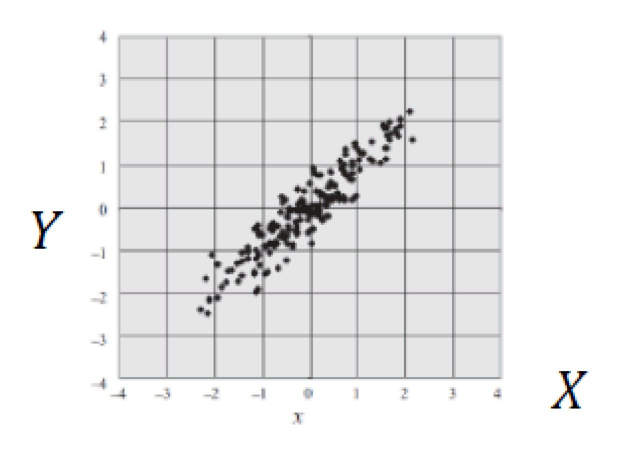
  By the linearity of expectation operation,

$$E[Z^2] = E\left[\left(\frac{X-E[X]}{\sigma_X}\right)^2\right] \pm 2E\left[\frac{X-E[X]}{\sigma_X} \times \frac{Y-E[Y]}{\sigma_Y}\right] + E\left[\left(\frac{Y-E[Y]}{\sigma_Y}\right)^2\right]$$

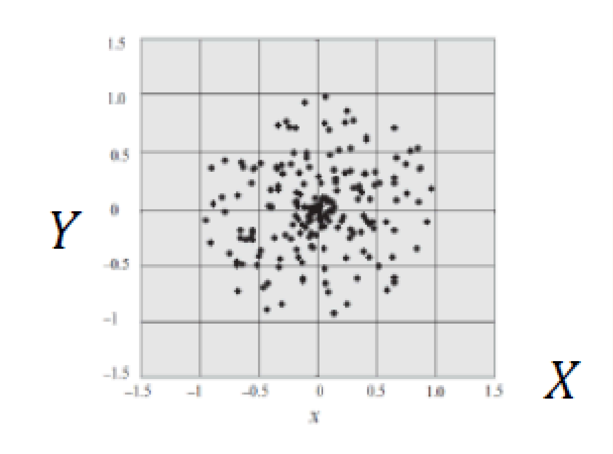
$$= \frac{E[(X-E[X])^2]}{\sigma_X^2} \pm 2\rho_{X,Y} + \frac{E[(Y-E[Y])^2]}{\sigma_Y^2} = 1 \pm 2\rho_{X,Y} + 1$$

  Since $E[Z^2] \ge 0$, $1 \pm 2\rho_{X,Y} + 1 \ge 0$ which shows $-1 \le \rho_{X,Y} \le 1$.

Example 1



- can fit in line ⇒ linearly correlated ⇒ must be dependent



- cannot fit in line ⇒ linearly uncorrelated
- Range of value for y depends on the value of x ⇒ must be dependent

Example 2

The PMF of X is shown in the table on the right. Another RV $Y = X^2$. Find the correlation coefficient between X and Y.

X	-1	0	1
$p_X(x)$	1/3	1/3	1/3

- Based on the PMF of X, the joint PMF and marginal PMFS of X and Y are derived in the table below.

$$E[X] = -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$$
$$E[Y] = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$$
$$E[XY] = \sum_{(x,y)\in S_{X,Y}} xy \times p_{X,Y}(x,y)$$
$$= -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$$

$p_{X,Y}$	-1	0	1	$p_Y(y)$
0	0	1/3	0	1/3
1	1/3	0	1/3	2/3
$p_X(x)$	1/3	1/3	1/3	

Cov$[X,Y] = E[XY] - E[X]E[Y] = 0$ and $\rho_{X,Y} = 0$. X and Y are uncorrelated.

But Y is derived from X and they are dependent.

Functions of one RV

- If X is a RV and $g(\cdot)$ is a deterministic function, then $Y = g(X)$ is also a RV.
  - $Y = aX + b, Y = X^2$, …
    - Temperature and sensor measurements.
    - Power produced by a circuit with random power level.
- What is the distribution of Y if the distribution of X is known?
  - Determine S_Y: The support of X is S_X. For each $x \in S_X$, $g(x) \in S_Y$. → $S_Y = \{y: y = g(x) \text{ for } x \in S_X\}$ - Make a table.
    If S_Y is countable, Y is discrete. → PMF?
    If S_Y is uncountable, Y is continuous → PDF? (next chapter)
  - Determine PMF/PDF of Y.

- Measurement of a quantity, is not the true value of the quality, it's the reading, so there is a transfer function, if quantity is RV, the reading is also RV.
- If Y is a function of RV X, Y is also a RV. Thus it is in interest to find the support and the PMF or PDF of y.

Example 1

X is a discrete uniformly distributed RV with $S_X = \{-1, 0, 1, 2\}$.

Let $Y = g(X) = X^2$. Find the PMF of Y.

- The support of Y is determined by finding all the values $y = g(x)$ for $x \in S_X$.

  A mapping table can be created to illustrate.

X	-1	0	1	2
Y	1	0	1	4

$\rightarrow S_Y = \{0, 1, 4\}$.

- To find the PMF of Y, refer to the mapping table above.

  $P(Y = 0) = P(X = 0) = 0.25$

  $P(Y = 1) = P(X = 1) + P(X = -1) = 0.5$

  $P(Y = 4) = P(X = 2) = 0.25$

  We can show the PMF by a table on the right.

Y	0	1	4
p_Y	0.25	0.5	0.25

Extra Q: Find the joint PMF for X and Y

P(X=x)	0.25	0.25	0.25	0.25	1
Y=4	0	0	0	0.25	0.25
Y=1	0.25	0	0.25	0	0.5
Y=0	0	0.25	0	0	0.25
	X=-1	X=0	X=1	X=2	P(Y=y)

P(Y=y, X=x) = P(Y=y | X=x) P(X=x)

Example 2

A fair dice is used in a game. Let X be the number in a roll. The game reward is $Y = (X-3)^2$. How much a player willing to pay to play the game?

- Determine the support of Y using the table below.

X	1	2	3	4	5	6
Y	4	1	0	1	4	9

- The PMF of Y can be shown using the table below.

Y	0	1	4	9
p_Y	1/6	1/3	1/3	1/6

- The expectation $E[Y] = 19/6$. If the fee is less than $E[Y]$, the player has an edge to play.

- Will be willing to pay to play if can win money ⇒ if the expected money price is more than fees
- Edge is called the fees - expected price

Two approaches to find the expectation of $Y = g(X)$.

Assume S_X, S_Y are the respective supports of X and Y.

– $E[Y] = \sum_{y \in S_Y} y \times p_Y(y)$

The expectation $E[Y]$ in the previous example is

$$E[Y] = 0 \times \frac{1}{6} + 1 \times \frac{1}{3} + 4 \times \frac{1}{3} + 9 \times \frac{1}{6} = \frac{19}{6}.$$

– $E[Y] = \sum_{x \in S_X} g(x) \times p_X(x)$

The expectation $E[Y]$ can also be computed as

$$E[Y] = \frac{1}{6}\left((1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2 + (6-3)^2\right) = \frac{19}{6}$$

Example 3

If $X \sim \text{Bin}(4, p)$, what is the PMFs for $Y = 2X + 1$ and $Z = |X - 2|$?

- We can list the values of X, Y and Z in a table.

  The last row shows the PMF of X, where

  $p_X(k) = \binom{4}{k} p^k (1-p)^{4-k}$

X	0	1	2	3	4
Y	1	3	5	7	9
Z	2	1	0	1	2
p_X	$p_X(0)$	$p_X(1)$	$p_X(2)$	$p_X(3)$	$p_X(4)$

- From this table, we have $S_Y = \{1, 3, 5, 7, 9\}$.

Y	1	3	5	7	9
p_Y	$p_X(0)$	$p_X(1)$	$p_X(2)$	$p_X(3)$	$p_X(2)$

$P(Y = 1) = P(X = 0) \rightarrow p_Y(1) = p_X(0)$

$\vdots \qquad\qquad \vdots$

$P(Y = 9) = P(X = 4) \rightarrow p_Y(9) = p_X(4)$ as shown in the table above.

We find that $p_Y(y) = p_X\left(\frac{y-1}{2}\right)$ for $y \in S_Y$ because $Y = 2X + 1$.

Can we generalize the idea to the PMF of Z?

X	0	1	2	3	4
Y	1	3	5	7	9
Z	2	1	0	1	2
p_X	$p_X(0)$	$p_X(1)$	$p_X(2)$	$p_X(3)$	$p_X(4)$

- From the right table, we see $S_Z = \{0, 1, 2\}$.

$P(Z = 0) = P(X = 2) \rightarrow p_Z(0) = p_X(2)$

$P(Z = 1) = P(X = 1) + P(X = 3)) \rightarrow p_Z(1) = p_X(1) + p_X(3)$

$P(Z = 2) = P(X = 0) + P(X = 4)) \rightarrow p_Z(2) = p_X(0) + p_X(4)$

Z	0	1	2
p_Y	$p_X(2)$	$p_X(1) + p_X(3)$	$p_X(0) + p_X(4)$

Comparing the PMF of Z with the PMF of Y, one cannot use the inverse function method because $Z = |X - 2|$ does not have an inverse function!

- Note that $P_Y(y)$ can be expressed in $P_X((y-1)/2)$ as Y = 2X + 1 has a inverse
  - an inverse is needed because X and Y needs to be one to one mapping so that the probability is the same
- $P_Z(z)$ can be expressed in $P_X(x)$, as Z = |X - 2| has no inverse
  - there is not one to one mapping from Z to X

General Case for Function of One RV

If the function is invertible

- $g(\cdot)$ is strictly increasing/decreasing.

  X is a discrete RV with support S_X and PMF $p_X(x)$.

  Y is also a discrete RV with $S_Y = \{g(x) : x \in S_X\}$ and its PMF is

  $$p_Y(y) = p_X(g^{-1}(y)) \text{ for all } y \in S_Y.$$

  In the previous example, $Y = 2X + 1$ is strictly increasing in S_Y.

  Its inverse function is $X = \frac{Y-1}{2}$. As discovered,

  $$p_Y(y) = p_X(g^{-1}(y)) = p_X\left(\frac{y-1}{2}\right) \text{ for all } y \in S_Y.$$

- Must be strictly increase/decreasing or monotonic, so can have inverse

If the function is not invertible

- $g(\cdot)$ is not strictly increasing/decreasing.
  1. Use the table to find the mapping from S_X to S_Y.
  2. Find the PMF by the mapping identified from the above table.

- if no inverse, no one to one, must use a table to list out, must have finite support.
- What if infinite support, or when it's continuous?
  - must divide the function such that each segment is invertible
  - Z = |X - 2| →
    - Z = - X + 2 for -inf < X < 2

- Z = X - 2 for 2 <= X < -inf
- X = 2, can be either

Functions of two RVs

- Functions of two RVs.
  - $Z = h(X, Y)$

  Such as $Z = X + Y, Z = X - Y, Z = XY, Z = X/Y, \ldots$

- PMF of Z can be found from the joint PMF of X and Y

$$P_Z(z) = \sum_{\{(x,y)|f(x,y)=z\}} P_{X,Y}(x, y) \quad for\ Z = f(X, Y)\ and\ z \in S_Z$$

Case of Independent RVs

- Sum of two independent discrete RVs $Z = X + Y$.
  - If two discrete RVs X and Y are independent, for all x and y,

    $$P(X = x, Y = y) = P(X = x)P(Y = y)$$

  - $Z = X + Y$ is also a discrete RV.
  - If the joint PMF or marginal PMFs of X and Y are known, how to find the PMF of $Z = X + Y$?

- If X and Y are independent, PMF of Z can be simplified by having the multiplication of the PMF of X and Y:

$$P_Z(z) = \sum_{\{(x,y)|f(x,y)=z\}} P_X(x)P_Y(y) \quad for\ Z = f(X, Y)\ and\ z \in S_Z$$

Example

Suppose we roll a pair of four-sided fair and independent dice. One of the dice is red and the other is black.

- RV X: the outcome of the red dice and $S_X = \{1,2,3,4\}$.

  RV Y: the outcome of the black dice and $S_Y = \{1,2,3,4\}$.
- X and Y are independent RVs. The possible outcomes are list by a table below. Each outcome occurs with the probability

  $p_{X,Y}(x,y) = p_X(x)p_Y(y) = 1/16$, and

  $P(\text{double}) = P(X = Y) = 1/4$

(X,Y)	1	2	3	4
1	(1,1)	(1,2)	(1,3)	(1,4)
2	(2,1)	(2,2)	(2,3)	(2,4)
3	(3,1)	(3,2)	(3,3)	(3,4)
4	(4,1)	(4,2)	(4,3)	(4,4)

- Let $Z = X + Y$. $S_Z = \{2,3,\ldots,8\}$ identified from the table below.
- $P(Z = 6) = P(\{(2,4),(3,3),(4,2)\}) = 3/16$
- $P(Z < 4) = P(\{(1,1),(2,1),(1,2)\}) = 3/16$
- $P(Z > 9) = P(\emptyset) = 0$
- How to find the PMF of Z ?
  - Find all pairs of (x,y) in the table that $z_0 = x + y$ for all $z_0 \in S_Z$.
  - Each such pair contributes to $p_Z(z_0)$ by $p_{X,Y}(x,y) = p_X(x) \times p_Y(y)$.
  - $p_Z(z_0) = \sum_{\{(x,y)|x+y=z_0\}} p_X(x) \times p_Y(y)$ for a specific z_0.
- In general, $p_Z(z_0) = \sum_{\{(x,y)|z_0=f(x,y)\}} p_{X,Y}(x,y)$ for function $Z = f(X,Y)$.

$X\backslash Y$	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

Z	(X,Y)	$P_Z(z)$
2	(1,1)	$p_X(1)p_Y(1) = 1/16$
3	(1,2), (2,1)	$p_X(1)p_Y(2) + p_X(2)p_Y(1) = 2/16$
4	(1,3), (2,2), (3,1)	$p_X(1)p_Y(3) + p_X(2)p_Y(2) + p_X(3)p_Y(1) = 3/16$
5	(1,4), (2,3), (3,2), (4,1)	$p_X(1)p_Y(4) + p_X(2)p_Y(3) + p_X(3)p_Y(2) + p_X(4)p_Y(1) = 4/16$
6	(2,4), (3,3), (4,2)	$p_X(2)p_Y(4) + p_X(3)p_Y(3) + p_X(4)p_Y(2) = 3/16$
7	(1,6), (6,1)	$p_X(1)p_Y(6) + p_X(6)p_Y(1) = 2/16$
8	(4,4)	$p_X(4)p_Y(4) = 1/16$

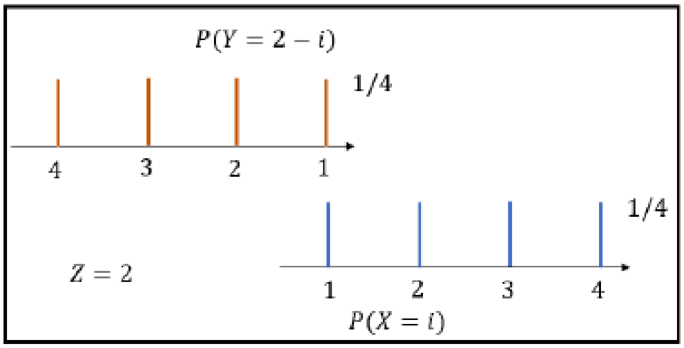
<u>Discrete Convolution Method</u>

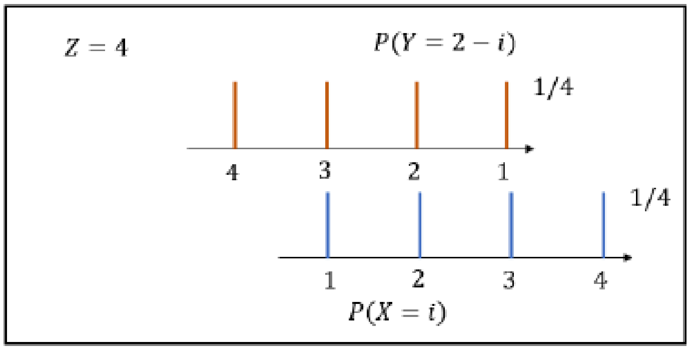
- The operation for cases $Z = 2$ and $Z = 4$ are illustrated in the figures below.
- The PMF of $Z = X + Y$

$$p_Z(j) = \sum_{\substack{k\in S_X \\ j-k\in S_Y}} p_X(k)p_Y(j-k)$$

Discrete convolution.

- When we apply this method,
  - the RVs X and Y must be independent;
  - X and Y can have different PMFs.





- k and j-k to get j
- -k so its reversed direction as k is incremented
- Graphical representation of multiplying to get the same value Z
  - because it's a product of marginal probability, only works if X and Y are independent
- When shifted until no alignment, means prob = 0

Example 2

Let X and Y be independent discrete RVs. Their distributions are

$p_X = \begin{pmatrix} 0 & 1 & 2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$ and $p_Y = \begin{pmatrix} 0 & 1 & 2 \\ 1/8 & 3/8 & 1/2 \end{pmatrix}$. Find the PMF of $Z = X + Y$.

- X and Y are discrete RVs. $Z = X + Y$ is also a discrete RV and its support is $S_Z = \{0, 1, 2, 3, 4\}$.
- The PMF of $Z = X + Y$ can be calculated using the discrete convolution.

$$p_Z(0) = p_X(0)p_Y(0) = \frac{1}{16}$$

Only $X = 0, Y = 0 \rightarrow Z = X + Y = 0$

$$p_Z(1) = p_X(0)p_Y(1) + p_X(1)p_Y(0) = \frac{3}{16}$$
$$p_Z(2) = p_X(0)p_Y(2) + p_X(1)p_Y(1) + p_X(2)p_Y(0) = \frac{1}{4} + \frac{1}{16} = \frac{5}{16}$$
$$p_Z(3) = p_X(1)p_Y(2) + p_X(2)p_Y(1) = \frac{3}{16}$$
$$p_Z(4) = p_X(2)p_Y(2) = \frac{4}{16} = \frac{1}{4}.$$

Expectation of RV Sum

$$E[Z] = E[X] + E[Y]$$

- X and Y need not be independent.
- This rule can be generalized to sum of multiple RVs.

