Chapter 1

Tuesday, 9 February 2021 1:06 PM

Combinatorial Analysis

Concept	General Counting	Unique Permutations	Permutation with alike objects
Condition	If there are *r* experiments being performed: Experiment 1 - n_1 possible outcomes Experiment 2 - n_2 possible outcomes $\vdots$ Experiment r - n_r possible outcomes	Suppose that there are *n* (distinct) objects,	For *n* objects of which n_1 are alike, n_2 are alike, $\cdots, n_r$ are alike
Formula	Total Possible Outcomes of the r experiments = $n_1 \times n_2 \times \cdots \times n_r$	$Total\ number\ of\ different\ arrangements =$ $n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1 = n!$	Total number of different arrangements $= \dfrac{n!}{n_1! \times n_2! \times \cdots \times n_r!}$
Note		Remark: $0! = 1$ This is used when the ordering of the objects matter	Just divide by the time number of repeats of each of the objects multiplied with each other This is used when the ordering of the objects matter

Concept	Permutations for People sitting in a circle	Combinations	Separating a set of items
Condition	For n people sitting in a circle	If there are *n* distinct objects, of which we choose a group of *r* items	Suppose there is a set of *n* antennas of which m are defective and n-m are functional. They are all assumed to be indistinguishable
Formula	$Total\ Possible\ Arrangements =$ $\dfrac{n!}{n} = (n-1)!$	Total number of possible groups $= \dfrac{n \times (n-1) \times (n-2) \times \cdots \times (n-r+1)}{r!}$ $= \dfrac{n!}{r! \times (n-r)!}$	Total number of linear orderings so that no two defectives are consecutive: 1) Line up the (n - m) functional antennas 2) There are (n - m + 1) possible positions to insert the defectives ones in between the functional ones including spaces before and after 3) Choose m positions out of these (n - m + 1) possible positions ^1^11 $\cdots$ ^ 1 1 = functional ^ = Place for at most one defective 4) Total Number of Possible Ordering $= \binom{n-m+1}{m}$
Note		Notation: Number of ways of choosing r items from n items: nC_r or $\binom{n}{r}$ Common Identity: 1) For r = 0, 1, 2, $\cdots$, n $\binom{n}{r} = \binom{n}{n-r}$ 2) $\binom{n}{0} = \binom{n}{n} = 1$ 3) For $1 \leq r \leq n$ $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$ Convention: When n is a nonnegative integer, and r <0 or r >n then $\binom{n}{r} = 0$	

Concept	Binomial Theorem	Multinominal Theorem	
Condition	Let n be a nonnegative integer	Let n be a nonnegative integer	
Formula	$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$	$(x_1 + x_2 + \cdots + x_r)^n = \sum_{n_1+n_2+\cdots+n_r=n} \binom{n}{n_1, n_2, \cdots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n^r}$ Multinominal Coefficient: - Number of divisions to divide n objects into r distinct groups of size $n_1, n_2, \cdots, n_r$ such that $\sum_{i=1}^{r} n_i = n$ $\binom{n}{n_1, n_2, \cdots, n_r} = \dfrac{n!}{n_1! \times n_2! \times \cdots \times n_r!}$	
Note	Interpretation: - Let the 2 terms be 2 distinct groups and n to be the number of distinct people - We are trying to see how many different ways we can split people into the various sizes of groups for each of the 2 groups - k is the number of people that goes into the first group and (n - k) is the number of people that goes into the second group - For each of the x^k & y^{n-k} terms, the total number of people is n the number of people in each group and the people in each group is different - For terms that have the same size for x and y	Interpretation: - Similar idea to Binomial Theorem, we have r distinct groups now and n number of people - We are trying to see how many different ways we can split people into the various sizes of groups for each of the r groups - The various $n_1, n_2, \cdots, n_r$ represents the number of people going into each of the groups for $x_1, x_2, \cdots, x_r$ and the total number of people must be n - The multinominal coefficient is whereby we pick n_1 people for the first group, out the of the remaining, n_2 for the second group, $\cdots$,out of the last n_r , choose all of them for the last group - Since the people are distinct, they can each choose to be in either of the few groups, therefore, the multinominal coefficient counts the different combinations we can have people of that group size	

(i.e. same x^k & y^{n-k}), they will add on to the binomial coefficient which tells us how many ways we can split people into the groups of k people in group x and (n-k) people in group y.

Remark:

$\binom{n}{k}$ is often referred to as the binomial coefficient because it is coefficient of the given combination of the powers of the 2 terms

Useful Identity:

1) $\sum_{k=0}^{n} \binom{n}{k} = 2^n$
   a. **Interpretation**: The total number of ways to split the n people is 2^n since each of the people have 2 choices, either to go to group x or group y. Therefore, total is 2^n since there are n people

2) $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$

3) $\binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots$

to be in that group for the various combination of sizes for other groups as well

Useful Identity:

1) $\sum_{k=0}^{n} \binom{n}{n_1, n_2, \cdots, n_r} = r^n$
   a. **Interpretation**: The total number of ways to split the n people is r^n since each of the people have r choices to go to either of the groups. Therefore, total is r^n since there are n people

Application:

- Number of integer solutions:
  - E.g. n = 5, r = 3
  - If we want to find the combination of 3 numbers to make up to 5
  - We imagine 5 1's and 2 + (2 because we want to make 3 numbers)
    - We have n - 1 spaces between the 5's to put the +
    - 1 _ 1 _ 1 _ 1 _ 1
  - We have $\binom{n-1}{r-1}$ ways to make 3 terms

Chapter 2

Thursday, 15 April 2021 6:34 PM

Probability:

Sample Space:
- The set of all possible outcomes of an experiment, usually denoted by S

Event:
- Any subset A of the sample space is an event

Important Expansion:
- Expansion of exponential functions
  For e^x:
  $$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$
  General Term: $\frac{x^r}{r!}$

Approaches to Solving Probabilities question:
1) Combinatorics:
   a. Think of all the possible combinations of events that could occur and divide by the total sample space
   $$P(A) = \frac{Number\ of\ ways\ for\ A\ to\ occur}{Total\ number\ of\ ways}$$

2) Looking at individual probabilities:
   a. Finding the individual probabilities that could make up the required probability and multiplying them together given that they are independent
   $P(A \cap B) = P(A) \times P(B)$