If $E[X_i]$ exists for $i \in \{1,2,\cdots,n\}$,

$$E[X_1 + X_2 \ldots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$ or

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i]$$

These RVs does not have to be independent.

$$Z = X + Y$$
$$E[Z] = E[X + Y] = E[X] + E[Y]$$
$$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} (E[X_i])$$

- no need to be independent
- Since expectation is a linear operation, works for multiple RVs also
- Take advantage of the linearity of expectation

<u>Proof</u>

- Expectation for $Z = g(X, Y)$.

$$E[Z] = \sum_{(x_i, y_j) \in S_{X,Y}} g(x_i, y_j) \times p_{X,Y}(x_i, y_j)$$

➢ $Z = X + Y$

$$E[Z] = E[X + Y] = \sum_{(x_i, y_j) \in S_{X,Y}} (x_i + y_j) \times p_{X,Y}(x_i, y_j)$$

$$= \sum_{(x_i, y_j) \in S_{X,Y}} x_i \times p_{X,Y}(x_i, y_j) + \sum_{(x_i, y_j) \in S_{X,Y}} y_j \times p_{X,Y}(x_i, y_j)$$

$$= \sum_{x_i \in S_X} x_i \sum_{y_j \in S_Y} p_{X,Y}(x_i, y_j) + \sum_{y_j \in S_Y} y_j \sum_{x_i \in S_X} p_{X,Y}(x_i, y_j)$$

$$= \sum_{x_i \in S_X} x_i \times p_X(x_i) + \sum_{y_j \in S_Y} y_j \times p_Y(y_j)$$

$$= E[X] + E[Y]$$

- for a fix x, sum of probability of y ⇒ marginal probability of P(X = x)
  - $$\sum_{y_j \in S_Y} p_{X,Y}(x_i, y_j)$$

Variance of RV Sum

Generalize to multiple RVs.

$$\mathrm{Var}[X_1 + X_2 + \cdots + X_n] = E\left[\sum_{j=1}^{n}\left(X_j - E\left[X_j\right]\right)\sum_{k=1}^{n}(X_k - E[X_k])\right]$$

$$= \sum_{j=1}^{n}\sum_{k=1}^{n} E[\,(X_j - E\left[X_j\right])(\,X_k - E[X_k])\,]$$

$$= \sum_{j=1}^{n}\mathrm{Var}\left[X_j\right] + \sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n}\mathrm{Cov}(X_j, X_k)$$

If $X_1, X_2, \cdots, X_n$ are independent,

$$\mathrm{Var}[X_1 + X_2 + \cdots + X_n] = \sum_{j=1}^{n}\mathrm{Var}\left[X_j\right]$$

$$Z = X + Y$$

$General\ Case$:

$$Var[Z] = Var[X + Y] = Var[X] + 2Cov[X, Y] + Var[Y]$$

$$Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} (Var[X_i]) + \sum_{j=1}^{n} \sum_{k=1, k\neq j}^{n} (Cov[X_j, X_k])$$

$$Var[\sum_{i=1}^{n} X_i] = \sum_{j=1}^{n} \sum_{k=1}^{n} (Cov[X_j, X_k])$$

$RVs\ are\ all\ independent \Rightarrow X\ and\ Y\ are\ uncorrelated \Rightarrow Covariance\ of\ X\ and\ Y\ is\ 0$:

$$Var[Z] = Var[X + Y] = Var[X] + Var[Y]$$

$$Var[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} (Var[X_i])$$

Proof

- Variance of $Z = X + Y$.

$$\begin{aligned}
\mathrm{Var}[Z] &= E[(Z - E[Z])^2] = E[(X + Y - E[X + Y])^2] \\
&= E[(X + Y - E[X] - E[Y])^2] \\
&= E[(X - E[X] + Y - E[Y])^2] \\
&= E[(X - E[X])^2 + 2(X - E[X])(Y - E[Y]) + (Y - E[Y])^2] \\
&= E[(X - E[X])^2] + 2E[(X - E[X])(Y - E[Y])] + E[(Y - E[Y])^2] \\
&= \mathrm{Var}[X] + 2\mathrm{Cov}(X, Y) + \mathrm{Var}[Y]
\end{aligned}$$

  - $\mathrm{Var}[Z] = \mathrm{Var}[X] + \mathrm{Var}[Y]$ only when $\mathrm{Cov}(X, Y) = 0$ (uncorrelated).
  - In general, the variance of Z is not the sum of the individual variances of X and Y!

Sample Average Example

Let $S_n = \sum_{i=1}^{n} X_i$ where X_i are identical and independent distributed (iid) RVs. All of them have expectation μ and variance σ^2. Find the expectation and variance of S_n.

- $E[S_n] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = n\mu$.
- Since X_i 's are independent, they are uncorrelated.

$$\text{Var}[S_n] = \sum_{i=1}^{n} \text{Var}[X_i] + \sum_{i=1}^{n} \sum_{k=1,k\neq i}^{n} \text{Cov}(X_i, X_k) = \sum_{i=1}^{n} \text{Var}[X_i] = n\sigma^2.$$

- Sample average $\frac{S_n}{n}$. $E\left[\frac{S_n}{n}\right] = \frac{n\mu}{n} = \mu$ and $\text{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\text{Var}[S_n] = \frac{\sigma^2}{n}$.

- IID since its from the same distribution (same quantity being measured, same measurement tools)
- covariance is 0 as each sample is independent
  - else will have n terms + n^2 - n terms
  - n for variance of n item
  - n^2 - n for covariance of n * n items - same item (imagine a square with diagonal removed)
- sample average means sum of item/ no of item
- data size big, variance small, means repeated measurement can reduce uncertainty, while the average approaches the actual value

Chapter 5: Continuous Random Variables

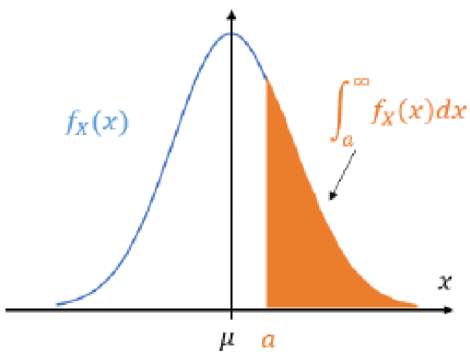
- Continuous sample space.
  - Singapore temperature on 9th of Aug, weight of product, time interval between two events and many, many more.
  - The number of outcomes are uncountable infinite.
- Continuous RVs represent continuous quantities from sample spaces with uncountable infinite outcomes.
  - The support S_X of a continuous RV X is uncountable.
  - $P(X = x_i) = 0$ for any $x_i \in S_X$. → $P\big(x \in (a, b)\big)$ is more meaningful.
  - Calculate probability related to X using probability density function (PDF) or cumulative distribution function (CDF).

- Support of discrete can be finite or countably infinite, and can be easily expressed in table. 1 to 1 mapping with integer
- Support of continuous is uncountable infinite, can map all numbers between 0 - 1 to 1 - inf
    - otherwise, if it's non-zero positive, it will sum to infinity as there is infinite amount of xi in S_X
- Probability of any single point in the support is 0, a range is needed
- There is no PMF, only PDF
    - PMF is for discrete point
    - PDF for continuous
        - something like density * volume get mass of irregular object
- (c) X is a continuous random variable. Can you think of a function $g(x)$ that will make $Y = g(X)$ a discrete random variable?

  Any type of quantization operation can map a continuous RV to a discrete RV. For instance, define a function mapping $g(x) = 1$ if $x > 2$, and $g(x) = -1$ if $x \leq 2$. Then $Y = g(X)$ can take only two values, $\{\pm 1\}$, with $P(Y = -1) = P(X \leq 2) = F_X(2)$ and $P(Y = 1) = 1 - F_X(2)$.

Probability Density Function (PDF)

- Probability density function (PDF)
    - To find $P(X \in (a, b))$, PDF is used.
        - Area under the curve in the range (a, b).



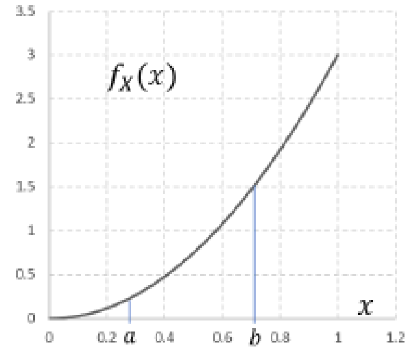
    - The PDF of a continuous RV X with support S is an integrable function $f(x)$ satisfying
        - $f(x) > 0$ for all $x \in S$. (It is possible that $f(x) > 1$.)
        - $\int_S f(x)dx = 1$.
        - For an interval A, $P(X \in A) = \int_A f(x)dx$.

- integral of the PDF in a certain range is the probability of X being in the range
    - area = probability
    - must integrate, the "raw" value of PDF is not probability
    - must integrate to 1, this determines whether a PDF is valid or not
        - validity of PMF depends on the domain and the function itself
        - must be zero outside the domain
        - integral for all x must be 1
    - means PDF must be integrable: finite number of undefined point (like asymptote)
    - all value of $f_X(x) >= 0$ for x in Sx, otherwise can get negative probability (negative area)
    - must know how to integrate for polynomial and exponential for EE2012
    - sketching of PDF can be good for health
- notation is of PDF of X: $f_X(x)$

$$P(x_{min} < X < x_{max}) = \int_{x_{min}}^{x_{max}} f_X(x)\, dx$$

<u>Example 1</u>

The PDF of a continuous RV X is $f(x) = 3x^2$ for $0 < x \leq 1$.



- $f(x) > 0$ when $x \in S = (0,1]$.
- $\int_0^1 f(x)\, dx = 1$.
- But $f(x_i)$ is not a probability.
  For example, $f(1) = 3 > 1$, which is obviously not a probability!
- An event is described by a range in S. The area under the curve in this range gives the probability of the event. For the event $\frac{1}{2} < X \leq 1$, its probability is $P\left(\frac{1}{2} < X \leq 1\right) = \int_{1/2}^{1} f(x)\, dx = \int_{1/2}^{1} 3x^2\, dx = \frac{7}{8}$.
- Note that $P(X = a) = 0$, $P(X = b) = 0$ for any a, b. So for continuous RV X (assuming $a < b$),
  $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$.

- 0 < x <= 1 limits the range of x, not the possible number of x, as it is still infinity
- point has no probability, the range <= and < means the same thing
- P(a range or a point) = P(a range), doesn't matter if the point intersects the range as the intersection is still a point
    - P(A U B) = P(A) + P(B) - P(A n B)

<u>Example 2</u>

Let X be a continuous random variable whose probability density function is $f(x) = x^3/4$ for $0 < x < c$. What is the value of the constant c that makes $f(x)$ a valid probability density function?

- Since the area under $f(x)$ for $0 < x < c$ should be 1,

$$\int_0^c f(x)\, dx = \frac{1}{16} x^4 |_0^c = \frac{1}{16} c^4 = 1 \rightarrow c = 2.$$

- there must be a fixed domain for any PDF as the integral of -inf to + inf of the PDF must be 1
- PDF implicitly states, f for domain, all else 0

Example 3

f = x^3 / C for 0<x<1, find c for PDF to be valid

int f = [x^4/4C] from 0 to 1 = 1
1/4c - 0 = 1
c = 1/4

Example 4

Find the real normalization constant c in the functions below that can be considered as valid PDFs. For those that cannot, explain why.

1. $f_1(x) = ce^{-x}, x \in R$.
2. $f_2(x) = cx(2-x),\ 0 < x \leq 2$.
3. $f_3(x) = 1 - cx,\ 0 < x \leq 1/c$.

- The first function $f_1\ (x)$ cannot be a PDF.

  Its area is infinite for all $c > 0$. It is uniformly zero everywhere if $c = 0$. It is negative everywhere if $c < 0$. Thus, no value of c will make $f_1\ (x)$ a valid PDF.

- f1
  - if C > 0: exponential is always positive, so the integral will be infinite for x in real
  - if C < 0: exponential is always negative, so the integral will be -infinite for x in real
  - if C = 0: exponential is always 0, so the integral will be 0 for x in real
  - no value of c such that f1 can be a PMF
  - what if for x in positive real
    - integral = [-ce^-x] from 0 to +inf = c
    - if c = 1, f1 is valid
- validity of PMF depends on the domain and the function itself

- A quick sketch shows that the function $f_2(x) > 0$ for all $x \in (0,2]$ when $c > 0$, and it has a finite area equal to

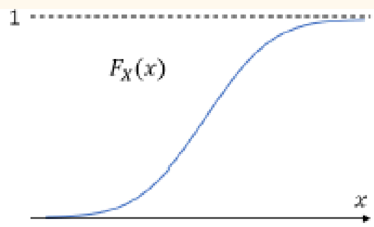
$$A = c\int_0^2 2x - x^2\,dx = c\left[x^2 - \frac{x^3}{3}\right]_0^2 = \frac{4c}{3}$$

  To be a valid PDF, c can be solved by the equation $A = 1$ and $c = \frac{3}{4}$.

- $f_3(x) > 0$ over its support. The region enclosed by $f_3(x)$ and the x-axis in $(0,1/c]$ is a triangle with area $B = \frac{1}{2c}$.

  To be a valid PDF, $B = 1$ and $c = 0.5$.

Cumulative Distribution Function (CDF)



- Cumulative distribution function (CDF).
  - CDF is defined as $F_X(x) \triangleq P(X \leq x)$
  - Discrete RV: CDF is a staircase shape and right continuous. It is clumsy when used for probability calculation.
  - Continuous RV X: $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$ for any real x.
  - CDF is much easier to use for continuous RVs.
  - $F_X(x)$ is continuous and non-decreasing for continuous RV X.

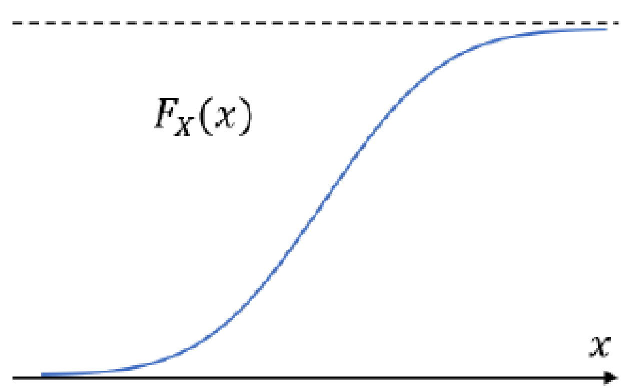
- $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$
  - use t as integration variable to not confused with x, as x is a variable to be plug in later
  - dummy variable
  - like int i in for loop, disappears after the for loop

- If F is the integral of f, F must be differentiable and thus must be smooth
  - unlike the jump for discrete case, here
  - both are non decreasing
- CDF is defined across all x
  - must give answer in the bracket thingy

Properties of CDF

Properties of CDF $F_X(x)$ for continuous RV X.

1. $0 \leq F_X(x) \leq 1, \forall x \in R$.
2. $\lim_{x \to \infty} F_X(x) = 1$.
3. $\lim_{x \to -\infty} F_X(x) = 0$.
4. If $a < b$, $F_X(a) \leq F_X(b)$.
5. Given $a < b$, $P(a < X \leq b) = F_X(b) - F_X(a)$. *



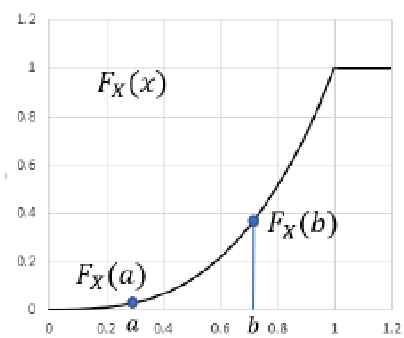
- $a < b$, $F_X(a) <= F_X(b)$, can equal for case where PDF is at 0, CDF is flat
- $P(a <= X < b) = P(a < X <= b) + P(X = a) - P(X = b) = P(a < X <= b)$

Example 1

The PDF of a continuous RV X is $f_X(x) = 3x^2$ for $0 < x \leq 1$. Find $F_X(x)$ and use it to find $P(0.7 \leq X \leq 0.9)$ and $P(0.5 < X \leq 3)$.

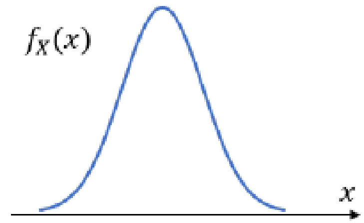
- $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_{0}^{x} 3t^2\,dt = x^3$ when $x \in (0,1]$.
  $F_X(x) = 0$ when $x \leq 0$.
  $F_X(x) = 1$ when $x \geq 1$.

$$F_X(x) = \begin{cases} 0, & x \in (-\infty, 0], \\ x^3, & x \in (0,1], \\ 1, & x \in (1, +\infty). \end{cases}$$

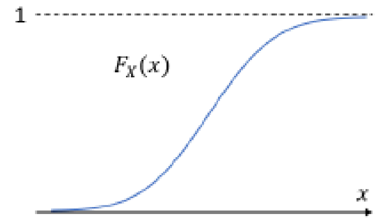


- $P(0.7 \leq X \leq 0.9) = F_X(0.9) - F_X(0.7) = 0.9^3 - 0.7^3 = 0.386$.
- $P(0.5 \leq X \leq 3) = F_X(3) - F_X(0.5) = 1 - 0.5^3 = 0.875$.

PDF and CDF relationship

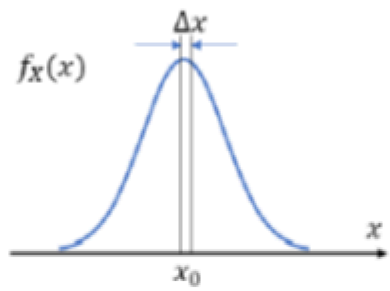


- Both PDF and CDF carry full information of a RV.
  - PDF curve is a convenient visualization tool.
  - CDF is not intuitive due to its "cumulative" nature.



- Relationship between the PDF and CDF of X.

  $f_X(x) = \frac{dF_X(x)}{dx}$ and $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$.
  - $P(X \le x) = F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ is the area under the curve $f_X(x)$ from $-\infty$ to x.
  - $P(a < X \le b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(t)\,dt$.



- CDF values are probability values.
- The values of PDF are not probability values. It is a density like the mass density. It only becomes a probability after integrating over a region of interest.
- $P(X = x_0) = 0$ and $f_X(x_0) \neq P(X = x_0)$.
- $f_X(a) > f_X(b)$ does not imply $P(X = a) > P(X = b)$.
- A larger value of $f_X(x_0)$ means X is more likely in the small neighborhood of x_0. → Small range that X is more likely to be.

- PDF can give visual information by seeing the area underneath a range, to see for a given range
- $f_X(a) > f_X(b)$ does not imply $P(X = a) > P(X = b)$. both equal to zero

- PDF and CDF range does not match is okay, match also ok
    - $$f_X(t) = \frac{1}{4}e^{-t/4}, t > 0.$$

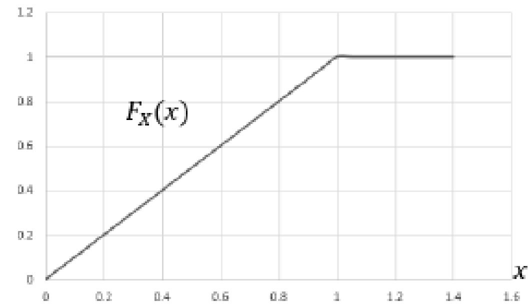
      continuous RV PDF and CDF, we h

      $$\begin{aligned} F_X(x) &= P(X \le x) \\ &= \int_{-\infty}^{x} f_X(t)dt \\ &= \int_{0}^{x} \frac{1}{4}e^{-t/4}dt, \quad \text{for } \boxed{x \ge 0} \\ &= 1 - e^{-x/4}, \quad \text{for } x \ge 0 \end{aligned}$$

Example

The CDF of RV X is given by the expression below.

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1. \\ 1, & x > 1 \end{cases}$$



- From the expression of $F_X(x)$, the attribute of this RV is not obvious.
- If we look at the PDF,

$$f_X(x) = \frac{dF_X(x)}{dx} = \begin{cases} 0, & x \in (-\infty, 0], \\ 1, & x \in (0,1], \\ 0, & x \in (1, +\infty). \end{cases}$$

It is clear that X has a constant PDF value 1 in the range $(0,1]$ and zero elsewhere. It is a continuous uniform distributed RV.

- this shows that having pdf is more informative

Expectation of Continuous RVs

- For a continuous RV X with PDF $f_X(x)$, its expectation is defined as

$$E[X] \triangleq \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

- X is a continuous RV with PDF $f_X(x)$. Expectation for $g(X)$ is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

  - If $g(x) = c$, a constant, $E[c] = \int_{-\infty}^{\infty} c f_X(x) dx = c$.
  - If $g(x) = cx$, $E[cX] = cE[X]$.

- Expectation operator $E[\cdot]$ is linear for discrete and continuous RVs.
  - For a set of n functions $g_i(x)$ and constants c_i, $i = 1, \ldots, n$, we have

$$E[c_1 g_1(X) + \cdots + c_n g_n(X)] = c_1 E[g_1(X)] + \cdots + c_n E[g_n(X)].$$

- when integrating, ignore the zero part of PDF outside the support and integrate the PDF in the support

Integration Tricks

- When the above integration is complicated, some integration tricks can be helpful.
  - If $f_X(x)$ is even symmetry about $x = m$, then $E[X] = m$. ("Even symmetry" means that $f_X(m-a) = f_X(m+a)$ for any $a \in R$.)
  - If X is a non-negative random variable, then $E[X] = \int_0^\infty \left(1 - F_X(x)\right)dx$.

- Even symmetry ⇒ support must be symmetric
  - both (m - a) and (m + a) must be in S_X
  - point of symmetry is E[X]
- Support must be non-negative

Proof for Tricks

1. **If $f_X(x)$ has even symmetry about $x = m$, then $E[X] = m$. "Even symmetry" means that $f_X(m-a) = f_X(m+a)$ for any $a \in R$.**

   It can be proved as follows. Let $f_Y(y) = f_X(m+y)$. Since $f_X(m-a) = f_X(m+a)$, then $f_Y(y) = f_X(m+y) = f_X(m-y) = f_Y(-y)$. So $f_Y(y)$ is an even function. Let $X = m + Y$.