Terms	Sets	Probability	Mutually Exclusive Events	Continuous Set Function
Description	Collection of items	1) Classical Approach: Assume all the sample points are likely to occur $P(E) = \frac{\lvert E\rvert}{\lvert S\rvert}$ |E| - Number of sample points in event E |S| - Number of sample points in S 2) Relative frequency approach: Try it multiple times using empirical data $P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$ n(E) - Number of times E occurs n - Number of repetitions of the experiment being carried out 3) Subjective Approach: Probability considered as a measure of belief. Start off with a belief and test out the belief and modify it after carrying out the experiment $P(E) = \frac{1}{2}$ for rolling a 6 on a dice roll because I believe that it is my lucky number	Events whereby they do not have a common sample point	Increasing Sequence: - For a sequence of events $\{E_n\}, n \geq 1$ $E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$ $\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$ Since it is an increasing sequence, if we want to find out E_n just have to get the union of all the previous events since they are smaller than the current event Decreasing Sequence: - For a sequence of events, $\{E_n\}, n \geq 1$ $E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$ $\lim_{n\to\infty} E_n = \bigcap_{i=1}^{n} E_i$ Since it is a decreasing sequence, if we want to find out E_n just have to get the intersection of all the previous events since they are all bigger than the current event so the intersection will be the current event
Properties	1) Commutative Laws: $E \cap F = F \cap E$ $E \cup F = F \cup E$ 2) Associative Laws: $(E \cap F) \cap G = E \cap (F \cap G)$ $(E \cup F) \cup G = E \cup (F \cup G)$ 3) Distributive Laws: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$ 4) DeMorgan's Law: 1. Complement of the union of items is the intersection of the complement of the events $\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} {E_i}^c$ 2. Complement of the intersection of items is the union of the complement of the events $\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$	Axioms of Probability: 1) For any event E $0 \leq P(E) \leq 1$ 2) Let S be the sample space $P(S) = 1$ 3) For any sequence of mutually exclusive events $E_1, E_2, \cdots$ $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$ Properties of Probability: 1) Impossible Event $P(\emptyset) = 0$ 2) For any infinite sequence of mutually exclusive events, $E_1, E_2, \cdots, E_n$ $P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$ 3) Complement events Let E be an event and E^c be the complement event $P(E^c) = 1 - P(E)$ 4) Subset of Events If $A \subset B$ $P(A) \leq P(B)$ 5) Relation between the union and intersection of events Let A and B be any two events $P(A \cup B) = P(A) + P(B) + P(A \cap B)$ 6) Inclusion and Exclusion Principle: Let $E_1, E_2, \cdots, E_n$ be any events	$E_i \cap E_j = \emptyset$ - There is no intersection between the 2 sets, the intersection is an empty set	Properties: Probability of an event which is the limit of sequence of monotone events is equal to the limit of the probability of these events If $\{E_n\}, n \geq 1$ is either an increasing or decreasing sequence of events $P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$

$$P(E_1 \cup E_2 \cup \cdots \cup E_n)$$
$$= \sum_{i=1}^{n} P(E_i) - \sum_{1 \le i_1 < i_2 \le n} P\left(E_{i_1} \cap E_{i_2}\right) + \cdots$$
$$+ (-1)^{r+1} \sum_{1 \le i_1 < \cdots < i_r \le n} P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right)$$
$$+ \cdots + (-1)^{n+1} P(E_1 \cap \cdots \cap E_n)$$

Because of double counting, we will need to minus and add back

Even terms minus, odd terms plus. Need to find all the combinations

7) <u>Probability of events:</u>

If event A has |A| outcomes and all in the outcomes in S are equally likely to occur

$$P(A) = \frac{number\ of\ outcomes\ in\ A}{number\ of\ outcomes\ in\ S}$$

8) <u>Multiplication of Probabilities:</u>

If we have 2 events that are occurring, assuming they are independent

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

Chapter 3

Thursday, 15 April 2021 6:57 PM

Conditional Probabilities:

Partitions:
- $A_1, A_2, \cdots, A_n$ parition the sample space S if:
  - They are **mutually exclusive events** of the sample space
    - The partitions do not intersect one another
    - $A_i \cap A_j = \emptyset, for\ all\ i \neq j$
  - They are **collectively exhaustive**
    - Any of one of the events in the sample space must lie within one of the partitions
    - $\bigcup_{i=1}^{n} A_i = S$

Steps to solving Probability Questions:
1) First Step Analysis:
   a. Understand what goes on at first
   b. Write out some base cases and find out the interaction between the various probabilities
2) Recursive relations:
   a. Find out if there are any recursive relations
3) Generalise to N terms, where N is the total number of terms
4) Generalise to i terms, where i is any arbitrary try
5) Solve for the required equation

Terms	**Conditional Probability**	**Odds**	**Independent Events**
Description	The probability of an event occurring given the occurrence of another event	Ratio of the positive outcomes to the remaining outcomes	Events are independent of each other if the occurrence of one event does not affect the probability of the other event occurring
Conditions	Let A be an event with $P(A) > 0$, all 3 conditions must hold: 1) For any event B, we have: $0 \leq P(B\|A_i) \leq 1$ 2) $P(S\|A) = 1$ Because given any condition, the probability of the sample space happening is always 1 since it compasses the whole space 3) Let $B_1, B_2, B_3, \cdots$ be a sequence of mutually exclusive events, then $P\left(\bigcup_{k=1}^{\infty} B_k \| A\right) = \sum_{k=1}^{\infty} P(B_k\|A)$ If they are mutually exclusive events, even if they are conditioned, they are still mutually exclusive since the they do not occur at the same time	A - Event that we want A^c - Complement of the event	A and B are events
Formula	Let E and F be two events. Suppose that P(F) > 0 $P(E\|F) = \frac{P(E \cap F)}{P(F)}$	$Odd(A) = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$	Independent: $P(A \cap B) = P(A) \times P(B)$ Dependent: $P(A \cap B) \neq P(A) \times P(B)$
Note	Interpretation: - Given that F has occurred, our sample space is reduced to F. Also, since F has occurred, the probability of E occurring is the fraction of the events that E occurs concurrently with F Multiplication Rule: Suppose that P(A) > 0 $P(A \cap B) = P(A) \times P(B\|A)$ General Multiplication Rule: Let $A_1, A_2, \cdots, A_n$ be n events $P(A_1 \cap A_2 \cap \cdots \cap A_n)$ $= P(A_1) \times P(A_2\|A_1) \times P(A_3\|A_2 \cap A_1) \times \cdots \times P(A_n\|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$ $\because P(A_1) \times \frac{P(A_1 \cap A_2)}{P(A_1)} \times \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \times \cdots \times \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n)}{P(A_1 \cap A_2 \cap \cdots \cap A_{n-1})}$ $= P(A_1 \cap A_2 \cap \cdots \cap A_n)$ $\because$All the other terms cancel out Bayes' Theorem: 1) Bays' First Formula (Law of Total Probability): Suppose that the events $A_1, A_2, \cdots, A_n$ partition the sample space Assume further that $P(A_i) > 0\ for\ 1 \leq i \leq n$ For any event B: $P(B) = P(B\|A_1)P(A_1) + \cdots + P(B\|A_n)P(A_n)$ $= P(B \cap A_1) + \cdots + P(B \cap A_n)$ Common Applications:	Odds of an event A tells us how much more likely it is that the event A occurs compared to the event that it does not occur	Conditional Probabilities: Since they do not affect the probability of each other occurring, even if one event occurred, the conditional probability is the unconditioned probability $P(A\|B) = P(A)$ $P(B\|A) = P(B)$ Phrasing: - A and B are independent - A is independent of B - B is independent of A Complement Events: If A and B are independent, the following are independent as well: 1) $A\ and\ B^c$ 2) $A^c\ and\ B$ 3) $A^c\ and\ B^c$ Independence between multiple events: If A and B are independent, A and C are independent, does not mean that A and $B \cap C$ are independent because we don't know if B and C are independent to each other. For n Events: - Number of equations to check: $2^n - n - 1$ - Need to check every combination for n wise combinations down to pairwise combinations For 3 Events:

- Compute P(B) when we know all of its conditional probabilities

2) Bays' Second Formula (Inverse Probabilities):

Suppose that the events $A_1, A_2, \cdots, A_n$ partition the sample space
Assume further that $P(A_i) > 0 \; for \; 1 \le i \le n$

For any event B:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)}$$

$$= \frac{P(A_i \cap B)}{P(B)} \leftarrow P(B) \; is \; using \; Bays' First \; Formula$$

Common application:

- Sensitivity and Specificity Questions
- Used to compute $P(A_i|B)$ when we know what is $P(B|A_i)$

For 3 Events:
For A, B, C to be independent (4 Equations to check)

1) Triple-wise
$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

2) Pair-wise (All 3 below)
$P(A \cap B) = P(A) \times P(B)$

3) $P(A \cap C) = P(A) \times P(C)$

4) $P(B \times C) = P(B) \times P(C)$

Chapter 4

Tuesday, 9 February 2021 5:59 PM

Discrete Random Variables:

- Random Variables are discrete if the range of X is finite or countably infinite
- There are gaps between the values of x even if it is very small like 0.1 0.2 etc.