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\infty} (m+y) f_X(x)\,dx \\
&= \int_{-\infty}^{\infty} m f_X(x)\,dx + \int_{-\infty}^{\infty} y f_X(m+y)\,d(m+y) \\
&= m\int_{-\infty}^{\infty} f_X(x)\,dx + \int_{-\infty}^{\infty} y f_X(m+y)\,dy \\
&= m + \int_{-\infty}^{\infty} y f_Y(y)\,dy = m + 0 = m
\end{aligned}$$

   In the above proof, we make use of the following relationships.

   a. $d(m+y) = dy$ because m is a constant.

   b. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$ because $f_X(x)$ is a PDF function.

   c. $\int_{-\infty}^{\infty} y f_Y(y)\,dy = 0$ because $f_Y(y)$ is an even function, y is an odd function.

2. If X is a non-negative random variable, then $E[X] = \int_0^\infty (1 - F_X(x))dx$.

The proof starts from the RHS and applies the integration by parts. Let $u = x$ and $v = 1 - F_X(x)$. Since $f_X(x) = \frac{dF_X(x)}{dx}$, we have $\frac{dv}{dx} = \frac{d(1-F_X(x))}{dx} = 0 - f_X(x) = -f_X(x)$.

$$\int_0^\infty (1 - F_X(x))dx = \int_0^\infty \mathrm{u'v}dx = uv|_0^\infty - \int_0^\infty \mathrm{uv'}dx$$
$$= x(1 - F_X(x))|_0^\infty - \int_0^\infty x(-f_X(x))dx$$
$$= x(1 - F_X(x))|_0^\infty + \int_0^\infty x f_X(x)dx$$

YTTRIUM 10:21 AM

1?

@ Cancel Post

In the equation above, when $x = \infty, x(1 - F_X(x)) = 0$ because $F_X(x)$ is a CDF and $F_X(\infty) = 0$. When $x = 0, x(1 - F_X(x)) = 0$. By the definition of expectation, $\int_0^\infty x f_X(x)dx = E[X]$.

Therefore, the above integration is equal to $0 + \int_0^\infty x f_X(x)dx = E[X]$.

A similar result can be derived for discrete RVs with non-negative values. It is as follows.

If X is a non-negative discrete RV, then $E[X] = \sum_{k=0}^\infty P[X > k]$.

- works for discrete also

Variance of Continuous RVs

- When $g(X) = (X - E[X])^2$, $E[g(X)]$ is the variance of X. That is
$$\mathrm{Var}[X] = E[(X - E[X])^2] = \int_{-\infty}^\infty (X - E[X])^2 f_X(x)dx.$$
- $\mathrm{Var}[X] = E[X^2] - (E[X])^2$.
- Variance operator $\mathrm{Var}[\cdot]$ is not linear for the discrete and the continuous RVs.
  - $\mathrm{Var}[c_1 g_1(X)] = c_1^2 \mathrm{Var}[g_1(X)]$.
  - But $\mathrm{Var}[X + Y] \neq \mathrm{Var}[X] + \mathrm{Var}[Y]$.

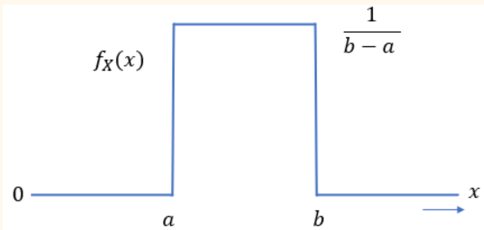
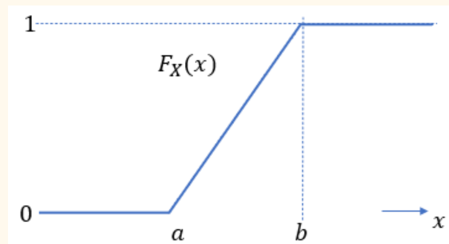
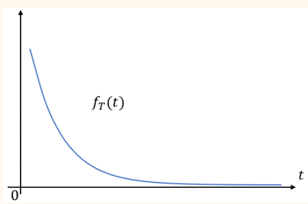
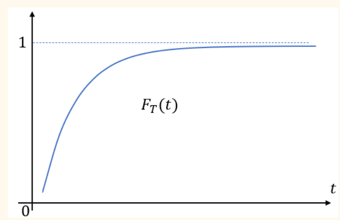
$$Var[X] = E[X^2] - m_X^{\ 2} = \sum_{x \in S_X} x^2 p_X(x) - (\sum_{x \in S_X} x p_X(x))^2 \ for\ discrete$$

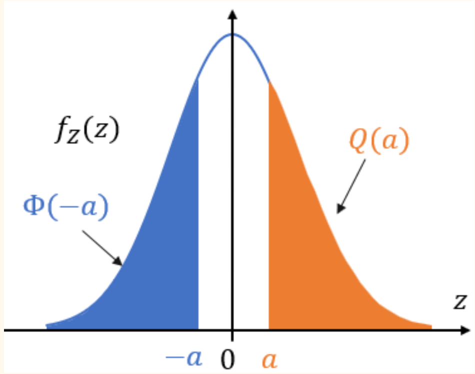
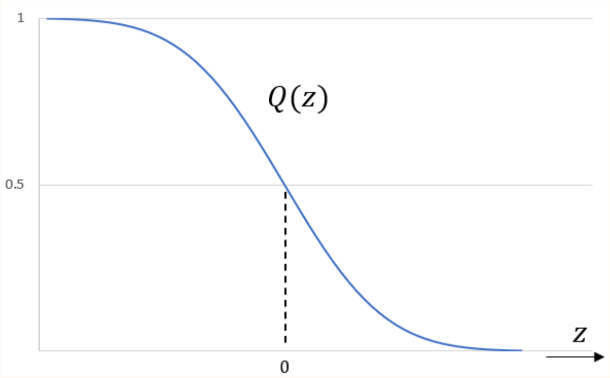
$$Var[X] = E[X^2] - m_X^{\ 2} = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx - (\int_{-\infty}^{\infty} x f_X(x)\, dx)^2 \ for\ continuous$$

$Var[X + Y] = Var[X] + Var[Y]$	X and Y must be independent or uncorrelated, can generalised to multiple RVs

- when integrating, ignore the zero part of PDF outside the support and integrate the PDF in the support

Summary for Common Continuous RVs

	PMF	CDF	E[X]	Var[X]
Continuous Uniform $X \sim U(a,\ b)$	$f_X(x) = \frac{1}{b-a}$, $S_X = (a, b)$ 	$F_X(x) = \int_a^x \frac{1}{b-a} dx = ratio\ of\ triangle\ of\ area\ 1$ $F_X(x) = \int_a^x \frac{1}{b-a} dx = 1 * \frac{x-a}{b-a} = \frac{x-a}{b-a}$ 	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponential $X \sim Exp(\lambda)$ Gives the distribution of interarrival time for X ~ Poison(λt), λ is the arrival rate and t is the time period	$f_T(t) = \lambda e^{-\lambda t}\ for\ t > 0$ 	$F_T(t) = P(T \le t) = \int_0^t \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t}$ 	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gaussian $X \sim N(\mu,\ \sigma^2)$ $X \sim N(E[X],\ Var[X])$	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $S_X = (-\infty, +\infty)$	$P(X < \mu) = 0.5$ $P(X < x) = P(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma})$ $= P(Z < \frac{x-\mu}{\sigma}) = 1 - Q(\frac{x-\mu}{\sigma})$ convert to standard Gaussian and use Q-function table	μ	σ^2
Standard Gaussian $Z \sim N(0,\ 1)$	$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, $S_Z = (-\infty, +\infty)$	$P(Z > a) = Q(a)$ $P(Z < b) = \Phi(b)$ $P(Z > a) = Q(a) = 1 - \Phi(a) = 1 - Q(-a)$ $P(Z < 0) = 0.5$	0	1

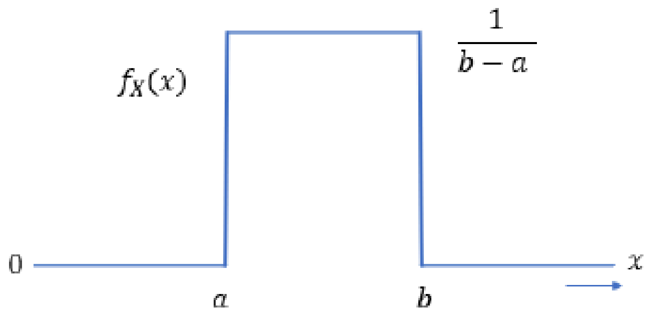
	 $f_Z(z)$ $Q(a)$ $\Phi(-a)$ z $-a$ 0 a	 1 $Q(z)$ 0.5 z 0		

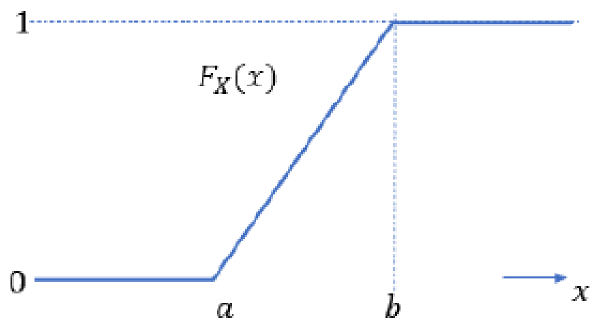
Continuous Uniform RV

(Continuous) uniform RV X in the range (a, b).

$$X \sim U(a, b)$$

- $S_X = (a, b)$ and $f_X(x) = \frac{1}{b-a}$ for $a < X < b$.
- $E[X] = \frac{a+b}{2}$ and $\text{Var}[X] = \frac{(b-a)^2}{12}$
- Picking a number in (a, b) at random, quantization noise…





A continuous uniform RV X has support $S_X = (a, b)$. Its PDF is

$$f_X(x) = \frac{1}{b-a} \text{ for } a < X < b.$$

- Its expectation is

$$E[X] = \int_a^b x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{2} \frac{1}{b-a} (b^2 - a^2) = \frac{b+a}{2}.$$

$E[X]$ is the middle point of S_X for continuous uniform RV.

- same for discrete
- b+a/2 midpoint of the even symmetric PDF

- To find the variance, apply $\text{Var}[X] = E[X^2] - (E[X])^2$.

$$E[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3(b-a)} x^3 |_a^b = \frac{b^3 - a^3}{3(b-a)}.$$

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}$$

- something something used in quantization error
  - noise due to quantization error has variance of Var of continuous uniform RV
  - variance give the noise power, reduce noise but reduce the range b - a
    - ⇒ the more bits used or less range use the higher the accuracy
  - https://www.ni.com/docs/en-US/bundle/ni-daqmx/page/mxcncpts/quanterror.html#:~:text=Quantization%20error%20is%20the%20inherent,errors%2C%20noise%2C%20and%20nonlinearities.

Relationship Between Poisson and Exponential Distribution

1. Introduction

The Poisson distribution is a discrete distribution with probability mass function

$$P(x) = \frac{e^{-\mu}\mu^x}{x!},$$

where $x = 0, 1, 2, \ldots$, the mean of the distribution is denoted by μ, and e is the exponential. The variance of this distribution is also equal to μ.

The exponential distribution is a continuous distribution with probability density function

$$f(t) = \lambda e^{-\lambda t},$$

where $t \geq 0$ and the parameter $\lambda > 0$. The mean and standard deviation of this distribution are both equal to $1/\lambda$.

The cumulative exponential distribution is

$$F(t) = \int_0^\infty \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda t}. \tag{1}$$

2. Relation between the Poisson and exponential distributions

An interesting feature of these two distributions is that, if the Poisson provides an appropriate description of the number of occurrences per interval of time, then the exponential will provide a description of the length of time between occurrences. To understand this, consider that, in a Poisson process, if events occur on average at the rate of λ per unit of time, then there will be on average λt occurrences per t units of time. The Poisson distribution describing this process is therefore $P(x) = e^{-\lambda t}(\lambda t)^x/x!$, from which $P(x = 0) = e^{-\lambda t}$ is the probability of no occurrences in t units of time.

Another interpretation of $P(x = 0) = e^{-\lambda t}$ is that this is the probability that the time, T, to the first occurrence is greater than t, i.e.

$$P(T > t) = P(x = 0 \mid \mu = \lambda t) = e^{-\lambda t}.$$

Conversely, the probability that an event does occur during t units of time is given by

$$P(T \leq t) = 1 - P(x = 0 \mid \mu = \lambda t) = 1 - e^{-\lambda t}.$$

Note that this is the cumulative exponential distribution which, when differentiated with respect to t, produces the probability density function of the exponential distribution $f(t) = \lambda e^{-\lambda t}$.

- Let T be the time between two occurrences, distributed by Poisson, then

$$P(0 < T <= t) = P(T <= t) = 1 - P(T > t) = 1 - e^{-\lambda t}$$

- thus the PDF for T is an exponential distribution in the following form

$$f_T(t) = \lambda e^{-\lambda t} \; for \; t \geq 0$$

- λ is arrival rate in exponential is not the same λ in Poisson which is average arrival in the given time period
  - λ in exponential is α in Poisson
  - If given time period is 1 time unit, magnitude of arrival rate = average arrival

Exponential RV T with parameter λ. $\quad T \sim \mathrm{Exp}(\lambda)$

- $S_T = (0, \infty)$ and $f_T(t) = \lambda e^{-\lambda t}$ for $t \in S_T$.
- $E[T] = \frac{1}{\lambda}$ and $\mathrm{Var}[T] = \frac{1}{\lambda^2}$ (as derived in the previous example).
- T: model the inter-arrival time in the Poisson distribution.
  - Assume a Poisson RV $N \sim \mathrm{Poisson}(\lambda t)$, where λ is the arrival rate (previous α) and t is the time duration considered in N.
  - The probability of no arrival in this period is

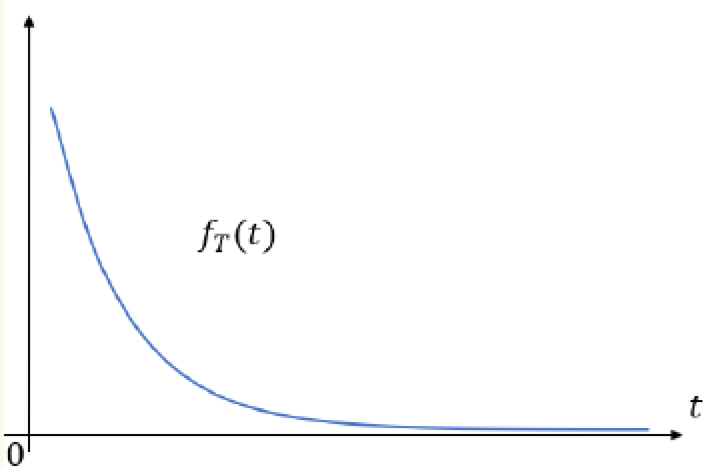
$$P(N = 0) = p_N(0) = \frac{(\lambda t)^0}{0!} e^{-\lambda t} = e^{-\lambda t}.$$

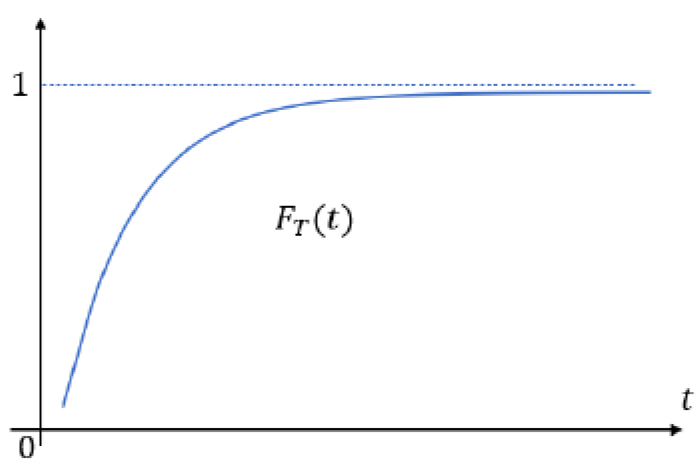
  - Denote the inter-arrival time by continuous RV T.

    $P(T > t) = P(N = 0) = e^{-\lambda t}$.

    CDF of T $\quad F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - e^{-\lambda t}, \; t > 0.$
  - Distribution for the arrival-interval time of two consecutive data packets, calls, buses, clients,...

- model for malfunction interval and inter arrival time





- approach 0 and 1
- exponential decrease in PDF

- malfunction of product has high chance when first bought, so checking is important,
- given that it didn't malfunction in the first t years, the chance of malfunctioning after some t is lower
- so buying second hand is better

The CDF for an exponential RV T has support $S_T = (0, +\infty)$. Its CDF is

$$F_T(t) = \int_0^t \lambda e^{-\lambda x}\, dx = 1 - e^{-\lambda t}, \qquad t > 0.$$

- Using the second trick for expectation, one can find its expectation as

  $E[T] = \int_0^\infty (1 - F_T(t))dt = \int_0^\infty e^{-\lambda t} dt = -\frac{1}{\lambda} e^{-\lambda t}|_0^\infty = \frac{1}{\lambda}.$

- The PDF of T is $f_T(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t}$ for $t > 0$.

- The variance of T is shown in next slides as $\text{Var}[T] = \frac{1}{\lambda^2}$.

- To find the variance of T, we have

  $\text{Var}[T] = E[T^2] - (E[T])^2 = E[T^2] - \frac{1}{\lambda^2}$, where

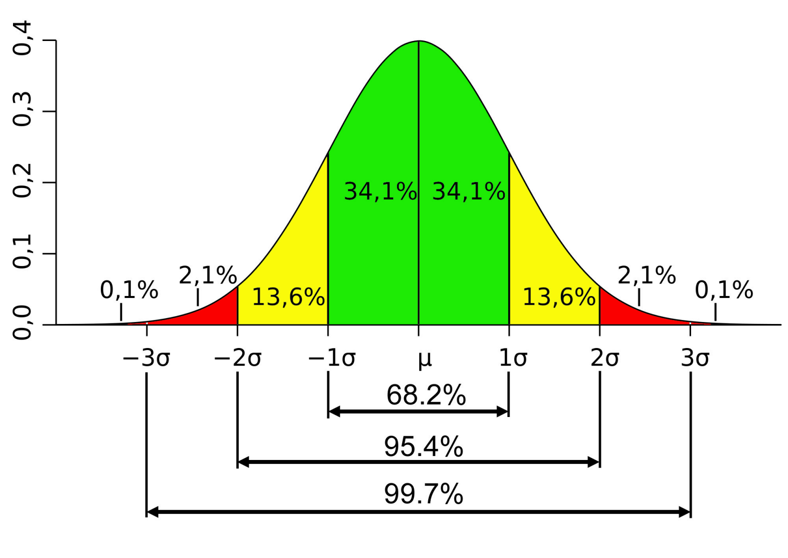
  $E[T^2] = \int_{-\infty}^\infty t^2 f_T(t)dt = \int_0^\infty t^2 \lambda e^{-\lambda t}\, dt$ and integration by parts is needed. Let $u = t^2$ and $v = -e^{-\lambda t}$. And $uv|_0^\infty = -t^2 e^{-\lambda t}|_0^\infty = 0$.

  $E[T^2] = uv|_0^\infty + \int_0^\infty 2te^{-\lambda t}\, dt = \frac{2}{\lambda}\int_0^\infty t\lambda e^{-\lambda t}\, dt = \frac{2}{\lambda} E[T] = \frac{2}{\lambda^2}$, where

  $\int_0^\infty t\lambda e^{-\lambda t} = E[T]$ by definition.

  Therefore, $\text{Var}[T] = E[T^2] - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$.

Gaussian RV



Gaussian (normal) RV X with $E[X] = \mu$, variance $\mathrm{Var}[X] = \sigma^2$.

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

– $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ and $S_X = (-\infty, \infty)$.

– $E[X] = \mu$ and $\mathrm{Var}[X] = \sigma^2$.

– Very popular RV model for many random quantities.

- Noise in communication channel, dynamics in stock market,...
- The distribution for the sum of a large number of random variables converges to a Gaussian distribution. ← Central limit Theorem (CLT)
- Easy to manipulate: $aX + b$ is also Gaussian, PDF is symmetric,...

- sigma ^2 not sigma
- single comms noise may not be gaussian, but adding all signal in a channel gives gaussian distribution for the noise
    - summing multiple independent bernoulli is binomial which is gaussian ish
- additive of same distribution, as bell curve is concentrated at the expected value
- summation of independent term
    - if binomial term alot, ncr overflow, need to transform into gaussian
- linear transformation gaussian is still gaussian
    - measurement has linear transfer function
- NOTE: the variance is taken out of the sqrt to be standard deviation, be careful for sum of independent Gaussian

A Gaussian RV X has support $S_X = (-\infty, +\infty)$. Its PDF is

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

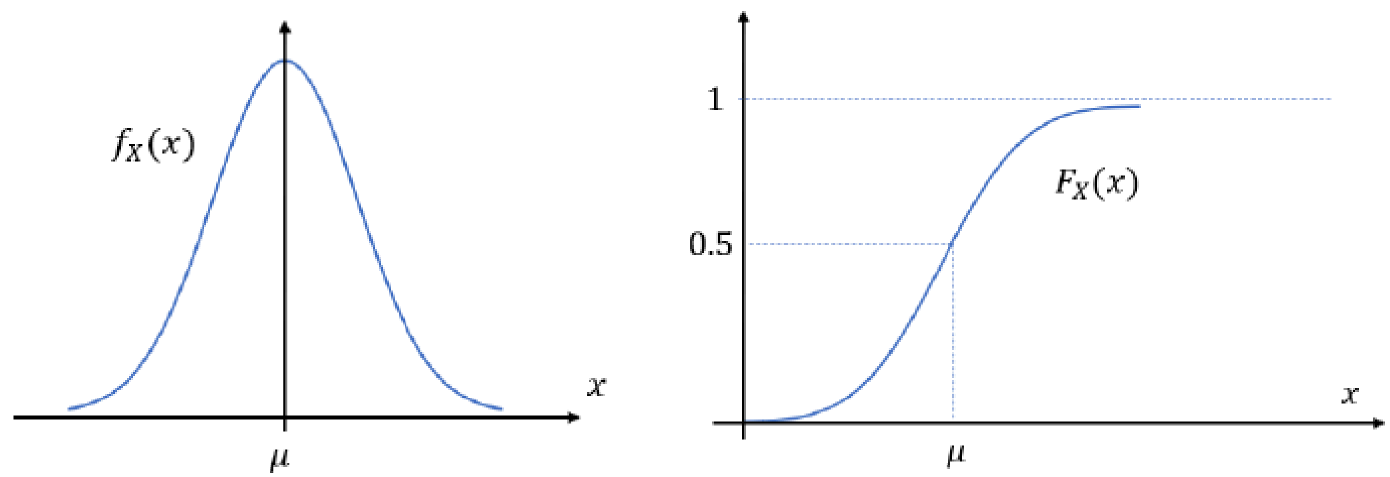
- $f_X(x)$ is even symmetry around $x = \mu$. Using the first trick for expectation, one can find its expectation as $E[X] = \mu$.

- (x - u)^2 shows that the distribution is even symmetric about u, support is also symmetric ⇒ E[X] = u

- Its variance is σ^2, which needs more integration tricks. Refer to the [proof](https://proofwiki.org/wiki/Variance_of_Gaussian_Distribution/Proof_1).
- A Gaussian RV has two parameters: expectation μ and variance σ^2. It is often denoted as $X \sim N(\mu, \sigma^2)$.

- https://statproofbook.github.io/P/norm-mean.html
- https://proofwiki.org/wiki/Variance_of_Gaussian_Distribution/Proof_1

– PDF and CDF for a Gaussian RV.



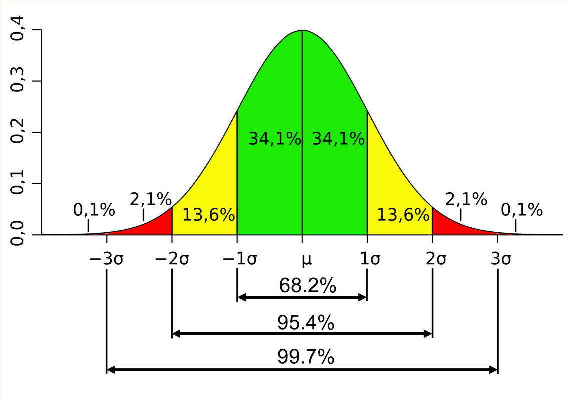
– The probability of event $\{a < X < b\}$

$$P(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$ → No close form!

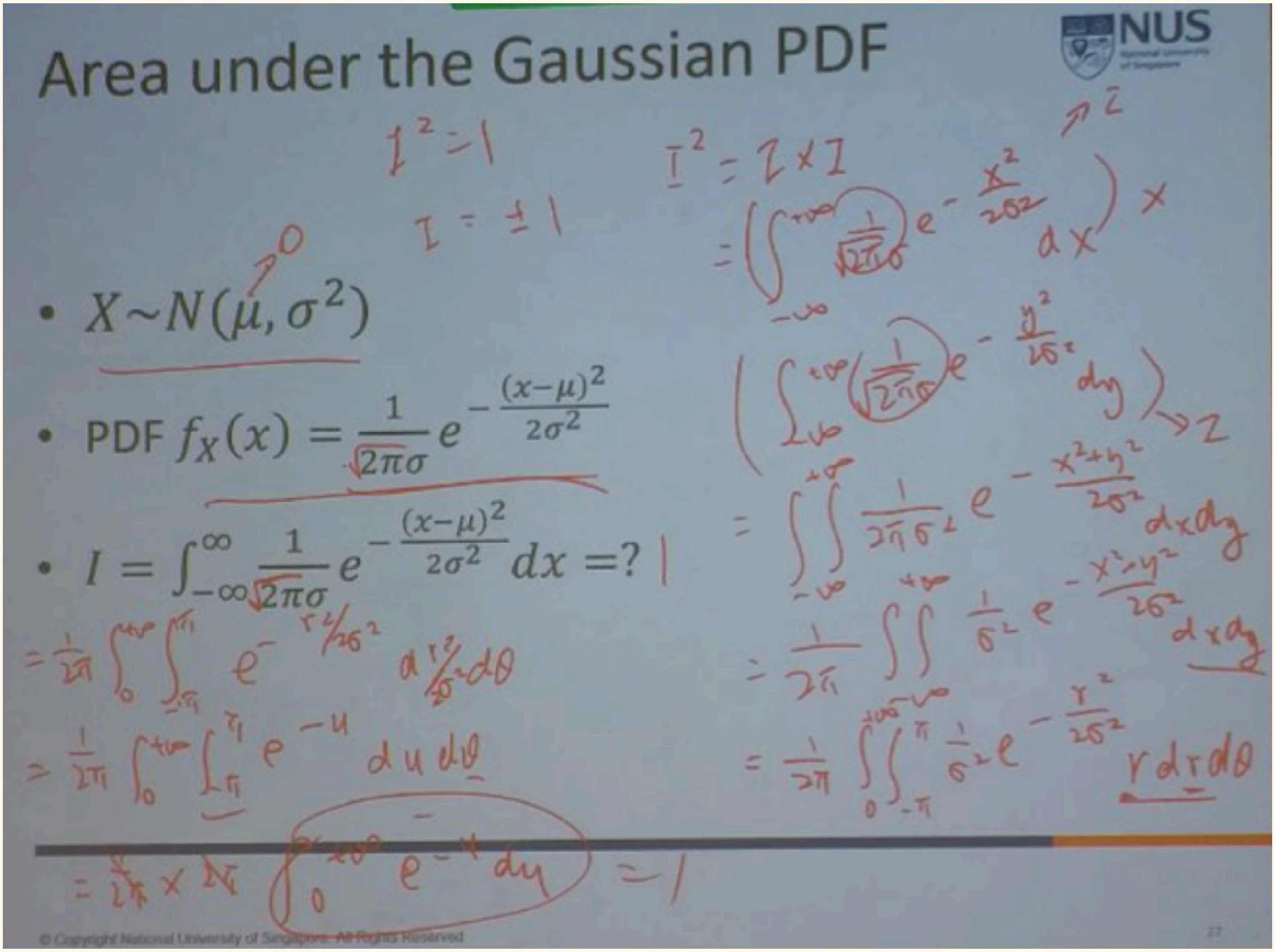
– Apply numerical method every time for different a and b, different μ and σ^2. -- inconvenient!

- no close form
    - if can find the probability for a standard gaussian
        - numerical methods to find common value and make the Q-function table

- if can transform all arbitrary gaussian into standard gaussian
    - for Y = aX + b
    - E[Y] = aE[X] + b
    - Var[Y] = a^2Var[X]
    - to transform arbitrary gaussian X~N(μ, σ^2), to the standard gaussian Z ~ N(0, 1)
    - $P(X < x) = P(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}) = P(Z < \frac{x-\mu}{\sigma}) = 1 - P(Z \geq \frac{x-\mu}{\sigma}) = 1 - Q(\frac{x-\mu}{\sigma})$
- ⇒ can find any gaussian

- This property for integer multiple of standard deviation away can be used to solve probability also
    - 

<u>Proving Area under Gaussian PDF Sums to 1</u>



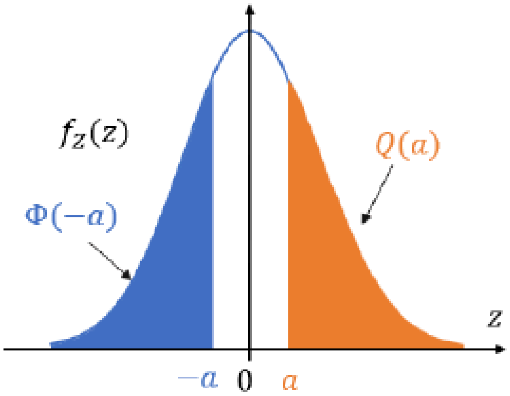
- Find I^2 instead and use change of variable
-

<u>Standard Gaussian</u>

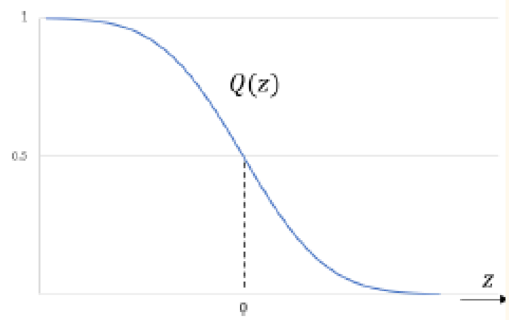
A special Gaussian RV: Standard Gaussian RV $Z \sim N(0,1)$.

- Since $\mu = 0$ and $\sigma^2 = 1$, $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.
- Denote the CDF Z by $\Phi(z) = P(Z \leq z)$
- Q function definition.

$$Q(a) \triangleq P(Z > a) = \int_a^{\infty} f_Z(z)dz$$



- Properties of Q function.
  1) $0 < Q(z) < 1$ for any $z \in R$.
  2) $Q(0) = 0.5$.
  3) $Q(a) > Q(b)$ if $a < b$. (strictly decreasing).
  4) $Q(a) + Q(-a) = 1$.



- special notation for CDF, Q is the complement of phi
- Q(a) = 1 - Φ(a) ⇒ Q(a) + Φ(a) = Q(a) + Q(-a) = 1
- Q(a) = Φ(-a) as it is symmetric
- Area always gets smaller with increase in x, strictly decreasing

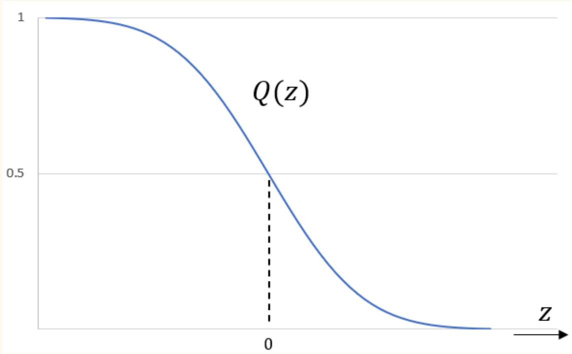
Q-Function Table

E&CE 411, Spring 2009, Table of Q Function

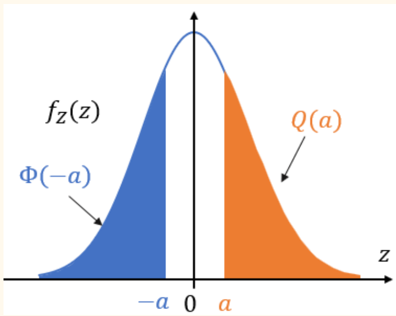
Table 1: Values of $Q(x)$ for $0 \le x \le 9$

x	$Q(x)$	x	$Q(x)$	x	$Q(x)$	x	$Q(x)$
0.00	0.5	2.30	0.010724	4.55	2.6823×10^{-6}	6.80	5.231×10^{-12}
0.05	0.48006	2.35	0.0093867	4.60	2.1125×10^{-6}	6.85	3.6925×10^{-12}
0.10	0.46017	2.40	0.0081975	4.65	1.6597×10^{-6}	6.90	2.6001×10^{-12}
0.15	0.44038	2.45	0.0071428	4.70	1.3008×10^{-6}	6.95	1.8264×10^{-12}
0.20	0.42074	2.50	0.0062097	4.75	1.0171×10^{-6}	7.00	1.2798×10^{-12}
0.25	0.40129	2.55	0.0053861	4.80	7.9333×10^{-7}	7.05	8.9459×10^{-13}
0.30	0.38209	2.60	0.0046612	4.85	6.1731×10^{-7}	7.10	6.2378×10^{-13}
0.35	0.36317	2.65	0.0040246	4.90	4.7918×10^{-7}	7.15	4.3389×10^{-13}
0.40	0.34458	2.70	0.003467	4.95	3.7107×10^{-7}	7.20	3.0106×10^{-13}
0.45	0.32636	2.75	0.0029798	5.00	2.8665×10^{-7}	7.25	2.0839×10^{-13}
0.50	0.30854	2.80	0.0025551	5.05	2.2091×10^{-7}	7.30	1.4388×10^{-13}
0.55	0.29116	2.85	0.002186	5.10	1.6983×10^{-7}	7.35	9.9103×10^{-14}
0.60	0.27425	2.90	0.0018658	5.15	1.3024×10^{-7}	7.40	6.8092×10^{-14}
0.65	0.25785	2.95	0.0015889	5.20	9.9644×10^{-8}	7.45	4.667×10^{-14}
0.70	0.24196	3.00	0.0013499	5.25	7.605×10^{-8}	7.50	3.1909×10^{-14}
0.75	0.22663	3.05	0.0011442	5.30	5.7901×10^{-8}	7.55	2.1763×10^{-14}
0.80	0.21186	3.10	0.0009676	5.35	4.3977×10^{-8}	7.60	1.4807×10^{-14}
0.85	0.19766	3.15	0.00081635	5.40	3.332×10^{-8}	7.65	1.0049×10^{-14}
0.90	0.18406	3.20	0.00068714	5.45	2.5185×10^{-8}	7.70	6.8033×10^{-15}
0.95	0.17106	3.25	0.00057703	5.50	1.899×10^{-8}	7.75	4.5946×10^{-15}
1.00	0.15866	3.30	0.00048342	5.55	1.4283×10^{-8}	7.80	3.0954×10^{-15}
1.05	0.14686	3.35	0.00040406	5.60	1.0718×10^{-8}	7.85	2.0802×10^{-15}
1.10	0.13567	3.40	0.00033693	5.65	8.0224×10^{-9}	7.90	1.3945×10^{-15}
1.15	0.12507	3.45	0.00028029	5.70	5.9904×10^{-9}	7.95	9.3256×10^{-16}
1.20	0.11507	3.50	0.00023263	5.75	4.4622×10^{-9}	8.00	6.221×10^{-16}
1.25	0.10565	3.55	0.00019262	5.80	3.3157×10^{-9}	8.05	4.1397×10^{-16}
1.30	0.0968	3.60	0.00015911	5.85	2.4579×10^{-9}	8.10	2.748×10^{-16}
1.35	0.088508	3.65	0.00013112	5.90	1.8175×10^{-9}	8.15	1.8196×10^{-16}
1.40	0.080757	3.70	0.0001078	5.95	1.3407×10^{-9}	8.20	1.2019×10^{-16}
1.45	0.073529	3.75	8.8417×10^{-5}	6.00	9.8659×10^{-10}	8.25	7.9197×10^{-17}
1.50	0.066807	3.80	7.2348×10^{-5}	6.05	7.2423×10^{-10}	8.30	5.2056×10^{-17}
1.55	0.060571	3.85	5.9059×10^{-5}	6.10	5.3034×10^{-10}	8.35	3.4131×10^{-17}
1.60	0.054799	3.90	4.8096×10^{-5}	6.15	3.8741×10^{-10}	8.40	2.2324×10^{-17}
1.65	0.049471	3.95	3.9076×10^{-5}	6.20	2.8232×10^{-10}	8.45	1.4565×10^{-17}
1.70	0.044565	4.00	3.1671×10^{-5}	6.25	2.0523×10^{-10}	8.50	9.4795×10^{-18}
1.75	0.040059	4.05	2.5609×10^{-5}	6.30	1.4882×10^{-10}	8.55	6.1544×10^{-18}
1.80	0.03593	4.10	2.0658×10^{-5}	6.35	1.0766×10^{-10}	8.60	3.9858×10^{-18}
1.85	0.032157	4.15	1.6624×10^{-5}	6.40	7.7688×10^{-11}	8.65	2.575×10^{-18}
1.90	0.028717	4.20	1.3346×10^{-5}	6.45	5.5925×10^{-11}	8.70	1.6594×10^{-18}
1.95	0.025588	4.25	1.0689×10^{-5}	6.50	4.016×10^{-11}	8.75	1.0668×10^{-18}
2.00	0.02275	4.30	8.5399×10^{-6}	6.55	2.8769×10^{-11}	8.80	6.8408×10^{-19}
2.05	0.020182	4.35	6.8069×10^{-6}	6.60	2.0558×10^{-11}	8.85	4.376×10^{-19}
2.10	0.017864	4.40	5.4125×10^{-6}	6.65	1.4655×10^{-11}	8.90	2.7923×10^{-19}
2.15	0.015778	4.45	4.2935×10^{-6}	6.70	1.0421×10^{-11}	8.95	1.7774×10^{-19}
2.20	0.013903	4.50	3.3977×10^{-6}	6.75	7.3923×10^{-12}	9.00	1.1286×10^{-19}
2.25	0.012224						

- Limited range of x given, but x increase, Q(x) gonna decrease and be damn small
  - only up to 9, 9 standard deviation away, is sufficient
- Limited resolution in x, increment in 0.05
  - just assume linear as the 0.05 is small is ok

- 

- for tut 5, the answer given is by software, thus the value gotten from the table is slightly different

* Only give x >= 0.5, just use the symmetric property to get the other side
  - Q(-a) = 1 - Q(a)
  - 

- $\Phi(z) = 1 - Q(z)$
- The proability $P(a < Z < b)$ can be found from a table of $Q(z)$ or $\Phi(z)$.

  $P(a < Z < b) = P(Z > a) - P(Z > b) = Q(a) - Q(b)$ or

  $P(a < Z < b) = P(Z < b) - P(Z < a) = \Phi(b) - \Phi(a)$
- The Q-function table only list the probabilities for $z \geq 0$.
  - Make use of $Q(-z) = 1 - Q(z)$.
  - No need to tabulate values of $Q(x)$ for $x < 0$.
  - Express the probability related to standard Gaussian RVs by Q-functions with nonnegative argument.
  - Make use of the symmetry of the Standard Gaussian PDF curve.

z	$Q(z)$
0	5.00E-01
0.1	4.60E-01
0.2	4.21E-01
0.3	3.82E-01
0.4	3.45E-01
0.5	3.09E-01
0.6	2.74E-01
0.7	2.42E-01

Converting Arbitrary Gaussian to Standard Gaussian

- Linear transform of $X \sim N(\mu, \sigma^2)$ to $Z \sim N(0,1)$.
  - $Y = aX + b$ is also Gaussian distributed for any $a, b \in R$ and $a \neq 0$.
  - $E[Y] = aE[X] + b = a\mu + b$
  - $\text{Var}[Y] = a^2\sigma^2$
  - Transform X to standard Gaussian RV Z.