Probability Mass Functions (PMF):

$$p_x(x) = \begin{cases} P(X = x) \; if \; x = x_1, x_2, \cdots \\ 0 \; otherwise \end{cases}$$

Properties of Probability Mass Functions:

1) $p(x_i) \geq 0; for \; i = 1, 2, \cdots;$
2) $p(x) = 0$ for other values of x
3) Since X must take on one of the values of x_i
$$\sum_{i=1}^{\infty} p_x(x_i) = 1$$

Cumulative Distribution Function (CDF):

$F_x(x) = P(X \leq x)$, $x \in \mathbb{R}$

Properties of Distribution function:

1) F_x is a nondecreasing function, i.e. if a < b, then $F_x(a) \leq F_x(b)$
2) $\lim_{b \to \infty} F_x(b) = 1$
3) $\lim_{b \to -\infty} F_x(b) = 0$
4) F_x has left limits, i.e.
$$\lim_{x \to b^-} F_x(x) \; exists \; for \; all \; b \in \mathbb{R}$$
5) F_x is right continuous, i.e.
$$\lim_{x \to b^+} F_x(x) = F_x(b) \; for \; any \; b \in \mathbb{R}$$

Calculation of probabilities using distribution function:

1) $P(a < X \leq b) = F_x(b) - F_x(a)$
2) $P(X = a) = F_x(a) - F_x(a^-), where \; F_x(a^-) = \lim_{x \to a^-} F_x(x)$
3) $P(a \leq X \leq b) = F_x(b) - F_x(a^-)$
4) $P(a \leq X < b) = F_x(b^-) - F_x(a^-)$
5) $P(a < X < b) = F_x(b^-) - F_x(a)$

Probability from PMF:

$$P(A) = \sum_{x \in A} p_x(x)$$

PMF to CDF:

$$F_x(x) = P(X \leq x) = \sum_{i=1}^{x} p_x(x_i)$$

CDF to PMF:

$p_x(x) = P(X \leq x) \; - P(X < x)$
$= F_x(x) - F_x(x^-)$

Expected Values:

- Contribution of each of the x values to the total score

$$E(X) = \sum_{x} x p_x(x)$$

Tail Sum Formula for Expectation:

- For nonnegative integer-valued random variable X

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

Expectation of a Function of a Random Variable:

$$E[g(x)] = \sum_{i} g(x_i)p_x(x_i) = \sum_{x} g(x)p_x(x)$$

Properties of Expectation:

1) E[aX] = aE[X]
2) E[X + b] = E[X] + b
3) E[aX + b] = aE[X] + b

K^{th} moment of a random variable:

$$E(X^k) = \sum_{x} x^k p_x(x)$$

K^{th} central moment:

$E[(X-\mu)^k]$

Note:

1) Expected value of a Random Variable, X is the first moment or mean of X
2) First central moment is 0
3) Second central moment is $E(X-\mu)^2$ is the variance of X

Variance:

- Measure of scattering (or spread) of the values of X around its expected value, μ

  Formula:

  $Var(X) = E[(X-\mu)^2]$

  $Var(X) = E(X^2) - [E(X)]^2$

Properties of Variance:

1) $Var(aX) = a^2 Var(X)$
2) $Var(X + b) = Var(X)$

Note:

1) $Var(X) \geq 0$
2) Var(X) = 0 if and only if X is a degenerate random variable (whereby it only takes only value and the value is the expected value, μ)
3) $E(X^2) \geq [E(X)]^2 \geq 0$

Standard Deviation

$\sigma_x = \sqrt{Var(X)}$

Properties of Standard Deviation:

1) $SD(aX) = |a|SD(X)$
2) $SD(X + b) = SD(X)$

Types of Discrete Random Variable Distributions:

Distribution	**Bernoulli Random Variable**	**Binomial Random Variable**	**Geometric Random Variable**
Description	The random variable only 2 possible outcomes and the probability of one event is p	Number of successes in n Bernoulli trials	Number of Bernoulli trials to obtain the first success
Parameters	p - Probability of success	n - Number of trials p - Probability of success	p - Probability of success
Probability	$P(X = 1) = p$	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$	$P(X = k) = pq^{k-1}$

Mass Function	$P(X = 0) = 1 - p$		
Notations	Be(p)	Bin(n, p)	Geom(p)
E(X)	$E(X) = p$	$E(X) = np$	$E(X) = \frac{1}{p}$
Var(X)	$Var(X) = p(1-p)$	$Var(X) = np(1-p)$	$Var(X) = \frac{1-p}{p^2}$
Note		<u>Approximation of Binomial Random Variable:</u> 1) **Normal Approximation** a. <u>De-Moivre-Laplace Limit Theorem:</u> Where $n \to \infty, q = 1 - p$ $Bin(n, p) \approx N(np, npq)$ $\approx \frac{X - np}{\sqrt{npq}} \sim Z(0, 1)$ Note: When $np(1-p) \geq$ the approximation will be generally quite good b. <u>Continuity Correction:</u> i. <u>Note:</u> For the range, each of the discrete values will be extended by $\frac{1}{2}$ therefore draw out the graph and see what is the range of k ii. $P(X = k) = P\left(k - \frac{1}{2} < x < k + \frac{1}{2}\right)$ iii. $P(X \geq k) = P\left(X \geq k - \frac{1}{2}\right)$ iv. $P(X > k) = P\left(X > k + \frac{1}{2}\right)$ v. $P(X \leq k) = P\left(X \leq k + \frac{1}{2}\right)$ vi. $P(X < k) = P\left(X < k - \frac{1}{2}\right)$ 2) **Poisson Approximation:** a. When n is large and p is small, np is moderate b. <u>Working Rule:</u> i. $p < 0.1 \to \lambda = np$ ii. $p > 0.9 \to \lambda = n(1-p)$ Work in terms of failure c. $Bin(n, p) \to Poisson(np)$ with accordance to the working rule	

Distribution	**Poisson Random Variable**	**Hypergeometric Random Variable**	**Negative Binomial Random Variable**
Description	Number of occurrence in an interval for the given random variable Can be used to model number of people entering etc.	Number of successes in a sample size n from a population with size N with m successes	Number of Bernoulli trials required to obtain r success
Parameters	λ - Expected number of random variable	n - Sample size N - Population size m - Number of items selected that will equate to success	r - Number of Success p - Probability of success
Probability Mass Function	$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$	$P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$P(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$
Notations	$Poisson(\lambda)$	H(n, N, m)	NB(r, p)