    Let $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$. We have $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

- Linear transformation of Gaussian is also Gasussion
- Given the linear transformation, the resulting expectation and variance can be easily gotten from the expectation and variance of the original Gaussian
- Any Gaussian distribution is characterised by their expectation and variance
- ⇒ Any Gaussian can be converted to the standard Gaussian by linear transformation if its expectation and variance is known

  Probabilities for any other Gaussian RV $X \sim N(\mu, \sigma^2)$?
  - Probabilities for standard Gaussian RV Z ? ← Solved!
  - Find the probabilities related to X by transforming X to Z.

  $$P(X > a) = P\left(\frac{X-\mu}{\sigma} > \frac{a-\mu}{\sigma}\right) = P\left(Z > \frac{a-\mu}{\sigma}\right) = Q\left(\frac{a-\mu}{\sigma}\right)$$

  $$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z < \frac{a-\mu}{\sigma}\right) = 1 - Q\left(\frac{a-\mu}{\sigma}\right)$$

  $$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = Q\left(\frac{a-\mu}{\sigma}\right) - Q\left(\frac{b-\mu}{\sigma}\right)$$

- when normalising, make sure is divide by standard deviation not variance

Example

Let $X \sim N(2,4)$. Find the following probabilities:

1. $P(X > 5)$;
2. $P(0 < X \leq 3)$;
3. $P(X \leq 0)$

in terms of the Q function.

- $P(X > 5) = P\left(\frac{X-2}{2} > \frac{5-2}{2}\right) = Q\left(\frac{5-2}{2}\right) = Q(1.5)$.
- To find the second probability, we note that

  $P(0 < X \leq 3) = P(X > 0) - P(X > 3)$, and therefore

  $P(0 < X \leq 3) = Q\left(\frac{0-2}{2}\right) - Q\left(\frac{3-2}{2}\right) = Q(-1) - Q(0.5)$.

  By the property of Q-function, we have $Q(x) = 1 - Q(-x)$. We substitute $Q(-1) = 1 - Q(1)$ and obtain $P(0 < X \leq 3) = 1 - Q(1) - Q(0.5)$.

Example – Gaussian RVs

- $P(X \leq 0) = 1 - P(X > 0) = 1 - Q(-1) = Q(1).$
- From the Q function table, we can find

  $P(X > 5) = Q(1.5) = 6.68 \times 10^{-2},$

  $P(0 < X \leq 3) = 1 - Q(1) - Q(0.5) = 0.532,$

  $P(X \leq 0) = Q(1) = 0.159.$

x	$Q(x)$	x	$Q(x)$
0	5.00E-01	2.7	3.47E-03
0.1	4.60E-01	2.8	2.56E-03
0.2	4.21E-01	2.9	1.87E-03
0.3	3.82E-01	3.0	1.35E-03
0.4	3.45E-01	3.1	9.68E-04
0.5	3.09E-01	3.2	6.87E-04
0.6	2.74E-01	3.3	4.83E-04
0.7	2.42E-01	3.4	3.37E-04
0.8	2.12E-01	3.5	2.33E-04
0.9	1.84E-01	3.6	1.59E-04
1.0	1.59E-01	3.7	1.08E-04
1.1	1.36E-01	3.8	7.24E-05
1.2	1.15E-01	3.9	4.81E-05
1.3	9.68E-02	4.0	3.17E-05
1.4	8.08E-02	4.5	3.40E-06
1.5	6.68E-02	5.0	2.87E-07
1.6	5.48E-02	5.5	1.90E-08
1.7	4.46E-02	6.0	9.87E-10
1.8	3.59E-02	6.5	4.02E-11
1.9	2.87E-02	7.0	1.28E-12
2.0	2.28E-02	7.5	3.19E-14
2.1	1.79E-02	8.0	6.22E-16
2.2	1.39E-02	8.5	9.48E-18
2.3	1.07E-02	9.0	1.13E-19
2.4	8.20E-03	9.5	1.05E-21
2.5	6.21E-03	10.0	7.62E-24
2.6	4.66E-03		

<u>Example 2</u>

Aussume $X \sim N(\mu, \sigma^2)$. Find the probability that $|X - \mu| > n\sigma$ for positive integer n.

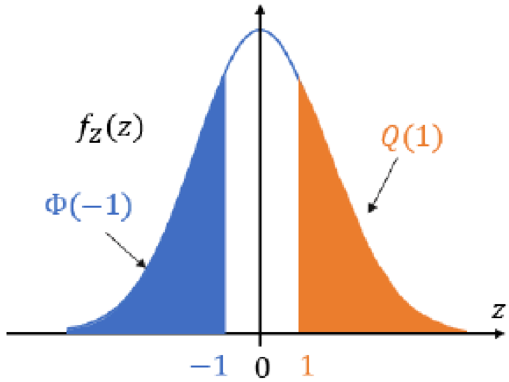
- $|X - \mu| > n\sigma$ means $X - \mu < -n\sigma$ or $X - \mu > n\sigma$. These two events are disjoint.

  $P(|X - \mu| > n\sigma) = P(X - \mu < -n\sigma) + P(X - \mu > n\sigma).$

  $P(X - \mu < -n\sigma) = P\left(\frac{X-\mu}{\sigma} < -n\right) = Q(n)$

  $P(X - \mu > n\sigma) = P\left(\frac{X-\mu}{\sigma} > n\right) = Q(n)$

  $P(|X - \mu| > n\sigma) = 2Q(n)$



- $|X-\mu| > n\sigma$ means X is more than $n\sigma$ away from its expectation. A table below shows the $P(|X-\mu| > n\sigma)$ for different n.

n	1	2	3	4
$2\,Q(n)$	0.318	4.56×10^{-2}	2.70×10^{-3}	6.34×10^{-5}

- It can be observed that the probability of X getting a value more than $n\sigma$ away from its expectation is very low once n exceeds 3.
  - For instance, if we model the average marks of students over a semester as Gaussian with expectation 60 and standard deviation 10, only about 1 in 185 students $(1/0.0027)$ will score higher than $90\ (= \mu + 3\sigma)$ or lower than $30\ (= \mu - 3\sigma)$.

- further away from expected, probability is very small
- within 3 sigma from expected are majority of the population

Sum of Independent Gaussian RVs

- The sum of two independent Gaussian RVs $U \sim N(\mu_1, \sigma_1^2)$ and $V \sim N(\mu_2, \sigma_2^2)$.
  - $H = U + V$ is also Gaussian.
    - $E[H] = \mu_1 + \mu_2$
    - $\mathrm{Var}[H] = \sigma_1^2 + \sigma_2^2$

$$f_H(x) = \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}\sqrt{2\pi}} e^{-\frac{(x-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}},\ S_X = (-\infty, +\infty)$$

- by linearity of expectation
- since independent, covariance is 0, variance is the sum of variance

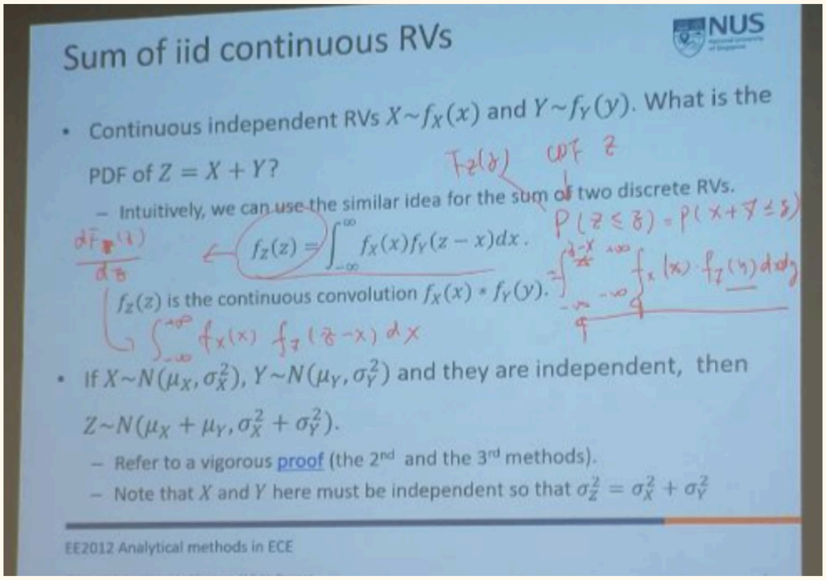
Chapter 6: Multiple Continuous Random Variables

Sum of Continuous Independent RVs

- Continuous independent RVs $X \sim f_X(x)$ and $Y \sim f_Y(y)$. What is the PDF of $Z = X + Y$?
  - Intuitively, we can use the similar idea for the sum of two discrete RVs.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx .$$

$f_Z(z)$ is the continuous convolution $f_X(x) * f_Y(y)$.

- [discrete convolution]
- Fz = P
- 

  - find pdf of Z by finding the CDF of Z first, through double integration to find the volume under the joint PDF of X and Y

Sum of Independent Gaussian RVs

- If $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ and they are independent, then $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.
  - Refer to a vigorous [proof] (the 2nd and the 3rd methods).
  - Note that X and Y here must be independent so that $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$

Sum of Identical and Independent (iid) Gaussian RVs

- When X and Y are identical and independent distributed (iid) Gaussian $N(\mu, \sigma^2)$, we have $Z = X + Y \sim N(2\mu, 2\sigma^2)$.

- Generalize to n iid Gaussian RVs $X_i \sim N(\mu, \sigma^2)$ for $1 \leq i \leq n$.
  - $Z = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$.
  - Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.
    - Increasing the number of measurements can improve the measurement accuracy.
    - Taking more samples in a survey to make the survey results more reliable.

- first measurement result does not affect the second measurement result ⇒ independant
- take n of a RV and divide it by n, the resultant RV is called sample average

Central Limit Theorem (CLT)

Let X_i for $1 \leq i \leq n$ be iid RVs with a finite expectation μ and a finite variance σ^2. And $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The RV

$$Z_n = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

distribution converges to the standard Gaussian RV as $n \to \infty$, that is

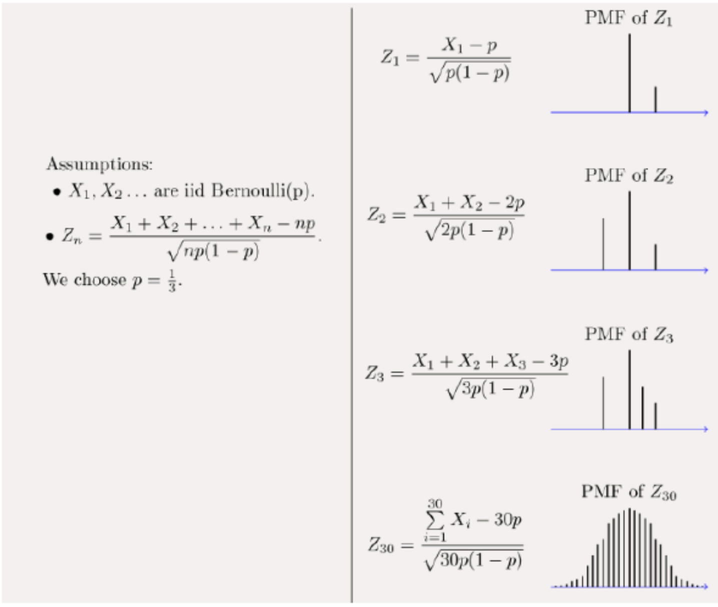
$$\lim_{n \to \infty} P(Z_n \leq z) = \Phi(z) \text{ for all } z \in R.$$

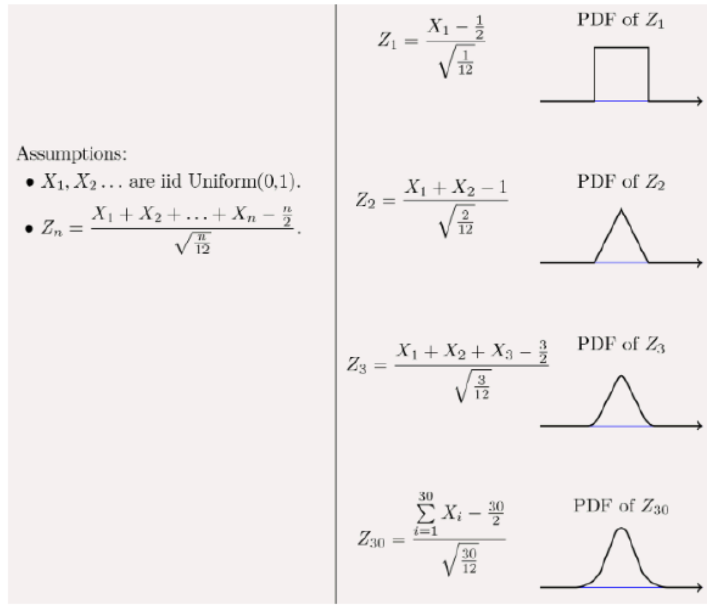
(Recall $\Phi(z)$ is the CDF of the standard Gaussian RV.)

- finite expectation and variance means the expectation and variance must exist, converge to a value and not diverge to infinity
- Xi can be any distribution, discrete or continuous, as long as there its iid and finite expectation and variance
  - also can estimate if just assume independence if an unknown RV is independent or not

In CLT, the iid X_i $(1 \le i \le n)$ can be any distribution. But they should all have finite expectation and variance.

– X_i's can be discrete distributions or continuous distributions.





- If the original variable is discrete, then the envelope of the plot is Gaussian
- convolution of two square pulses, get a triangular pulse

<u>Example</u>

A data packet of 1000 bits is transmitted in a wireless communication. Due to channel noise, each bit is received in error with probability 0.1. The bit errors occur independently. Find the probability of more than 120 errors received in the data packet.

- Let X_i be a Bernoulli RV indicate if i^{th} bit in error (1 – error, 0 – correct). Then $X_i \sim$Bernoulli (0.1). Its expectation $\mu = 0.1$ and variance $\sigma^2 = 0.09$.

- The total number of bits in error is $Y = \sum_{i=1}^{n} X_i$ and $n = 1000$. By CLT,

$$P(Y > 120) = P\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - n\mu}{\sqrt{n}\sigma}\right) = P\left(Z_{1000} > \frac{120-100}{\sqrt{90}}\right)$$

$$\approx Q\left(\frac{20}{\sqrt{90}}\right) = 0.0175.$$

Compare with the direct method

$$P(Y > 120) = \sum_{i=121}^{1000} \binom{1000}{i} (0.1)^i (0.9)^{1000-i}$$

- convert to standard Gaussian to estimate

- Potential Issue with this method
    - accuracy of estimation? n can be big but won't be infinite
    - how to find P(Y=120)?
    - ⇒ Continuity Correction
- without continuity correction, is just whack the same value in no need care about first discrete value

<u>Continuity Correction</u>

Continuity correction for discrete RVs when applying CLT.

– CLT requires $n \to \infty$ or at least n is large enough for a relatively accurate Gaussian estimation.

– The value of n for a good Gaussian estimation depends on the distribution of X_i's. Based on the experience, if $n \geq 30$, the Gaussian estimation is relatively accurate.

– When X_i distribution is discrete, part of the estimation error come from the fact that Y is discrete and it is estimated by a continuous distribution (Gaussian).

- only needed for discrete RVs
- its to account for the case between two discrete number
    - 0 - 1 , 0.5 area is taken in

Example 1

Assume that $Y \sim Bin\ (n = 20, p = 0.5)$. Find $P(8 \leq Y \leq 10)$.

- Apply the steps in the previous example. $X_i \sim \text{Bernoulli}\ (0.5)$ and $\mu = 0.5$, $\sigma^2 = 0.25$ for X_i. Then $Y = \sum_{i=1}^{n} X_i$ for $n = 20$.

$$P(8 \leq Y \leq 10) = P\left(\frac{8-n\mu}{\sqrt{n}\sigma} \leq \frac{Y-n\mu}{\sqrt{n}\sigma} \leq \frac{10-n\mu}{\sqrt{n}\sigma}\right)$$

$$= P\left(\frac{8-10}{\sqrt{5}} \leq Z_{20} \leq \frac{10-10}{\sqrt{5}}\right) \approx Q\left(-\frac{2}{\sqrt{5}}\right) - Q(0) = 0.314$$

- Since $n = 20$ is relatively small, we can find $P(8 \leq Y \leq 10)$ by applying the binomial PMF

$$P(8 \leq Y \leq 10) = \sum_{k=8}^{10} \binom{20}{k} p^k (1-p)^{20-k} = 0.4565$$

The estimation has a quite significant gap to this value.

- Since Y only can take integer values, we can improve the estimation by using the continuity correction, which compute $P(8 \leq Y \leq 10)$ as

$$P(8 \leq Y \leq 10) = P(8 - 0.5 \leq Y \leq 10 + 0.5)$$

$$= P\left(\frac{7.5 - n\mu}{\sqrt{n}\sigma} \leq \frac{Y - n\mu}{\sqrt{n}\sigma} \leq \frac{10.5 - n\mu}{\sqrt{n}\sigma}\right)$$

$$= Q\left(-\frac{2.5}{\sqrt{5}}\right) - Q\left(\frac{0.5}{\sqrt{5}}\right) = 0.4567$$

Compared with the accurate value 0.4565, the continuity correction significantly improves the estimation accuracy. It is working well when X_i is Bernoulli distributed.

- continuity correction is to account for the area under the stand Gaussian between 7-8 and 10-11 for the example of 8<=Y<=10
  - corrected by take 0.5 more
- Y is n bernoulli of p = 0.5, so to use CLT
  - n = 20
  - μ = 0.5

- $\sigma = \sqrt{pq} = \sqrt{0.25}$
- for P(Y=10) then the estimation will be P(9.5 < Y < 10.5)
  - $P(9.5 < Y < 10.5)$
  - $= P(\frac{9.5 - n\mu}{\sigma\sqrt{n}} < \frac{Y - n\mu}{\sigma\sqrt{n}} < \frac{10.5 - n\mu}{\sigma\sqrt{n}})$
  - $= P(\frac{9.5 - 20 * 0.5}{\sqrt{0.25}\sqrt{20}} < Z < \frac{10.5 - 20 * 0.5}{\sqrt{0.25}\sqrt{20}})$
  - $= P(-\frac{\sqrt{5}}{10} < Z < \frac{\sqrt{5}}{10})$
  - $= 1 - 2Q(\frac{\sqrt{5}}{10})$
  - $= 1 - 2Q(0.2236)$
  - $= 1 - 2(0.4116)$ $\quad linear\ interpolated\ between\ 0.20\ and\ 0.25$
  - $= 0.1768$
- compare with Binomial calculation (exact)
  - P(Y=10) = 20C10 (0.5)^20 = 0.1762

<u>Example 2</u>

- ▪ The total number of bits in error is $Y = \sum_{i=1}^{n} X_i$ and $n = 1000$. By CLT,

  $$P(Y > 120) = P\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - n\mu}{\sqrt{n}\sigma}\right) = P\left(Z_{1000} > \frac{120-100}{\sqrt{90}}\right)$$

  $$\approx Q\left(\frac{20}{\sqrt{90}}\right) = 0.0175.$$

  Compare with the direct method

  $P(Y > 120) = \sum_{i=121}^{1000} \binom{1000}{i} (0.1)^i (0.9)^{1000-i}$
- > 120.5 cuz first y is 121
- must be careful of this bullshit, y > 120 ⇒ y >= 121, > 120.5

<u>Summary for Gaussian Approximation</u>

Whatever Distribution	Find Mean and Var	Approximate	Calculate (Note Continuity Correction)
$any\ bullshit,\ X,\ with\ n > 30$	$E[X] = \mu,\ Var[X] = \sigma^2$	$X \sim N(\mu, \sigma^2)$	$P(X > x)$ $\approx P(\frac{X-\mu}{\sigma} > \frac{(x+0.5)-\mu}{\sigma})$ $= P(Z > \frac{(x+0.5)-\mu}{\sigma})$
$X \sim B(n, p)$	$E[X] = np,\ Var[X] = npq$	$X \sim N(np, npq)$	$P(X > x)$ $\approx P(\frac{X-np}{\sqrt{npq}} > \frac{(x+0.5)-np}{\sqrt{npq}})$ $= P(Z > \frac{(x+0.5)-np}{\sqrt{npq}})$
$X_i \sim Bernoulli(p)$ $Y = \sum_{i=0}^{n} X_i$ (sum of iid)	$E[X] = p,\ Var[X] = pq$ $E[Y] = np,\ Var[X] = npq$	$Y \sim N(np, npq)$	$P(Y > y)$ $\approx P(\frac{Y-np}{\sqrt{npq}} > \frac{(y+0.5)-np}{\sqrt{npq}})$ $= P(Z > \frac{(y+0.5)-np}{\sqrt{npq}})$
$X \sim Geo(p)$	$E[X] = \frac{1}{p},\ Var[X] = \frac{1-p}{p^2}$	$X \sim N(\frac{1}{p}, \frac{1-p}{p^2})$	$P(X > x)$ $\approx P(\frac{X-\frac{1}{p}}{\sqrt{\frac{1-p}{p^2}}} > \frac{(x+0.5)-\frac{1}{p}}{\sqrt{\frac{1-p}{p^2}}})$ $= P(Z > \frac{(x+0.5)-\frac{1}{p}}{\sqrt{\frac{1-p}{p^2}}})$

$X \sim Poisson(\lambda)$	$E[X] = \lambda,\ Var[X] = \lambda$	$X \sim N(\lambda, \lambda)$	$P(X > x)$ $\approx P(\frac{X-\lambda}{\sqrt{\lambda}} > \frac{(x+0.5)-\lambda}{\sqrt{\lambda}})$ $= P(Z > \frac{(x+0.5)-\lambda}{\sqrt{\lambda}})$

Normal Approximation (w/ 5 Step-by-Step Examples!)
The Normal Approximation to the Binomial Distribution - The Continuity Correction
The Normal Approximation to the Poisson Distribution - The Continuity Correction
The Normal Approximation to the Geometric Distribution - The Continuity Correction this is wrong

Multiple Continuous RVs

Multiple RVs are characterized by their joint distribution.

- Bivariate continuous RVs X and Y.
  - Joint PDF $f_{X,Y}(x, y)$
  - Joint CDF $F_{X,Y}(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) dv du$.
- To be a joint PDF, $f_{X,Y}(x, y)$ should satisfies
  - $f_{X,Y}(x, y) \ge 0$ for all $(x, y) \in S_{X,Y}$.
  - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy dx = \int_{S_{X,Y}} f_{X,Y}(x, y) dy dx = 1$
  - $P\big((x, y) \in A\big) = \int_{(x,y)\in A} f_{X,Y}(x, y) dy dx$ for an event A.

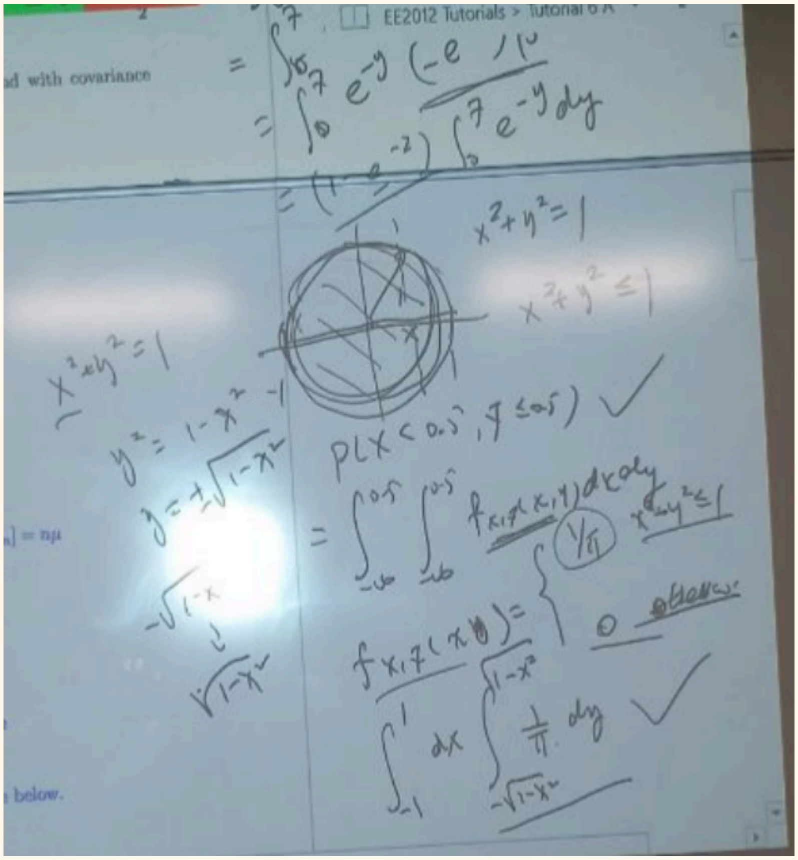
- the PDF is a surface, where the PDF give the height, probability is then the volume under the surface
  - height can be more than 1, but the total volume under the support is 1
- For bivariate, support is an area.

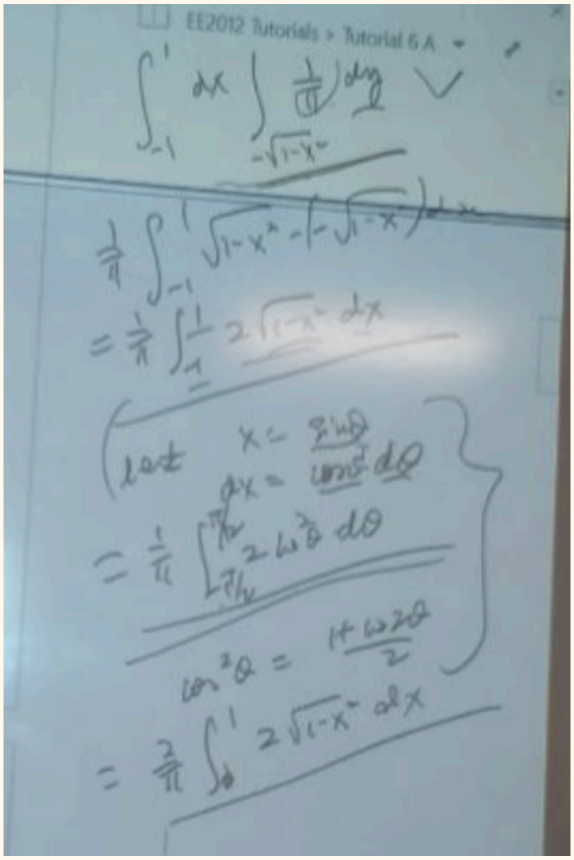
- The marginal PDF can be defined in a similar way as the marginal PMF for discrete RVs.

$$f_X(x) = \int_{y \in S_Y} f_{X,Y}(x, y) dy \text{ and } f_Y(y) = \int_{x \in S_X} f_{X,Y}(x, y) dx$$

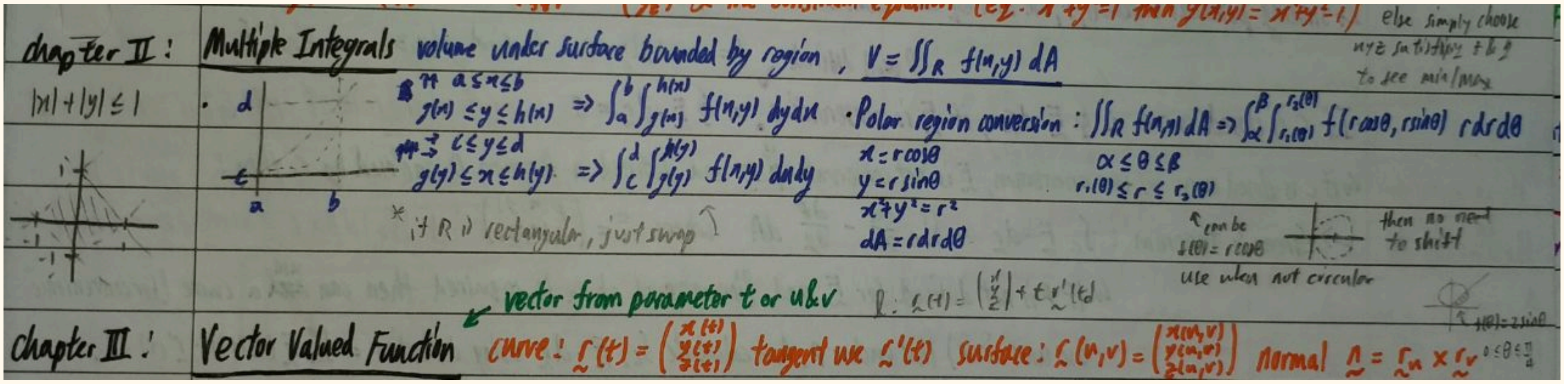
- integrate with respect with Y to get rid of y
- the marginal PDF is used for expectations, variance and covariance

Finding CDF from PDF for non rectangular domain





- even function, integral can cut in half



Expectation

- Expectations and variances of X and Y can be defined based on their marginal PDFs.
- Expectation involving two RVs
  - $E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y) \times f_{X,Y}(x,y)dydx$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dydx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dydx$$

$$= \int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\right)dx + \int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx\right)dy$$

$$= \int_{-\infty}^{\infty} x f_X(x)dx + \int_{-\infty}^{\infty} y f_Y(y)dy = E[X] + E[Y]$$

whatever the relationship between X and Y.

- True no matter the relationship

  - $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \times f_{X,Y}(x,y)dydx \neq E[X] \times E[Y]$ in general.

- This relationship is only true if X and Y are uncorrelated
  - or if its independence since the criteria is uncorrelated
- see discrete case

Covariance and Correlation Coefficient

Covariance and correlation coefficient between X and Y.

- $\text{Cov}[X,Y] \triangleq E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$
- $\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$ where σ_X=$\sqrt{\text{Var}[X]}$ and $\sigma_Y = \sqrt{\text{Var}[Y]}$ are the standard deviations of X and Y.
- $-1 \leq \rho_{X,Y} \leq 1$.
- If $Y = aX + b$ for with a and b being real constants, then $\rho_{X,Y} = 1$ when $a > 0$ and $\rho_{X,Y} = -1$ when $a < 0$.
- If $\rho_{X,Y} = 0$, then X and Y are uncorrelated.
- If X, Y are independent, then $\text{Cov}[X,Y] = 0$ and X, Y are uncorrelated. But the reverse is not true in general.

- for joint gaussian, independent and correlation are the same thing, independent ⇔ uncorrelated, no case where dependent but not uncorrelated

– The necessary and sufficient condition for two continuous RVs X and Y being independent is that their joint PDF is the product of their marginal PDFs. That is

$$f_{X,Y}(x,y) = f_X(x) \times f_Y(y).$$

Similar with the results for discrete RVs.

Example

A and B visit the staff lounge between 10 am and 11 am. Their arrival time is uniformly distributed in this period. Each of them will stay for 10 minutes after they arrive. What is the probability they will see each other?

time of arrival A ~ U(0, 60) , B~ U(0, 60) independent and identical RV

will meet with |A-B| < 10

A B axis with

fA(a) is a flat line, a constant, fB(b) is a flat line, a constant, joint PDF is a flat plane, finding volume to find probability, or area ratio

P(|A-B| < 10) = area for which this is true in the plot

Bivariate Gaussian RVs

Assume ρ is the correlation coefficient of two continuous RVs X and Y. They are joint Gaussian distributed if their joint PDF is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

where μ_X and μ_Y are the expectations of X and Y, σ_X and σ_Y are the standard deviations of X and Y.

– Marginal distribution $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$.

(refer to the supplementary materials for proof.)

- to get $f_X(x)$, integrate $f_{X,Y}(x,y)$ from -inf to +inf with respect to y
  - sum for all y to get marginal distribution of x

Proof

The joint PDF for two joint Gaussian RVs X and Y is

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]},$$

where μ_X and σ_X^2 are the expectation and variance of X, μ_Y and σ_Y^2 are the expectation and variance of Y.

To simplify the derivation, let $u=\frac{x-\mu_X}{\sigma_X}$ and $v=\frac{y-\mu_Y}{\sigma_Y}$. Then the joint PDF above becomes

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}[u^2-2\rho uv+v^2]}.$$

Since $u^2-2\rho uv+v^2=(u-\rho v)^2+(1-\rho^2)v^2$, we have

$$e^{-\frac{1}{2(1-\rho^2)}[u^2-2\rho uv+v^2]}=e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2}\times e^{-\frac{1}{2(1-\rho^2)}\times(1-\rho^2)v^2}=e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2}\times e^{-\frac{1}{2}v^2}.$$

The marginal for Y is

$$f_Y(y)=\int_{-\infty}^{\infty}f_{X,Y}(x,y)\,dx=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2}\times e^{-\frac{1}{2}v^2}\,dx.$$

Since $du=\sigma_X dx$,

$$f_Y(y)=\frac{1}{2\pi\sigma_Y\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2}\times e^{-\frac{1}{2}v^2}\,du$$

$$=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{1}{2}v^2}\times\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2(1-\rho^2)}(u-\rho v)^2}\,du. \qquad (*)$$

Let $h=\frac{u-\rho v}{\sqrt{1-\rho^2}}$. Because the integral in Equ (*) is about u, we have $du=\sqrt{1-\rho^2}dh$. Equ (*) becomes

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{1}{2}v^2}\times\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}h^2}\,dh \qquad (**)$$

In (**), $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}h^2}\,dh=1$ since it is the integral of the standard Gaussian PDF. So the marginal PDF of Y is

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{1}{2}v^2}=\frac{1}{\sqrt{2\pi}\sigma_Y}e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}.$$

This shows the marginal PDF of Y is Gaussian and $Y \sim N(\mu_Y, \sigma_Y^2)$. With the similar argument, we can find the marginal PDF of X is also Gaussian and $X \sim N(\mu_X, \sigma_X^2)$. Therefore, we show the marginal PDFs of joint bivariate Gaussian RVs are Gaussian.

Bivariate Independent Gaussian RVs

If two joint Gaussian RVs X and Y are uncorrelated, they are independent.

When $\rho = 0$,

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \times \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

$$= f_X(x) \times f_Y(y)$$

- all the ρ becomes 0
- can be separated to product of two one-dimensional gaussian ⇒ independent
  - hence uncorrelated ⇒ independent
  - only true for gaussians

When X and Y are both $\sim N(0, \sigma^2)$,

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

- noise is zero mean and with power of sigma squared

Complex Gaussian RVs

Let $i = \sqrt{-1}$ and X, Y be two Gaussian RVs. $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. $H = X + iY$ is a complex Gaussian RV.

– Expectation $E[H] = E[X] + iE[Y] = \mu_X + i\mu_Y = \mu$ (complex)

– Variance

$$\Gamma = E[(H - \mu)(H - \mu)^*]$$

$$= E\left[\left(X - \mu_X + i(Y - \mu_Y)\right)\left(X - \mu_X - i(Y - \mu_Y)\right)\right]$$

$$= E[(X - \mu_X)^2 + (Y - \mu_Y)^2] = \sigma_X^2 + \sigma_Y^2 \triangleq \sigma^2 \qquad \text{(real)}$$

- variance can be compared to quantify the spread, but complex numbers cant be compared. Hence the variance definition has to be modified to give a real variance
    - hence the conjugate, to get the magnitude,
    - variance is real and positive
- Use the fact that ZZ* = |Z|^2 to simplify calculation
- Variance of complex = variance of real + variance of imaginary

Circularly Symmetric Complex Gaussian RVs.

An important family of complex Gaussian RVs are the circularly symmetric complex Gaussian RVs.

- $\mathrm{Re}[H] = X \sim N\left(0, \frac{\sigma^2}{2}\right)$, $\mathrm{Im}[H] = Y \sim N\left(0, \frac{\sigma^2}{2}\right)$ and they are independent.

  $H = X + iY \sim CN(0, \sigma^2)$ → Circularly symmetric complex Gaussian RV

- Models for the complex additive white Gaussian noise and the fading coefficient in wireless transmission without LOS.

- $H = |H| e^{i\angle H} = R\, e^{i\Theta}$, where the magnitude $R = \sqrt{X^2 + Y^2}$ is a continuous nonnegative RV and the phase $\Theta = \tan^{-1}\left(\frac{Y}{X}\right)$ is a continuous RV in $[-\pi, \pi)$.

- both real and imaginary has the same variance and mean of 0
- Uses:
  - noise of independent and whatever distribution will be complex Gaussian when added up, by CLT
  - multipath signal will get intensity and phase delay difference, complex gaussian
    - LOS means line of sight, if has LOS, then modelled by another thing since has dominant signal
    - fading coefficient is for the the intensity and phase delay difference from multipath

- Generalizing, a complex, jointly gaussian random vector $\underline{\mathbf{z}} = \underline{\mathbf{x}} + \mathrm{i}\underline{\mathbf{y}}$ is circularly symmetric when the vector $e^{\mathrm{i}\theta}\underline{\mathbf{z}}$ has the same multivariate probability density function for all θ.

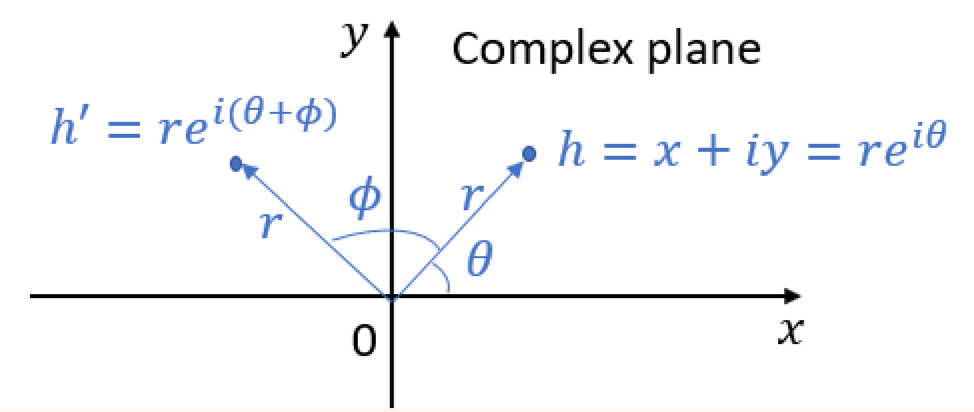
- Such a probability density function must have a zero mean.

- no matter the theta, the distribution is even symmetric at the origin

Meaning of Circularly Symmetric

- Circularly symmetric:

  For a constant $\phi \in [-\pi, \pi)$, $e^{i\phi} H$ has the same distribution as H.



- rotating the joint PDF along the z axis, results in the same distribution
  - same distribution ⇒ same joint PDF
  - ⇒ symmetrical when rotate
  - ⇒ circularly symmetric

The PDF $f_H(r,\theta)$ of H, where $r \geq 0$ and $\theta \in [-\pi, \pi)$.

- Let $A = \{(x,y): x \leq x_0, y \leq y_0\}$. Event A includes all points on the subarea that are to the left and lower side of the point (x_0, y_0) on the complex plane.

- X and Y are iid and uncorrelated. Their joint PDF is

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

- $P(A) = \int_{(x,y)\in A} f_{X,Y}(x,y)\, dxdy = \int_A f_H(r,\theta)\, drd\theta \quad (*)$
- The Cartesian coordinate to the polar coordinate: $dxdy = r\, drd\theta$.
- Comparing the two sides of $(*)$ and since $r^2 = x^2 + y^2$, we have

$$f_H(r,\theta) = f_{X,Y}(x,y) \times r = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}.$$

- note the dxdy = r drdθ, r is called the Jacobian
  - can be explained from area of rectangle and arc length
  - or from x=rcos theta and somehow 2 by 2 matrix det is r
- integration in x y plane is hard, but easy in polar coordinates
- can equate because the integral of (x, y) in A is equivalent to the integral of (r, theta) in A
  - same surface
  - same portion integrated
  - just different coordinates system

PDF $f_H(r, \theta + \phi)$ for $e^{i\phi}H = e^{i\phi}|H|e^{i\angle H} = Re^{i(\Theta+\phi)}$

- In $f_H(r, \theta)$, the phase variable θ does not appear at all.

$$f_H(r, \theta + \phi) = f_H(r, \theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

- $e^{i\phi}H$ has the same distribution as the distribution of H, which is the circularly symmetric property of $H \sim CN(0, \sigma^2)$.