E(X)	$E(X) = \lambda$	$E(X) = \frac{nm}{N}$	$E(X) = \frac{r}{n}$
Var(X)	$Var(X) = \lambda$	$Var(X) = \frac{nm}{N}[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}]$	$Var(X) = \frac{r(1-p)}{p^2}$
Note	Can be used to approximate Binomial Random Variable when n is large and p is small enough: For Bin(n, p): $Poisson(\lambda = np)$ $P(X = k) \approx e^{-\lambda}\frac{\lambda^k}{k!}$ $E(X) \approx np$ $Var(X) \approx np$	For a fixed N, E(X) is large is either n or m or both are large	Geom(p) = NB(1, p)

Chapter 5

Thursday, 15 April 2021 2:18 PM

Continuous Random Variable:
- There exists a nonnegative function f_x defined for all real $x \in \mathbb{R}$, such that
$P(a < X \le b) = \int_a^b f_x(x)dx\,, for\ -\infty < a < b < +\infty$

Probability Density Function (PDF):
$f_x(x) = \frac{P(x < X < x + \delta x)}{\delta x}, where\ \delta\ is\ very\ small$

Cumulative Distribution Function (CDF):
- The distribution function is continuous

$F_x(x) = P(X \le x\,), for\ x \in \mathbb{R}$

Conversion from PDF to CDF:
$F_x(x) = \int_{-\infty}^{x} f_x(t)dt\,, x \in \mathbb{R}$

Conversion from CDF to PDF:
$f_x(x) = \frac{\partial}{\partial x}F_x(x), x \in \mathbb{R}$

Note:
1) For any $a, b \in (-\infty, \infty)$:
$$P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$$
2) If a and b converges:
$P(X = x) = 0$
3) For $a \to -\infty\ and\ b \to \infty$:
$$1 = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_x(x)dx$$

Expectation:
$E(X) = \int_{-\infty}^{\infty} x f_x(x)dx$

Expectation of function of Continuous Random Variable:
$E[g(x)] = \int_{-\infty}^{\infty} g(x)f_x(x)\,dx$

Properties of Expectation:
1) E(aX) = aE(X)
2) E(X + b) = E(X)

Variance:
$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_x(x)dx$
$Var(X) = E(X^2) - [E(X)]^2$
$E(X^2) = Var(X) + [E(X)]^2$

Properties of Variance:
1) $Var(aX) = a^2Var(X)$
2) $Var(X + b) = Var(X)$

Standard Deviation:
$SD = \sqrt{Var(X)}$

Properties of Standard Deviation:
1) SD(aX) = |a|SD(X)
2) SD(X+b) = SD(X)

Tail Sum Formula:
$E(X) = \int_0^{\infty} P(X > x)dx = \int_0^{\infty} P(X \ge x)\,dx$

Fundamental Theorem of Calculus:
$If\ F(x) = \int_a^x f(t)\,dt,$
$\frac{d}{dx}F(x) = f(x)$

Random Variable of a function of a random variable:
- Conditions:
  - g(x) is strictly **monotonic** (increasing or decreasing)
    - If it is not monotonic, we may have to break up the range so that it is monotonic, i.e ranges whereby if we cut a line across we will not get 2 values
  - **Differentiable** (Continuous)
- Random Variable Y defined by Y = g(X) has a pdf:
$$f_y(y) = \begin{cases} f_x\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right|, if\ y = g(x) for\ some\ x; \\ 0, if\ y \ne g(x)\ for\ all\ x \end{cases}$$
$g^{-1}(y)\ is\ the\ value\ of\ X\ such\ that\ g(x) = y$
- For CDF can just integrate from the pdf like normal

Cumulative Probability for Standard Normal Distribution i.e. P(Z < z)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999
3.1	0.99903	0.99906	0.9991	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.9994	0.99942	0.99944	0.99946	0.99948	0.9995
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.9996	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.9997	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.9998	0.99981	0.99981	0.99982	0.99983	0.99983

Types of Continuous Random Variable Distributions:

Distribution	**Uniform Distribution**	**Normal Distribution**	**Standard Normal Random Variable**	**Exponential Distribution**	**Chi-square (χ^2) Random Variable**
Description	The probability is uniformly distributed between an interval of values	The density function of such a distribution has a bell-shape, it is always positive and symmetric at μ and the maximum is obtained at $x = \mu$	It is a special case of Normal Distribution whereby N(0, 1)	It is a distribution whereby it also possess the memoryless property Used to model simple lifetime or just one component of a system whereby it is the time till the failure	Sum of k number of independent standard normal variables and squaring them
Parameters	a - Lower bound of the range b - Upper bound of the range	$\mu\ - Mean$ $\sigma^2\ - Variance$	$\mu = 0$ $\sigma^2 = 1$	λ > 0 $- Time\ to\ the\ 1st\ occurrence\ of\ the\ random\ variable$	k - Degrees of Freedom or also the number of independent standard normal variables
Probability Distribution Function (PDF)	$f_x(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & otherwise \end{cases}$	$f_x(x) = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ , -\infty < x < \infty$	$\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$	$f_x(x) = \begin{cases} \lambda e^{-\lambda x}, & if\ x \ge 0 \\ 0, & if\ x < 0 \end{cases}$	$f_x(x) = \begin{cases} 0, x \le 0 \\ \frac{e^{-\frac{x}{2}}x^{\frac{k}{2}-1}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}\ for\ y \ge 0 \end{cases}$ $\Gamma(a) = \int_0^{\infty} t^{a-1}\,e^{-t}\,dt$
Continuous Distribution Function (CDF)	$F_x(x) = \int_{-\infty}^{x} f_x(y)\ dy = \begin{cases} 0, if\ x < a \\ \frac{x-a}{b-a}, if\ a \le x < b \\ 1, if\ b \le x \end{cases}$	$F_x(x) = \int_{-\infty}^{\infty} f_x(x)\ dx\,, -\infty < x < \infty$	$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,dt$	$F_x(x) = \begin{cases} 0, & if\ x \le 0 \\ 1 - e^{-\lambda x} & if\ x > 0 \end{cases}$	$F_x(x) = \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}\ , for\ x \ge 0$ $\gamma_x(a) = \int_0^{x} t^{a-1}\,e^{-t}\ dt$
Notation	U (a, b)	$N(\mu, \sigma^2)$	N(0, 1)	$Exp(\lambda)$	$\chi^2(k)$
E(X)	$E(X) = \frac{a+b}{2}$	$E(X) = \mu$	E(Z) = 0	$E(X) = \frac{1}{\lambda}$	E(X) = k
$\boldsymbol{E(X^2)}$				$E(X^2) = \frac{2}{\lambda}$	
Var(X)	$Var(X) = \frac{(b-a)^2}{12}$	$Var(X) = \sigma^2$	Var(Z) = 1	$Var(X) = \frac{1}{\lambda^2}$	Var(X) = 2k
Note	Connection with Beta Distribution: $U(0, 1\,) \equiv Beta(1, 1)$	Standardising a Normal Distribution: $Y \sim N(\mu, \sigma^2) \to Z \sim N(0, 1)$ $P(a < Y \le b) = P\left(\frac{a-\mu}{\sigma} < Z \le \frac{b-\mu}{\sigma}\right)$ $= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$ *σ Rule*: $P(\lvert X - \mu\rvert \le 3\sigma) = 99.74\%$ $P(\lvert X - \mu\rvert \le 2\sigma) = 95.6\%$ $P(\lvert X - \mu\rvert \le \sigma) = 68\%$ Binomial Distribution: Can be used to approximate Binomial Distribution, more details under Chapter 4 $Bin(n, p) \to N(np, np(1-p)$	Properties of the Standard Normal Distribution: 1) **Symmetric at the point 0** $P(Z \ge 0) = P(Z \le 0) = 0.5$ 2) **Negative Standard Normal still has the same distribution** $-Z \sim N(0, 1)$ 3) **Converse Property** $P(Z \le x) = 1 - P(Z > x), for\ -\infty < x < \infty$ 4) **Symmetric Property** $P(Z \le -x) = P(Z \ge x), for\ -\infty < x < \infty$ 5) **Standardising from Normal Distribution** $If\ Y \sim N(\mu, \sigma^2), then\ X = \frac{Y - \mu}{\sigma} \sim N(0, 1)$ 6) **Going to Normal Distribution and setting own parameters**	Memoryless Property of Exponential Distribution: $P(X > s + t\ \lvert X > s) = P(X > t), for\ s, t > 0$ Connection with Gamma Distribution: $Gamma(1, \lambda) = Exp\ (\lambda)$ If we have n independent Exponential Random Variable $X_i \sim Exp(\lambda)$, the sum of them will be $\sim Gamma\ (n, \lambda)$ Connection with Weibull Distribution: $Exp(\lambda) = W(1, \lambda, 0)$ Whereby $\alpha = 1, \beta = \lambda, v = 0$	Gamma Distribution: $\chi^2(1) \equiv Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$

			$If\ X \sim N(0,1), then\ Y = aX + b \sim N(b, a^2)$ $for\ a, b\ in\ \mathbb{R}$		

Distribution	Gamma Distribution	Weibull Distribution	Cauchy Distribution	Beta Distribution	Lognormal Random Variable
Description	If events are occurring randomly in time and they all follow an Exponential Distribution with parameter, λ, the amount of time one has to wait until a total of α events occuring follows a Gamma Distribution	Used to model the life time of a system of many components			It is the random variable whereby we have a normal variable and we take it to the power of e (e^X)
Parameters	$\alpha\ - Shape\ Paramter, > 0$ $\lambda\ - Rate\ Paramter\ \left(\frac{1}{\lambda}\right), > 0$	$v\ - Location\ Parameter\ (Waiting\ time\ parameter\ or\ sometimes\ the\ shift\ paramter.\ Indicated\ the\ time\ to\ failure)$ $\alpha\ - Scale\ Parameter\ (Characteristic\ life\ paramter)$ $\beta - Shape\ Parameter\ (Weibull\ Slope\ or\ the\ threshold\ parameter)$	$-\infty < \theta < \infty$ $\alpha > 0$	a b $-\infty < a, b < \infty$	$\mu \in (-\infty, \infty)$ $\sigma > 0$
Probability Distribution Function **(PDF)**	$f_x(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & if\ x \geq 0; \\ 0, & if\ x < 0 \end{cases}$ Note: Below for the Gamma Function Expression $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1}\, dy$	$f_x(x) = \begin{cases} \frac{\beta}{\alpha}\left(\frac{x-v}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x-v}{\alpha}\right)^\beta\right], if\ x > v \\ 0 \quad if\ x \leq v \end{cases}$	$f_x(x) = \frac{1}{\pi\alpha\ [1 + \left(\frac{x-\theta}{\alpha}\right)^2}, for\ -\infty < \theta < \infty$	$f_x(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, if\ 0 < x < 1 \\ 0, \quad otherwise \end{cases}$ Note: Beta Function, B(a, b) $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt$	$f_x(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(lnx-\mu)^2}{2\sigma^2}}$
Continuous Distribution Function **(CDF)**					$F_x(x) = \frac{1}{2} + \frac{1}{2}\int_{-\infty}^{\ln(x)} \frac{\ln(x) - \mu}{\sqrt{2}\sigma}\, dx$
Notation	$Gamma(\alpha, \lambda)$	$W(v, \alpha, \beta)$	$Cauchy(\theta, \alpha)$	$Beta(a, b)$	Lognormal(μ, σ^2)
E(X)	$E(X) = \frac{\alpha}{\lambda}$	$E(X) = a\Gamma\left(1 + \frac{1}{\beta}\right) + v$	Does not exist	$E(X) = \frac{a}{a+b}$	$E(X) = e^\mu$
Var(X)	$Var(X) = \frac{\alpha}{\lambda^2}$	$Var(X) = \alpha^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2\right]$	Does not exist	$Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$	$Var(X) = \left[e^{\sigma^2} - 1\right] e^{2\mu+\sigma^2}$
Note	1) $\Gamma(1) = \int_0^\infty e^{-y}\, dy = 1$ 2) By integration by parts, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ 3) For integral values of $\alpha, say\ \alpha = n$, only applicable for integer values of gamma $\Gamma(n) = (n-1)\Gamma(n-1)$ $= (n-1)(n-2)\Gamma(n-2)$ $\cdots$ $= (n-1)!$ 4) $Gamma(1, \lambda) = Exp(\lambda)$ 5) If $X_i \sim Exp(\lambda)\ independently, then\ X_1 + \cdots + X_n \sim Gamma\ (n, \lambda)$ 6) If X $\sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right), then\ X \sim \chi^2(n)$ 7) $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-y} y^{\frac{1}{2}}\, dy = \sqrt{\pi}$	Connection with Exponential Distribution: $Exp(\lambda) = W(1, \lambda, 0)$ Whereby $\alpha = 1, \beta = \lambda, v = 0$		Connection with Uniform Distribution: $U(0,1) \equiv Beta(1,1)$ Alternative expression for Beta Function: $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$	

Distribution	Bivariate Normal Distribution	Rayleigh Distribution			
Description	Made up of two independent normal random variables and adding them together forms a normal random variable as well Normally, the bivariate normal distribution is a three-dimensional bell curve	Used in communications theory, to model multiple paths of dense scattered signals reaching a receiver			
Parameters	$\mu_x\ - Mean\ of\ X, \mu_x > 0$ $\mu_y\ - Mean\ of\ Y, \mu_y > 0$ $\sigma_x\ - Standard\ Deviation\ of\ X, \sigma_x > 0$ $\sigma_y\ - Standard\ Deviation\ of\ Y, \sigma_y > 0$ $\rho\ - Correlation\ of\ X\ and\ Y, -1 < \rho < 1$	$\sigma\ - Shape\ Parameter$			
Probability Distribution Function **(PDF)**	$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{z}{2(1-\rho^2)}}$ $z = \left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - \frac{2\rho\left((x-\mu_x)(y-\mu_y)\right)}{\sigma_x\sigma_y}$ $\rho = Cor(x,y) = \frac{Cov(x,y)}{\sigma_x\sigma_y} = \frac{E(XY) - E(X)E(Y)}{\sigma_x\sigma_y}$	$f_x(x) = \frac{x}{\sigma^2} e^{-\left(\frac{x^2}{2\sigma^2}\right)}$			
Continuous Distribution Function **(CDF)**					
Notation	$X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2 + 2\rho(X,Y)\sigma_x\sigma_y)$	$Rayleigh(\sigma)$			
E(X)		$E[X] = \sigma\sqrt{\frac{\pi}{2}}$			
Var(X)		$Var(X) = \frac{\sigma^2(4-\pi)}{2}$			
Note	Both X and Y are normally distributed and are independent with each other The Bivariate normal distribution is the sum of the 2 of them	Weibull Distribution: Special Case of Weibull Distribution: Scale Parameter of Weibull is 2 Chi-Squared Distribution: Special Case of Chi-Squared Distribution: When the shape parameter, $\sigma = 1$, it is a chi square distribution with 2 degrees of freedom			