- joint pdf does not depend on θ,
  - ⇒ independent on θ
  - ⇒ distribution is the same regardless of θ
  - hence circularly symmetric
- when rotate is the same

<u>Marginal Distribution of Random Variable R and Θ in Polar Version of Circularly Symmetric Complex Gaussian</u>

Find the marginal distributions of $f_H(r, \theta)$

- $f_R(r) = \int_{-\pi}^{\pi} f_H(r, \theta) d\theta = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \times \int_{-\pi}^{\pi} d\theta$

$$= \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \times 2\pi = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \leftarrow \text{Rayleigh distribution PDF}$$

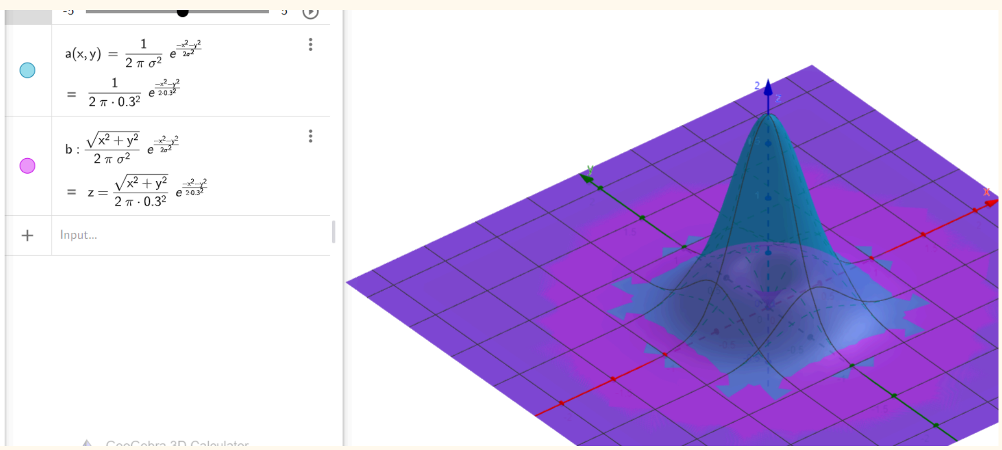
The magnitude R (of the complex Gaussian RV H) has the Rayleigh distribution. Its support S_R is all nonnegative real numbers.

- PDF of R, is the PDF of the magnitude, or distance from origin
- R is nonnegative, hence, the PDF is valid
- magnitude of circularly symmetric complex gaussian follows the Rayleigh distribution

$$f_\Theta(\theta) = \int_0^\infty f_H(r,\theta)dr = \int_0^\infty \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr = \int_0^\infty \frac{1}{2\pi} e^{-\frac{r^2}{2\sigma^2}} d\left(\frac{r^2}{2\sigma^2}\right) = \frac{1}{2\pi}.$$

The phase Θ of the complex Gaussian RV H is a continuous uniform distributed RV with $S_\Theta = [-\pi, \pi)$.

- phase of circularly symmetric complex gaussian follows a continuous uniform distribution
- make sense as the distribution is the same when rotated, shape is the same in all direction (angle)

- [ ] 

- [ ] at r = 0, the pdf fr is 0, how can if its suppose to be a peak
- [ ] https://www.geogebra.org/3d/bnakymut

Tutorial

Tutorial 1

Q1. In a group of 10 students, 6 students play basketball, 7 students play soccer ball and 1 student plays neither.

(a) Find the probability that a randomly selected student plays both of the sports.

(b) If two students A and B are selected from the group in order, what is the probability that A is a basketball player and B is a soccer player?

n = 10

nB = 6
nS = 7
n(B n S) = 4

b)
P(A = basketball n B = soccer)
= P(B = soccer n A = also soccer) + P(B = soccer n A = !also soccer)
= P(B = soccer | A = also soccer)P(A = also soccer) + P(B = soccer | A = !also soccer)P(A = !also soccer)
= 4/10 * 6/9 + 2/10 * 7/9
= 19/45 #

Q3. Assume a fair six-sided dice is rolled.

(a) If the dice is rolled 10 times, find the probability that seven shows at least once.

(b) Find the probability that one shows twice, three shows twice, and six shows once if the dice is rolled 5 times.

(c) Let event A = "the number is less than 3". If the dice is rolled an infinite number of times, what is the probability that A occurs k times at the n^{th} roll but not earlier?

P(A) = 2/6 = 1/3
A occurs k times at nth roll ⇒ kth times at nth roll
P = P(k-1 times in n-1 rolls) P(A)
= (n-1)C(k-1)P(A)$^{k-1}$(1-P(A))$^{n-1-(k-1)}$P(A)
= (n-1)C(k-1)P(A)^{k}(1-P(A))$^{n-k}$
= (n-1)C(k-1)(1/3)^{k}(2/3)$^{n-k}$ #

Q5. Each of the five players competing in a quiz is asked 10 quick-fire questions. The score of each player depends on the number of correct answers, and each player gets all scores with equal probability. What is the probability that

(a) all players get a different score?

scores 0 - 10, 11 possibility

method 1
Find all possible combination of 5 different scores, permutation is taken care off by this method when doing 11*10…
P(all diff) = 11*10*9*8*7/11^5 = 5040/14641 #

method 2
from 11 possible marks choose 5 then permute
P(all diff) = 11C5 * 5! / 11^5 = 5040/14641 #

(b) exactly 2 of them get the same score?

method 1
Find all possible combination of 4 different scores * combination of players where 2 has the same score.

Scores	M1	M2	M3	M4	
Players	A	B	C	D	E

P(exactly 2 same score) = 11*10*9*8*5C2 / 11^5 = 7200/14641

- from 5 choose two to have same score, the this will take care of possible pairs of same scores
- Multiply by 11 possibilities
- Multiply by 10 9 8 possibilities

method 2
Pick 2 out of 5 players to have 1 of 11 possible scores, 3 players left can get 3 different scores out of the remaining 10 and get permuted. 5C2 takes care of permutation, so no need 4!, it's 3!
P(exactly 2 same score) = 5C2*11C1*10C3*3! / 11^5= 7200/14641

(c) exactly 3 of them get the same score?

method 1
Find all possible combination of 3 different scores * combination of players where 3 has the same score.

Scores	M1	M2	M3		
Players	A	B	C	D	E

P(exactly 2 same score) = 11*10*9*5C3 / 11^5 = 900/14641

method 2
Pick 3 out of 5 players to have 1 of 11 possible scores, 2 players left can get 2 different scores out of the remaining 10 and get permuted. 5C3 takes care of permutation, so no need 3!, it's 2!
P(exactly 2 same score) = 5C3*11C1*10C2*2! / 11^5= 900/14641

Q6. One rolls two six-sided fair dices n times.

(a) Find the probability of two numbers being the same in one roll.

(b) Find the probabilities of their sum being 5 and 7, respectively.

(c) Find the probability sum 5 happens before sum 7.

P(sum = 5) = 4/36 = 1/9
P(sum = 7) = 6/36 = 1/6
P(sum != 5, 7) = 1 -1/9 -1/6 = 13/18

Sum 5 before sum 7 ⇒ 5 appeared in a series of trials where 7 never appeared, sum 5 is before sum 7
Let En = no sum 5 or 7 for n-1 times and sum 5 for nth time
P(En) = $(13/18)^{n-1}$1/9

E = sum 5 before sum 7
P(E) = sum P(En) from n = 1 to infinity
= 1/9 * (1/(1-13/18))
= 2/5 #

Q8. Suppose in an ECE graduation party there are 23 students. What is the probability that at least two students have the same birthday? That is if you randomly ask any two students. Assume that there are 365 days in a year and that all days are equally likely to be the birthday of a given student.

Simplified case of 3 students
Given 2 students, the total possible combination is 365*365, and probability of same = 365/365^2 = 1/365

The chance of choosing 2 student from 3 is
A B C
A B
B C
A C
3C2 combination
⅓ chance

P(same bd randomly ask 2 students) = ⅓ (1/365) + ⅓ (1/365) + ⅓ (1/365) = 1/365

Thus, "randomly ask 2 students" can be ignored as all students has equal chance of being chosen

P(at least 2 have same bd)
= 1 - P(all diff)
= 1 - (365*364*...*(365-23+1)/365^23)
= 1 - (365!/((365-23)! * 365^23))
= 1 - (365P23) / 365^23
= 0.507 #

In general

. That is, the probability of a different birthday is 363/365. For n students, we have

$$P(\text{no students share the same birthday}) = \frac{365!}{(365-n)!365^n}.$$

Tutorial 2

Q6. Symbols 0 and 1 are transmitted in a noisy channel with probability p and $1-p$, respectively. Due to the channel noise, they are received correctly with probability u and v, respectively. Assume the transmissions for any two symbols are independent.

(a) What is the probability that the k^{th} symbol is received correctly?

(b) What is the probability that the sequence 1101 is received correctly?

(c) To reduce the error probability, each symbol is transmitted three times and the receiver applies a majority vote to decode it. What is the probability that 0 is correctly decoded? If 101 is received, what is the probability this symbol is 0?

Tutorial 5

Q3. An average of 50 pedestrians pass a given point on a sidewalk per hour in the day, and the average drops to 10 in the night. Assuming that the number of pedestrians per hour is Poisson, find the probability density function of the time between successive pedestrians passing this point.

The number of pedestrians per hour follows Poisson distribution. The time between successive pedestrians follows the exponential distribution in day and night based on the relation between the Poisson distribution and the exponential distribution. Let T be the time interval between successive pedestrians in hours. In the day, T is exponentially distributed with $\lambda = 50$ per hour. In the night, T is exponentially distributed with λ= 10 per hour. Their conditional PDFs can be represented as follows.

$$f_T(t|\text{“day”}) = 50e^{-50t}, \quad t > 0,$$
$$f_T(t|\text{“night”}) = 10e^{-10t}, \quad t > 0.$$

With equal chance, a randomly inter-pedestrian interval can picked in the day or in the night. Therefore we have the overall PDF as

$$\begin{aligned} f_T(t) =& 0.5 f_T(t|\text{“day”}) + 0.5 f_T(t|\text{“night”}) \\ =& 25e^{-50t} + 5e^{-10t}, t > 0. \end{aligned}$$

- See relationship between poisson and exponential
- conditional probability can be used in CDF and will give the same final PDF after differentiation

Q5. Two chips are being considered for use in a certain system. The lifetime of chip 1 is modelled by a Gaussian RV with expectation 20,000 hours and standard deviation 5,000 hours. The lifetime of chip 2 is also Gaussian, with expectation 22,000 hours and standard deviation 1,000 hours. Which chip is preferred if the target lifetime is 20,000 hours? What if the target is 24,000 hours?

Denote the lifetime of Chip 1 and Chip 2 by the random variables X_1 and X_2 respectively. We have $X_1 \sim \mathcal{N}(20000, 5000^2)$ and $X_2 \sim \mathcal{N}(22000, 1000^2)$. Their PDFs is illustrated in Figure 4.

If the target lifetime of the chip is 20,000 hours, we need to compare $P(X_1 > 20000)$ against

stable life time means variance small

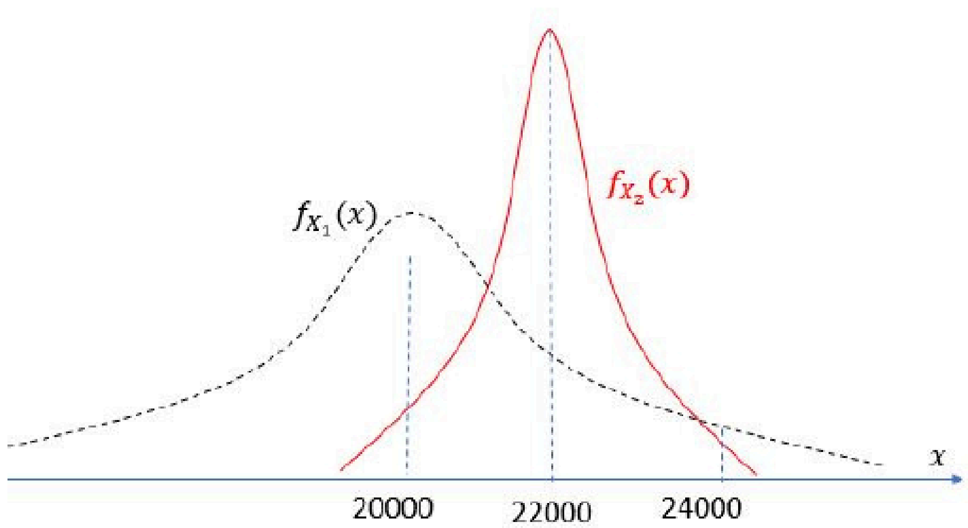


FIG. 4. PDF of $f_{X_1}(x)$ and $f_{X_2}(x)$.

$P(X_2 > 20000)$ and pick the one with the higher probability.

$$P(X_1 > 20000) = Q\left(\frac{20000 - 20000}{5000}\right) = Q(0) = 0.5$$
$$P(X_2 > 20000) = Q\left(\frac{20000 - 22000}{1000}\right) = 1 - Q(2) = 0.9772.$$

Therefore Chip 2 is preferred.

If the target lifetime is 24,000 hours, we need to compare following two probabilities and choose the chip with relatively higher probability.

$$P(X_1 > 24000) = Q\left(\frac{24000 - 20000}{5000}\right) = Q\left(\frac{4000}{5000}\right) = Q(0.8) = 0.212$$
$$P(X_2 > 24000) = Q(\frac{24000 - 22000}{1000}) = Q\left(\frac{2000}{1000}\right) = Q(2) = 0.0228.$$

Thus Chip 1 is preferred this time, which is surprising. The reason is Chip 1's lifetime has a much larger standard deviation than Chip 2's. Although on average Chip 2 lasts longer than Chip 1, it is more likely for Chip 1 to last more than 4000 hours beyond its average than for Chip 2 to last more than 2000 hours beyond its average. Note that both probabilities in this case are quite small. The relatively larger one in Figure 4 should be selected.

- distribution can be compared this way to get the preferred distribution base on the probability

Q6. It is found that the 3G network is available with probability 0.8; independently, the 4G network is available with probability 0.8. A smartphone first tries to connect to the 4G network, then if that is not available, it connects to the 3G network. If neither network is available then there is no connectivity. Let X be some performance measure of the network, with the following conditional distributions:

$$f_X(x|\text{“connected to 4G”}) = 0.1[u(x-10) - u(x-20)]$$
$$f_X(x|\text{“connected to 3G”}) = 0.2[u(x-7) - u(x-12)]$$
$$f_X(x|\text{“neither”}) = \delta(x).$$

Find the PDF of X.

- two unit steps can be used to create a rectangular pulse, once normalised to area = 1 can be used for modelling uniform RV

Tutorial 6

Q2. Let $X_1, X_2, ..., X_n$ be random variables with the same expectation μ and with covariance function

$$\mathrm{Cov}[X_i, X_j] = \begin{cases} \sigma^2, & i = j \\ \rho\sigma^2, & |i-j| = 1 \\ 0, & \text{otherwise} \end{cases}$$

where $|\rho| < 1$, find the expectation and variance of $S_n = X_1 + X_2 + \cdots + X_n$.

- from the first line, infers that all Xi have the same variance
- make sense if i is time, for example voice amplitude cannot change fast, so the one before and after must be correlated somehow

The following two results are known to us after last question.

(a) $\mathrm{Var}[X_i] = \mathrm{Cov}[X_i, X_i] = \sigma^2$, for all i;

(b) $\mathrm{E}[X_i] = \mu$, for all i.

Therefore,

$$\mathrm{E}[S_n] = E[X_1 + X_2 + \cdots + X_n] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n] = n\mu$$

$$\mathrm{Var}[S_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Cov}[X_i, X_j]$$

$$= n\sigma^2 + \sum_{i=1}^{n} \sum_{|i-j|=1}^{n} \mathrm{Cov}[X_i, X_j] = n\sigma^2 + 2(n-1)\rho\sigma^2$$

- just whack general formula for covariance taking note that for cov(Xi, Xj) = cov(Xj, Xi) so need double count

Therefore,

$$\mathrm{E}[S_n] = E[X_1 + X_2 + \cdots + X_n] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n] = n\mu$$

$$\mathrm{Var}[S_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Cov}[X_i, X_j]$$

$$= n\sigma^2 + \sum_{i=1}^{n} \sum_{|i-j|=1}^{n} \mathrm{Cov}[X_i, X_j] = n\sigma^2 + 2(n-1)\rho\sigma^2$$

To see this derivation, we use the case of $n = 4$ as an example in the table below.

$\mathrm{Cov}[X_i, X_j]$	X_1	X_2	X_3	X_4
X_1	σ^2	$\rho\sigma^2$	0	0
X_2	$\rho\sigma^2$	σ^2	$\rho\sigma^2$	0
X_3	0	$\rho\sigma^2$	σ^2	$\rho\sigma^2$
X_4	0	0	$\rho\sigma^2$	σ^2

As you can see, in this table, each column, except for the first and the last column, has three nonzero elements, which have a sum of $\sigma^2 + 2\rho\sigma^2$. The first and the last column only has two nonzero elements with a sum of $\sigma^2 + \rho\sigma^2$. So $\mathrm{E}[S_4] = 4\sigma^2 + 8\rho\sigma^2 - 2\rho\sigma^2 = 4\sigma^2 + 6\rho\sigma^2$.

- draw out to avoid mistake

Q3. The lifetime of the batteries produced by a company is $X \sim N(\mu = 150, \sigma^2 = 16)$. Their quality control department randomly tests 9 battery samples from the market for their lifetime. Assume all the testings are independent. independent sample from populatopn

(a) Denote the lifetime of these 9 samples by X_i for $1 \le i \le 9$. What is the distribution of the average lifetime from the samples?

Denote the sample average by $S = \frac{1}{9}\sum_{i=1}^{9} X_i$. sum of xi does not mean 9*x,

Since $X_i \sim N(150, 16)$, we have

$$\mathrm{E}[S] = \mathrm{E}\left[\frac{1}{9}\sum_{i=1}^{9} X_i\right] = \frac{1}{9}\sum_{i=1}^{9} \mathrm{E}[X_i] = 150.$$

Since X_i are also independent, we have

$$\mathrm{Var}[S] = \mathrm{Var}\left[\frac{1}{9}\sum_{i=1}^{9} X_i\right] = \left(\frac{1}{9}\right)^2 \sum_{i=1}^{9} \mathrm{Var}[X_i] = \frac{16 \times 9}{81} = \frac{16}{9}.$$

X_i's are independent Gaussian RVs, their average $S \sim N(150, 16/9)$. And it has a much smaller variance due to multiple sample average.

- Sum of iid not the same as sum of a single RV

(b) To improve its battery lifetime, the company applies a new manufacturing technology. The quality department randomly tests the lifetime of 9 new batteries and finds out their average life time is 155 hours. If the battery lifetime is unchanged, what is the probability for getting a lifetime sample average greater than or equal to 155? Based on the result, do you think the new technology helps improve the battery lifetime?

If the battery lifetime is unchanged, the sample average lifetime is $S \sim N(150, 16/9)$ as derived in the last question. Then

$$P(S \geq 155) = P\left(\frac{S-150}{\sqrt{\frac{16}{9}}} \geq \frac{155-150}{\sqrt{\frac{16}{9}}}\right) = P(Z \geq 3.75) = 0.8842 \times 10^{-4}.$$

This is a small probability. It implies the event "a sample average of 155" unlikely happens if the battery lifetime is unchanged. This testing result suggests the new technology probably helps improve the battery lifetime.

probability for getting a lifetime sample average greater than or equal to 155? means witht he old distrubtion whats th eporb

observation and disturbution doesnt match emans distrubtuion is differnet

hypthesis tesiting requres prob threshold

must be independent samples

Q4. Tom uses a coin in a game. But he does not know if it is fair. To test it, he flips it 100 times and observes the number of heads. Then he makes a decision by the rule that if the number of head observed is greater than 55, the coin is biased.

(a) If the coin is fair, what is the probability estimation that he makes an incorrect decision?

If the coin is fair, he will make an incorrect decision when the number of heads is greater than 55 based on the decision rule. Denote such event probability by P_e. We already discussed in previous chapters that the number of heads in 100 flips of a fair coin is binomial distributed with $n = 100$ and $p = 0.5$. But we need to estimate P_e because there are 45 terms to add if we use binomial distribution.