Jointly Distributed Random Variables:

2 Random Variables:

Note: Need to always take note of the range in which the distribution is valid, when there are conditions, we will set one of the variables under the condition first and then we look at the condition of the other variable to see the range of the sum or integration

Important Calculations:
Let $a, b, a_1 < a_2, b_1 < b_2$ *be real numbers*:

1) $P(X > a, Y > b) = 1 - F_x(a) - F_y(b) + F_{X,Y}(a,b)$

2) $P(a_1 < X \le a_2, b_1 < Y \le b_2)$
$= P(X \le a_2, Y \le b_2) - P(X \le a_1, Y \le b_2) + P(X \le a_1, Y \le b_1) - P(X \le a_2, Y \le b_1)$
$= F_{X,Y}(a_2, b_2) - F_{X,Y}(a_1, b_2) + F_{X,Y}(a_1, b_1) - F_{X,Y}(a_2, b_1)$

Type of Random Variable	Discrete Random Variable	Continuous Random Variable
Joint Probability Mass/Density Function	$p_{X,Y}(x,y) = P(X = x, Y = y)$	*For every set C of pairs of real numbers:* $P((X,Y) \in C) = \int\int_{X,Y \in C} f_{X,Y}(x,y)\, dx\, dy$
Marginal Probability Mass Function	$p_x(x) = P(X = x) = \sum_{y \in \mathbb{R}} p_{X,Y}(x,y)$ $p_y(y) = P(Y = y) = \sum_{x \in \mathbb{R}} p_{X,Y}(x,y)$	$f_x(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$ $f_y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$
Joint Distribution Function	$F_{X,Y}(a,b) = P(X \le a, Y \le b) = \sum_{X \le a} \sum_{y \le b} p_{X,Y}(x,y)$	$F_{X,Y}(x,y) = P(X \le x, Y \le y), \quad for\ x, y \in \mathbb{R}$ Conversion to PDF from CDF: $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$
Marginal Distribution Function	$F_x(X) = \sum_{y \in \mathbb{R}} F_{X,Y}(x,y)$ $F_y(Y) = \sum_{x \in \mathbb{R}} F_{X,Y}(x,y)$	$F_x(X) = \lim_{y \to \infty} F_{X,Y}(x,y)$ $F_y(Y) = \lim_{x \to \infty} F_{X,Y}(x,y)$
Independence	*For random variables X and Y to be independent* *For all* $x, y \in \mathbb{R}$, *we have:* ***PMF***: $p_{X,Y}(x,y) = p_x p_y(y)$ ***CDF***: $F_{X,Y}(x,y) = F_x(x) F_y(y)$	*For random variables X and Y to be independent* *For all* $x, y \in \mathbb{R}$, *we have:* **PDF**: $f_{x,y}(x,y) = f_x(x) f_y(y)$ **CDF**: $F_{X,Y}(X,Y) = F_x(x) F_y(y)$
Note	$\{X > a, Y > b\} \neq \{X \le a, Y \le b\}^c$ Useful Formulas: 1) $P(a_1 < X \le a_2, b_1 < Y \le b_2) = \sum_{a_1 < x \le a_2} \sum_{b_1 < y \le b_2} p_{X,Y}(x,y)$ 2) $P(X > a, Y > b) = \sum_{X > a} \sum_{Y > b} p_{X,Y}(x,y)$	Useful Formulas: 1) Let $A, B \subset \mathbb{R}$, take C = A x B above (Because it should be within the product space of the given event C) $P(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x,y)\, dy\, dx$ *2)* In particular, Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ *where* $a_1 < a_2$ *and* $b_1 < b_2$ $P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x,y)\, dy\, dx$ *3)* Let $a, b \in \mathbb{R}$ $F_{X,Y}(a,b) = P(X \le a, Y \le b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x,y)\, dy\, dx$

Independence:
Can be applied to any number of random variables to get the PMF/PDF/CDF so long as each of the random variables are independent to each other
$P(X \in A, Y \in B) = P(X \in A) P(Y \in B)\ for\ any\ A, B \subset \mathbb{R}$

Random Variables X and Y are independent if and only if there exist functions $g, h : \mathbb{R} \to \mathbb{R}$ such that for all x, y ∈ ℝ, we have
- We can split them into separate functions of X and Y, but we have to ensure that the product space is a rectangle so that they are independent

$$f_{X,Y}(x,y) = h(x) g(y)$$

Sum of Independent Random Variables:

Continuous and Independent:
CDF:
$$F_{X+Y}(a) = \int_{-\infty}^{\infty} F_x(a-y) f_y(y)\, dy = \int_{-\infty}^{\infty} F_y(a-x) f_x(x)\, dx$$
PDF:
$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_x(a-y) f_y(y)\, dy = \int_{-\infty}^{\infty} f_x(x) f_y(a-x)\, dx$$

Distribution	Uniform Random Variable	Gamma Random Variable	Exponential Random Variable	Normal Random Variable
Condition	X and Y are independent X ~ U(0, 1) Y ~ U(0, 1)	X and Y are independent $X \sim Gamma\ (\alpha, \lambda)$ $Y \sim Gamma\ (\beta, \lambda)$	If there are n independent exponential random variables each having parameter λ $X_1, X_2 \cdots, X_n$	$X_i, i = 1, \cdots, n$ are independent normal random variables with respective parameters μ_i, σ_i^2 $i = 1, \cdots, n$
Sum	$f_{X+Y}(a) = \begin{cases} a, & 0 \le a \le 1; \\ 2-a, & 1 < a < 2; \\ 0, & elsewhere \end{cases}$	$X + Y \sim Gamma(\alpha + \beta, \lambda)$	$X_1 + X_2 + \cdots + X_n \sim Gamma\ (n, \lambda)$	$X_1 + X_2 + \cdots + X_n \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma^2\right)$
Note	Can be done for any uniform random variable 1) Take note of the total range after summing them up 2) Find the PDF of the sum of the random variables using the formula 3) Take note of the range in which they are different (i.e. Positive or negative) 4) Find the corresponding PDF after figuring out the range	Both of the random variables must have the same λ	This follows from the sum of independent Gamma distributions $X_1 \sim Gamma(1, \lambda)$ $X_1 + X_2 \sim Gamma(2, \lambda)$	The mean and variance does not have to be the same, they can all be different and we can just sum of all them together

Jointly Distributed Random Variables: $n \ge 3$

Joint Probability Density Function of X, Y, Z: $f_{X,Y,Z}(x,y,z)$

1) For any $D \subset \mathbb{R}^3$, *we have*
$$P((X,Y,Z) \in D) = \iiint_{(x,y,z) \in D} f_{X,Y,Z}(x,y,z)\, dx\, dy\, dz$$
2) Let A, B , C ⊂ ℝ, take D = $A \times B \times C$ above
$$P(X \in A, Y \in B, Z \in C) = \int_C \int_B \int_A f_{X,Y,Z}(x,y,z)\, dx\, dy\, dz$$

Marginal Probability Density Functions of X, Y, Z:
1) $f_x(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dy\, dz$
2) $f_y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dx\, dz$
3) $f_z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dx\, dy$
4) $f_{X,Y}(x,y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dz$
5) $f_{X,Z}(x,z) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dy$
6) $f_{Y,Z}(y,z) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dx$

Joint Distribution Function of X, Y, Z:
$F_{X,Y,Z}(x,y,z) = P(X \le x, Y \le y, Z \le z)$

Marginal Distribution Function of X, Y, Z:
1) $F_{X,Y}(x,y) = \lim_{Z \to \infty} F_{X,Y,Z}(x,y,z)$
2) $F_{X,Z}(x,z) = \lim_{Y \to \infty} F_{X,Y,Z}(x,y,z)$
3) $F_{Y,Z}(y,z) = \lim_{X \to \infty} F_{X,Y,Z}(x,y,z)$
4) $F_x(x) = \lim_{y \to \infty, z \to \infty} F_{X,Y,Z}(x,y,z)$
5) $F_y(y) = \lim_{x \to \infty, z \to \infty} F_{X,Y,Z}(x,y,z)$
6) $F_z(z) = \lim_{x \to \infty, y \to \infty} F_{X,Y,Z}(x,y,z)$

Independence:

For Jointly Continuous Random Variable, these 3 statements are equivalent:

1) Random Variables X, Y, Z are independent

2) For all $x, y, z \in \mathbb{R}$, we have
$f_{X,Y,Z}(x,y,z) = f_x(x) f_y(y) f_z(z)$

3) For all $x, y, z \in \mathbb{R}$, we have
$F_{X,Y,Z}(x,y,z) = F_X(x) F_Y(y) F_Z(z)$

Checking for independence:
Random variables X, Y, Z are independent if and only if there exist functions $g_1, g_2, g_3: \mathbb{R} \to \mathbb{R}$ such that for all $x, y, z \in \mathbb{R}$

$f_{X,Y,Z}(x,y,z) = g_1(x) g_2(y) g_3(z)$

Conditional Distributions:
$f_{X,Y|Z}(x,y|z) = \frac{f_{X,Y,Z}(x,y,z)}{f_z(z)}$
$f_{X|Y,Z}(x,|y,z) = \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}$

Discrete and Independent:

Distribution	Poisson Random Variable	Binomial Random Variables
Condition	X & Y are independent $X \sim Poisson(\lambda)$ $Y \sim Poisson(\mu)$	X & Y are independent $X \sim Bin(n, p)$ $Y \sim Bin(m, p)$
Sum	$X + Y \sim Poisson\ (\lambda + \mu)$	$X + Y \sim Bin(n + m, p)$
Note	The means of the independent Poisson random variables can be different The mean of the new Poisson random variable is the sum of the independent Poisson random variables	The probability of the independent Binomial random variables must be the same, the number of trials can be different The number of trials of the new Binomial random variables is the sum of the independent Binomial random variables and it will have the same probability of success

Conditional Distribution:

Note: Take note of the form of the result that is calculated, see if it resembles other distribution functions which will make computation much simpler

Type of Random Variable	Discrete Random Variables	Continuous Random Variables
Conditions	X given Y = y	X given Y = y
Conditional Probability Mass/Density Function	$p_{X\|Y}(x\|y) = P(X = x\|Y = y)$ $= \dfrac{p_{X,Y}(x, y)}{p_y(y)}\ such\ that\ p_y(y) > 0$	$f_{X\|Y}(x\|y) = \dfrac{f_{X,Y}(x, y)}{f_y(y)}\ for\ all\ Y\ such\ that\ f_y(y) > 0$
Conditional Distribution Function	$F_{X\|Y}(x\|y) = P(X \le x\|Y = y)$ $= \sum_{a \le x} p_{X\|Y}(a\|y)\ such\ that\ p_y(y) > 0$	$F_{X\|Y}(x\|y) = P(X \le x\ \|Y = y) = \int_{-\infty}^{x} f_{X\|Y}(t\|y)\ dt$
Independence	For all $Y\ such\ that\ p_Y(y) > 0$ $p_{X\|Y}(x\|y) = p_X(x)$	For all Y such that $p_y(y) > 0$ $f_{X\|Y}(x\|y) = f_x(x)$
Note	The same logic goes for the condition the other way around	The same logic goes for the condition the other way around

Joint Probability Distribution Function of Functions of Random Variables:

Conditions:

1) Let X and Y be jointly continuous distributed random variables with known joint probability density function (pdf)

2) Let U and V be given functions of X and Y in the form:
   U = g(X, Y) , V = h(X, Y)
   And we can uniquely solve X and Y in terms of U and V,
   x = a(u, v) and y = b(u, v)
   This ensures that we can find the inverse function of g and h

3) The functions g and h have continuous partial derivates at all points (x, y) and

$$J(x, y) = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} = \frac{\partial g}{\partial x}\left(\frac{\partial h}{\partial y}\right) - \frac{\partial g}{\partial y}\left(\frac{\partial h}{\partial x}\right) \neq 0$$

   For all points (x, y), remember that Jacobian is absolute value

Joint Probability Density Function of U and V whereby they are both functions of X and Y:

$$f_{U,V}(u, v) = f_{X,Y}(x, y)\ [J(x, y)]^{-1}$$

$J(x, y)$ is the Jacobian determinant of g and h:

$$J(x, y) = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} = \frac{\partial g}{\partial x}\left(\frac{\partial h}{\partial y}\right) - \frac{\partial g}{\partial y}\left(\frac{\partial h}{\partial x}\right)$$

Multiple Random Variables:

Given the joint density function of n random variables $X_1, X_2 \cdots, X_n$

For $Y_1 = g_1(X_1, \cdots, X_n), Y_2 = g_2(X_1, \cdots X_n), \cdots Y_n = g_n(X_1, \cdots, X_n)$

Conditions still hold:

1) There must still be a unique solution for g_i
2) $J(x_1, x_2, \cdots, x_n) \neq 0$

Joint Probability Density Functions of the functions of $X_1, X_2, \cdots, X_n$:

$$f_{Y_1,Y_2,\cdots,Y_n}(y_1, y_2, \cdots, y_n) = f_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n)[J(x_1, x_2, \cdots, x_n)]^{-1}$$

Jacobian Determinant:

$$J(x_1, x_2, \cdots, x_n) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \frac{\partial g_n}{\partial x_2} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}$$

Chapter 7

Thursday, 15 April 2021 6:57 PM

Properties of Expectation:

Note:
$If\ a \le X \le b, then\ a \le E(X) \le b$

Boole's Inequality:

- Probability of union of a countable set of events is less than the sum of probabilities of each of these events. This is because we do not assume the A_i's are mutually exclusive. Even if they are mutually exclusive, the equation below will be equals to each other

$$P\left(\bigcup_{k=1}^{n} A_k\right) \le \sum_{k=1}^{n} P(A_k)$$

Calculating Probabilities by conditioning:

- Similar idea to Bayes First Theorem (Law of Total Probability)

Discrete Random Variables:

$$P(A) = \sum_y P(A|Y=y)P(Y=y)$$

Continuous Random Variables:

$$P(A) = \int_{-\infty}^{\infty} P(A|Y=y) f_y(y)$$

Example:

$$P(X<Y) = \int_{-\infty}^{\infty} P(X<Y|Y=y)\ f_y(y)\ dy$$
$$= \int_{-\infty}^{\infty} P(X<y|Y=y) f_y(y) dy$$
$$= \int_{-\infty}^{\infty} P(X<y)\ f_y(y) dy$$
$$= \int_{-\infty}^{\infty} F_x(x) f_y(y)\ dy$$