Let $X_i \sim \text{Bernoulli}(0.5)$ for $1 \leq i \leq 100$ and the number of heads in $n = 100$ flips is $Y = \sum_{i=1}^{n} X_i$. We have $\text{E}[X_i] = 0.5$ and the variance $\sigma^2 = 0.25$ for all X_i. By the central limit theorem, $Z = \frac{Y-0.5n}{\sqrt{n}\sigma} = \frac{Y-50}{5}$ can be estimated as a standard Gaussian RV. Therefore, the probability P_e can be estimated as

$$P_e = P(Y > 55) \approx P\left(Z > \frac{55-50}{5}\right) = P(Z > 1) = Q(1) = 0.1587,$$

where $Q(\cdot)$ is the Q-function defined in our lecture.

disicion error is P(Y>55)

decision error is the probability of conclusion being wrong, Tom will be wrong if the coin is fair and P(Y>55)

Q5. In a communication via a noisy channel, the signal detected by the receiver is modelled as $R = e^{-j\theta}M + H$. In this model, M is the transmitted message. The constant phase angle θ is due to the channel disturbance and it is perfectly estimated by the receiver. The channel additive noise H is a circularly symmetric complex Gaussian random variable with zero expectation and variance σ^2. Since θ is known to the receiver, the receiver can compensate the received signal as $R' = e^{j\theta}R = M + e^{j\theta}H$. Let $H' = e^{j\theta}H$.

(a) Find the expectation and variance of H'.

Let $H = X + jY$ with both X and Y being real RVs. Since H is a circularly symmetric complex Gaussian RV, we have $X \sim N(0, \sigma^2/2)$, $Y \sim N(0, \sigma^2/2)$ and they are also independent.

By Euler formula $e^{j\alpha} = \cos(\alpha) + j\sin(\alpha)$, we have

$$H' = (\cos(\theta) + j\sin(\theta)) \times (X + jY) = [X\cos(\theta) - Y\sin(\theta)] + j[X\sin(\theta) + Y\cos(\theta)].$$

Because both X and Y have zero expectation, we have

$$\begin{aligned}
\mathrm{E}[H'] &= \mathrm{E}[X\cos(\theta) - Y\sin(\theta)] + j\mathrm{E}[X\sin(\theta) + Y\cos(\theta)] \\
&= \cos(\theta)\mathrm{E}[X] - \sin(\theta)E[Y] + j\{\sin(\theta)\mathrm{E}[X] + \cos(\theta)\mathrm{E}[Y]\}\} \\
&= 0.
\end{aligned}$$

Since $\mathrm{E}[H'] = 0$, the variance of H' is

$$\begin{aligned}
\mathrm{Var}[H'] &= \mathrm{E}[H' \times (H')^*] \\
&= \mathrm{E}[(X\cos(\theta) - Y\sin(\theta))^2 + (X\sin(\theta) + Y\cos(\theta))^2] \\
&= \mathrm{E}[X^2 + Y^2] = \mathrm{E}[X^2] + \mathrm{E}[Y^2] = \sigma^2/2 + \sigma^2/2 = \sigma^2.
\end{aligned}$$

- Precalculate H' to simplify calculation
- use the fact that ZZ* = |Z|^2 to simplify calculation

(b) Show the real and imaginary parts of H' are independent.

From (a), the real part of H' is $H'_r = X\cos(\theta) - Y\sin(\theta)$, and the imaginary part of H' is $H'_i = X\sin(\theta) + Y\cos(\theta)$. Since X and Y are independent Gaussian RVs and θ is a constant, both H'_r and H'_i are real Gaussian RVs. If they are uncorrelated, they will be independent. From (a), we also have $\mathrm{E}[H'_r] = \mathrm{E}[H'_i] = 0$. The covariance of H'_r and H'_i is calculated as follows.

$$\begin{aligned}
\mathrm{Cov}[H'_r, H'_i] &= \mathrm{E}[(H'_r - 0)(H'_i - 0)] \\
&= \mathrm{E}[(X\cos(\theta) - Y\sin(\theta))(X\sin(\theta) + Y\cos(\theta))] \\
&= \mathrm{E}[X^2\cos(\theta)\sin(\theta) - Y^2\cos(\theta)\sin(\theta)] \\
&= \ cos(\theta)\sin(\theta) \times \{\mathrm{E}[X^2] - \mathrm{E}[Y^2]\} \\
&= \cos(\theta)\sin(\theta) \times \{\sigma^2/2 - \sigma^2/2\} = 0
\end{aligned}$$

So H'_r and H'_i are uncorrelated. Since they are both Gaussian, they are independent. Combining the results in (a) and (b), we have $H' \sim CN(0, \sigma^2)$. It has the same distribution as H. This is exactly the property of the cyclic symmetric Gaussian RVs.

- there is no special covariance formula for complex Gaussian like variance as the component of complex gaussian are real and the covariance between them have no difference with the covariance of any other RVs

Midterm

If P(H) = 0.8, and the coin is flipped continuously. What is the probability that 5 heads appear before 3 tails

(e) Let A = "5 heads appear before 3 tails". Note that A or $\bar{A}$ ("3 tails appear before 5 heads") will occur within the first 7 flips.

If there are at least 5 heads in the first 7 flips, the number of tails will be less than 3. → A

If there are less than 5 heads in the first 7 flips, the number of tails will be at least 3. → $\bar{A}$

Let $p = 0.8$.

$P(A) = \sum_{i=5}^{7}\binom{7}{i}p^i(1-p)^{7-i} = \binom{7}{5}p^5(1-p)^2 + \binom{7}{6}p^6(1-p) + \binom{7}{7}p^7 = 0.8520.$

- if n = 8 = 3 + 5, cannot determine which comes first, so must look at n - 1 = 7

$+P(X=7)$

Q.2 (10 marks)

$= {}^7C_5\, 0.8^5\, 0.2^2 + {}^7C_6\, 0.8^6\, 0.2 + {}^7C_7\, 0.8^7\, 0.2^0 = 0.8520$ #

In a random experiment, a coin is picked at random from 3 coins and the picked coin is flipped 5 times. The head probability of the first coin is 1/2, the head probability of the second coin is 2/3, and the third coin has head on both sides. Define events A = " 3 heads are observed", B = " 2 heads are observed in the first two flips" and $C = A \cap B$.

(a) Find $P(B|A)$. (2 marks)

(b) Find $P(A)$. (4 marks)

(c) Find $P(\text{"the first coin was picked"}\,|\,C)$. (4 marks)

b) $P(A) = \frac{1}{3}\left[{}^5C_3 (1/2)^3 (1/2)^2 + {}^5C_3 (2/3)^3 (1/3)^2 + 0\right] = \frac{2495}{11664} \approx 0.2139$ #

a) $P(B|A) = \frac{P(B\cap A)}{P(A)}$ Let $p = P(H)$ for the coin chosen

$= \frac{p^2\, {}^3C_1\, p(1-p)^2}{{}^5C_3\, p^3 (1-p)^2}$ HHXXX ∩ 3H2T

$= \frac{{}^3C_1}{{}^5C_3} = 0.3$ # ↑ independent on the coin chosen

c) $P(C) = P(A\cap B) = P(HH\,XXX \cap 3H2T)$

$= \frac{1}{3}\left[(\tfrac{1}{2})^2\, {}^3C_1 (\tfrac{1}{2})(\tfrac{1}{2})^2 + (\tfrac{2}{3})^2\, {}^3C_1 (\tfrac{2}{3})(\tfrac{1}{3})^2 + 0\right]$

$= \frac{499}{7776}$ OR $P(C) = P(B|A)P(A) = 0.3 \times \frac{2495}{11664} = \frac{499}{7776}$

$P(\text{1st coin}\,|\,C) = \frac{P(\text{1st coin} \cap C)}{P(C)} = \frac{\frac{1}{3}(\frac{1}{2})^2\, {}^3C_1 (\frac{1}{2})(\frac{1}{2})^2}{499/7776}$

$= 243/499 = 0.4870$ #

- Can see that P(B|A) is independent of the probability of head, and thus the coin chosen. Hence the 2 marks lmao.
  - can use counting method also

Quiz 1

Question 1 — 1 / 1 pts

Let S={1,2,3,4}, P({1,2,4})=0.7 and P({1,3,4})=0.8. What is P({1,4})?

Correct! ◉ 0.5

○ 0.3

○ -0.3

○ 0.2

Let S={1,2,3,4}, P({1,2,4})=0.7 and P({1,3,4})=0.8. What is P({1,4})?
disjoint => p(3) = 1-0.7 =0.3, P(2) = 1 - 0.8 = 0.2
p(1,4) = 1 - P(2,3) = 1- 0.2-0.3 = 0.5#

Question 2

1 / 1 pts

Flipping a fair coin 4 times, which a number of heads appears with the highest chance?

- 3
- 1
- Correct! 2
- 4

0	1	2	3	4
1/16	4/16	6/16	4/16	1/16

2#

also is E[X] = np

Question 3

1 / 1 pts

The product deficiency rate is 0.1 in an assembling line. What is the probability that the at least one defective product in out of 10 products?

- 0.3487
- Correct! 0.6513
- 0.3036
- 0.1

P(at least 1 in 10 defect) = 1 - P(all good) = 1 - 0.9^10 = 0.6513#

Question 4

1 / 1 pts

Urn A contains 3 red and 3 black balls, whereas urn B contains 4 red and 6 black balls. If a ball is randomly selected from each urn, what is the probability that the balls will be the same colour?

- ○ 0.6
- Correct! ◉ 0.5
- ○ 0.2
- ○ 0.3

P(same colour) = P(R, R) + P(B, B) = 0.5*4/10 + 0.5*6/10= 0.5#

Question 5

1 / 1 pts

5 one-dollar coins will be distributed to 4 kids A, B, C and D randomly. All distribution schemes have the same probability. What is the chance A is distributed exactly 2 dollars?

- ○ 0.25
- ○ 0.02637
- Correct! ◉ 0.1786
- ○ 0.1429

k= 5 votes, n= 4 candidates
total ways = (5+4-1)C(3) = 8C3 = 56

A	B	C	D
2			

k= 3 votes, n= 3 candidates
total ways = (3+3-1)C(3) = 5C3 = 10

P(A=2) = 10/56 = 0.1786

Quiz 2

Question 1

1 / 1 pts

Two coins A and B are flipped， and the results are recorded. Find the probabilities below.

1. The probability that both coins land heads given the coin A lands heads.
2. The probability that both coins land heads given at least one coin lands heads.

- ○ 1/2, 1/4
- ○ 1/4, 1/3
- ○ 1/4, 1/4
- Correct! ◉ 1/2 , 1/3

Question 2

1 / 1 pts

A coin with head probability 0.4 is flipped 5 times. Define event A = " 3 heads in 5 flips" and event B = "first 2 flips are heads". What is the probability P(A) and P(B|A).

- Correct! ◉ 0.2304, 0.3
- ○ 0.3125, 0.5
- ○ 0.3125, 0.3
- ○ 0.2304, 0.5

Question 3

1 / 1 pts

In a chess tournament, a player A may meet 3 types of opponents.

P(Type 1)=0.5 , P(A win| Type 1)=0.3

P(Type 2)=0.25 , P(A win| Type 2)=0.4

P(Type 3)=0.25 , P(A win| Type 3)=0.5

Which type of opponent did player A meet most probably if the player won?

- Type 2
- Type 3
- Correct! Type 1

Question 4

1 / 1 pts

There are two types of coin in a bag. Type A coin amount is half of the type B amount. When flipped, type A coin gives head with probability 3/4 and type B gives head with probability 4/5. A coin is picked from the bag randomly and flipped twice. The outcome sequence (Head, Tail) is observed. What is the probability coin type A is picked?

- 0.6098
- 0.1067
- Correct! 0.3695
- 0.3191

Question 5

1 / 1 pts

Assume P(A) =0.3, P(B)=0.4. If they are independent, what is $P(AB^c)$? If they are disjoint, what is $P(AB^c)$?

- 0.12, 0.18
- 0.12, 0.4
- Correct! 0.18, 0.3
- 0.18, 0.6

Quiz 3

Question 1

1 / 1 pts

A fair coin is tossed n times. If the probability that head occurs 6 times is equal to the probability that head occurs 8 times, the value of n is

- Correct! 14
- 24
- 16
- 48

Question 2

1 / 1 pts

There are 2 two-dollar coins and 3 one-dollar coins in bag. Two coins are picked at random from a bag. Let X be random variable representing the picked total amount. What is E[X]?

Correct!

- (•) $2.8
- ( ) $1.5
- ( ) $1.4
- ( ) $2.5

Question 3

1 / 1 pts

If a fair coin is flipped, what is the probability that more than 10 flips are needed to observe the first head?

Correct!

- (•) 1/1024
- ( ) 1/512
- ( ) 10/1024
- ( ) 10/512

Question 4

1 / 1 pts

At least how many flips of a fair coin is needed so that with probability at least 0.9 that one or more heads are observed?

Correct!

- (selected) 4
- 6
- 5
- 3

Question 5

1 / 1 pts

Let the support of the random variable X is {-1,0,1} and P{X = 0} = 0.5, P{X = 1} = 0.3. What is $E[X^2]$?

Correct!

- (selected) 0.5
- 0
- 1
- 2

Quiz 4

Question 1

1 / 1 pts

Let X be the number of heads and Y be the number of tails in 5 flips of a fair coin. What is the joint distribution P(X=2, Y=3) ?

- 5/8
- Correct! 5/16
- 1/2
- 1/8

X and Y are correlated with X + Y = 5, correlation coefficient = -1
P(X = 2, Y = 3) = P(X = 2) = P(Y = 3) = 5C3 (0.5)^5 = 5/16 #

P(X = 2, Y = 3) = P(X = 2 n Y = 3) = P(X = 2 | Y = 3) P(Y = 3) = 1 * P(Y = 3) = 5C3 (0.5)^5 = 5/16 #

X\Y	0	1	2	3	4	5
0	0	0	0	0	0	P(Y=5) = P(X=1)
1	0	0	0	0	P(Y=4) = P(X=1)	0
2	0	0	0	P(Y=3) = P(X=2)	0	0
3	0	0	P(Y=2) = P(X=3)	0	0	0
4	0	P(Y=1) = P(X=4)	0	0	0	0
5	P(Y=0) = P(X=5)	0	0	0	0	0

Question 2

1 / 1 pts

A fair dice is rolled 3 times, and the number is observed in each roll. X= "times of 1 observed" and Y= "times the observed number is less than 4". Find the joint probability P(X=1,Y=2).

- 0.1302
- 0.0625
- 0.0289
- Correct! 0.1736

- answer is wrong
- X = no of 1
- Y = no of 1, 2, 3
- P(X=1, Y=2) means 1 occurred once, 2/3 occurred once and 4/5/6 occurred once
- = 1/6 * 2/6 * 3/6 * 3! = 1/6 #

P(X=1, Y=2) = P(X=1 n Y=2) = P(Y=2|X=1)P(X=1) = (2C1 * 2/5 * 3/5)(3C1 * (1/6) * (5/6)^2) = 1/6

Question 3

0 / 1 pts

The joint PMF of two random variables are showed by the table below. Check their correlation and dependency.

Y\X	0	1	2
-1	1/4	0	1/8
0	0	1/4	0
-1	1/4	0	1/8

You Answered

- correlated and dependent
- uncorrelated and independent
- uncorrelated and dependent

Correct answer

- uncorrelated and dependent

- -1 is 1
- Value and the probability of taking the value of Y/X is different and depends on the value of X/Y ⇒ dependent
  - find the marginal probability first by summing across rows and column
  - the check does the product matches the joint probability, if one case doesn't, means not independent
- To check for correlation, find their covariance
  - $Cov[X, Y]$
  - $= E[(X - E[X])(Y - E[Y])]$
  - $= E[XY] - E[X]E[Y]$ use this
  - = sum of x*y*P(x,y) - product of their mean
  - = 0 - ¾ * 0
  - = 0 ⇒ not correlated

Question 4

0 / 1 pts

The joint PMF of two random variables are showed by the table below. Find E[X] and Var[Y].

Y\X	0	1	2
-1	1/4	0	1/8
0	0	1/4	0
-1	1/4	0	1/8

- 0.5 and 0.75
- 0.5 and 0

Correct answer

- 0.75 and 0.75

You Answered

- 0.75 and 0

- -1 is 1
- make sure is Y^2

Question 5 1 / 1 pts

Two variables X and Y are identical and independent distributed, which both have expectation 0 and variance 1. Find out $E[(X-Y)^2]$.

- 4
- 1
- Correct! 2
- 0

- $E[(X-Y)^2] = E[X^2] - 2E[XY] + E[Y^2]$
- make sure use $Var[X] = E[X^2] - E[X]^2$
  - this get $E[X^2] = 1$ and $E[Y^2] = 1$
- Since independent, $Cov(X, Y) = E[XY] - E[X]E[Y] = 0 \Rightarrow E[XY] = 0$
- thus $E[(X-Y)^2] = E[X^2] - 2E[XY] + E[Y^2] = 2$ #

Quiz 5

Question 1 1 / 1 pts

The probability density function of X is given as $f_X(x) = a + bx^2$ for $0 \le x \le 2$. It is known that $E[X] = 5$. What are the vlaue of a and b?

- Correct! -3.5 and 3
- 3 and -3.5
- 0.05 and 0.1
- 0.1 and 0.05

Question 2 1 / 1 pts

What is $F_X(0.5)$ if $f_X(x) = c(1 - x^2)$ for $0 < x \le 1$ and c is an unknown constant?

- None of the above.
- Correct! 0.6875
- 0.3125
- 0.875

Question 3

1 / 1 pts

Assume $X \sim Exp\,(0.1)$. What are $P\,(X > 10)$ and $P\,(X > 20|X > 10)$?

- 0.6321 and 0.6321.
- 0.6321 and 0.2231.
- Correct! 0.3679 and 0.3679.
- 0.3679and 0.2231.

P(X>10) = P(X>20|X>10) exponential has no memory

Question 4

1 / 1 pts

If $X \sim N\,(160, 25)$. what is $P(155 \leq X < 165)$? Express the results in terms of Q-function.

- Correct! 1-2Q(1)
- 1-2Q(0.2)
- 2Q(1)-1
- 2Q(0.2)-1

Question 5

1 / 1 pts

Let X be uniformly distributed from (1, 10] and $Y = \log_{10}(X)$. What is the probability $P(Y \leq 0.5)$?

- 0.3514
- 0.2162
- Correct! 0.2403
- 0.5

other stuff

Independent random variables [edit source]

Let X and Y be independent random variables that are normally distributed (and therefore also jointly so), then their sum is also normally distributed. i.e., if

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

then

$$Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2).$$

This means that the sum of two independent normally distributed random variables is normal, with its mean being the sum of the two means, and its variance being the sum of the two variances (i.e., the square of the standard deviation is the sum of the squares of the standard deviations).[1]

In order for this result to hold, the assumption that X and Y are independent cannot be dropped, although it can be weakened to the assumption that X and Y are jointly, rather than separately, normally distributed.[2] (See here for an example.)

The result about the mean holds in all cases, while the result for the variance requires uncorrelatedness, but not independence.

https://en.wikipedia.org/wiki/Sum_of_normally_distributed_random_variables

(b) Suppose you have an urn of 10 marbles, in which there are 4 blue marbles and 6 red marbles. You randomly pick a sample of 4 marbles <u>without replacement</u>. Let X be the random variable that describes the number of blue marbles found in the random sample. The probability of observing k blue marbles in a sample of 4 marbles is in fact a hypergeometric distribution:

$$P[X = k] = \frac{\binom{4}{k}\binom{6}{4-k}}{\binom{10}{4}}$$

for $k = 0,1,2,3,4$. The mean and standard deviation of X are 1.6 and 0.8, respectively. What is the expectation value of X^2? (5 marks)

can choose like this for picking without replacing

Admin Stuff

Assessments	Weightage
Formative quizzes	10%
Midterm exam	20%
Assignment	10%
Final exam	60%

Midterm exam on Monday of week 7, 2^{nd} Oct. Venue TBA.

Formative quizzes one for each chapter.
They are Canvas quizzes conducted during lecture

Tutorials are from week 4 to week 13.

- 6 sets of tutorial questions.
- 3 tutorial groups are planned.