Independence of Mean Variance from Normal Sample:
Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables with mean μ and variance σ^2
- Sample Mean $\bar{\boldsymbol{X}}$ and Sample Variance $\boldsymbol{S^2}$ are independent

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\ and\ \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Term	Expectation of 2 Random Variables	Mean of 2 Random Variables	Variance of 2 Random Variables	Covariance	Correlation/Correlation Coefficient	Conditional Expectation
Description	-	-		Measures the total variation of two random variables from their expected values	Correlation measures the strength of the relationship between variables It measures the degree of linearity between X and Y. • If X increase, Y increase → Positive Correlation • If X increase, Y decrease → Negative Correlation	-
Discrete Random Variable	If X and Y are jointly discrete with joint pmf, $p_{X,Y}(x,y)$ $E[g(X,Y)] = \sum_y \sum_x g(x,y) p_{X,Y}(x,y)$	-	-	Covariance of jointly distributed random variables X and Y, denoted by Cov(X, Y) $Cov(X,Y) = E\left[(X-\mu_x)(Y-\mu_y)\right]$ Or $Cov(X,Y) = E(XY) - E(X)E(Y)$ $\mu_x\ - Mean\ of\ X$ $\mu_y\ - Mean\ of\ Y$	Correlation Coefficient of random variables X and Y, denoted by $\rho(X,Y)$: $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$	If X and Y are jointly distributed discrete random variables: $E[X\|Y=y] = \sum_x x\, p_{X\|Y}(x\|y)\ , if\ p_y(y) > 0$ Function of a Random Variable: $E[g(X)\|Y=y] = \sum_x g(x) p_{X\|Y}(x\|y)$
Continuous Random Variable	If X and Y are jointly continuous with joint pdf, $f_{X,Y}(x,y)$ $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\ dx\ dy$	-	-	Covariance of jointly distributed random variables X and Y, denoted by Cov(X, Y) $Cov(X,Y) = E\left[(X-\mu_x)(Y-\mu_y)\right]$ Or $Cov(X,Y) = E(XY) - E(X)E(Y)$ $\mu_x\ - Mean\ of\ X$ $\mu_y\ - Mean\ of\ Y$	Correlation Coefficient of random variables X and Y, denoted by $\rho(X,Y)$: $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$	If X and Y are jointly continuous random variables: $E[X\|Y=y] = \int_{-\infty}^{\infty} x\, f_{X\|Y}(x\|y)\, dx\ \ if\ f_y(y) > 0$ Function of a Random Variable: $E[g(X)\|Y=y] = \int_{-\infty}^{\infty} g(x) f_{X\|Y}(x\|y)\, dx$
Sum	1) $E[g(X,Y)+h(X,Y)]$ $= E[g(X,Y)] + E[h(X,Y)]$ 2) $E[g(X)+h(Y)] = E[g(X)] + E[h(Y)]$	**Mean of Sum=Sum of Mean** $E(X+Y) = E(X) + E(Y)$ General Case: $E(a_1X_1 + \cdots + a_nX_n)$ $= a_1E(X_1) + \cdots + a_nE(X_n)$ Note: This relation does not require all the random variables to be independent to one another	Variance of a Sum: $Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k) + 2\sum_{1\le i<j\le n} Cov(X_i, X_j)$ $where\ \sum_{1\le i<j\le n} Cov(X_i,X_j)\ is\ the\ total\ number\ of\ Covariance\ terms\ and\ most\ of\ the\ time\ it\ should\ be$ $\frac{n(n-1)}{2}$ and when multiplied with 2 is $n(n-1)$	-	-	Sum of Conditional Probabilities: $E\left[\sum_{k=1}^{n} X_k\|Y=y\right] = \sum_{k=1}^{n} E[X_k\|Y=y]$
Independence between the Random Variables	Above holds regardless of independence	Above holds regardless of independence If X and Y are independent then for any functions $g, h: \mathbb{R} \to \mathbb{R}$ $E[g(X)h(y)]$ $= E[g(x)]E[h(y)]$	Variance of a Sum under independence: Let $X_1, \cdots, X_n$ be independent random variables, then **Variance of Sum = Sum of Variance** $Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k)$	Independence of 2 Random Variables: If X and Y are independent then for any functions $g, h: \mathbb{R} \to \mathbb{R}$ $E[g(X)h(y)] = E[g(x)]E[h(y)]$ $Cov(X,Y) = 0$ **Note**: Cov(X, Y) = 0 does not imply independence	-	Sum of Conditional Probabilities if they are independent: $E\left[\sum_{k=1}^{n} X_k\|Y=y\right] = \sum_{k=1}^{n} X_k$
Important Relations	1) $If\ g(x,y) \ge 0\ whenever\ p_{X,Y}(x,y) > 0, then\ E[g(X,Y)] \ge 0$ 2) Monotone Property: If jointly distributed random variables X and Y satisfy $X \le Y$ $E(X) \le E(Y)$	Sample Mean: $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$	Sample Variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ Law of Total Variance: $Var(X) = E[Var(X\|Y)] + Var[(E(X\|Y)]$	Properties of Covariance: 1) $Var(X) = Cov(X,X)$ 2) $Cov(X,Y) = Cov(Y,X)$ 3) $Cov\left(\sum_{i=1}^{n} a_iX_i, \sum_{j=1}^{m} b_jY_j\right)$ $= \sum_{i=1}^{n}\sum_{j=1}^{m} a_ib_jCov(X_i,Y_j)$ 4) $Cov(I_A, I_B) = P(AB) - P(A)(B)$ $= P(B)[P(A\|B) - P(A)]$ 5) $Cov(X_i - \bar{X}, \bar{X}) = 0$ 6) $Cov(X,a) = 0, where\ a\ is\ a\ constant$ 7) Scalar mutliple just have to bring out like Expectation $Cov(aX, bY) = abCov(X,Y)$ 8) Addition just have to take into account all possible combinations $Cov(X+Y, X-Y)$ $= Cov(X,Y) + Cov(X,-Y) + Cov(Y,X) + Cov(Y,-Y)$ Note: • If $Cov(X,Y) \ne 0$, X and Y are correlated • If $Cov(X,Y) = 0$, X and Y are uncorrelated	Range: $-1 \le \rho(X,Y) \le 1$ Results: 1 - Strong Positive Correlation 0 - No Correlation -1 - Strong Negative Correlation Note: 1) $\rho(X,Y)$ is dimensionless or does not depend on the magnitude of X and Y because it has been standardised already 2) $\rho(X,Y) = 0$ does not imply X and Y are independent, it only means they are uncorrelated 3) $\rho(X,Y) = 1\ if\ and\ only\ if\ Y = aX + b\ where\ a = \frac{\sigma_y}{\sigma_x} > 0$ 4) $\rho(X,Y) = -1\ if\ and\ only\ if\ Y = aX + b\ where\ a = -\frac{\sigma_y}{\sigma_x} < 0$ 5) Similar to covariance, if X and Y are independent then $\rho(X,Y) = 0$ However, the converse is not true	Expectation of Conditional Expectation: $E[X] = E[E[X\|Y]]$ Relationship between Conditional Distributions: 1) If X and Y are independent Binomial Random Variable with Bin(n, p) $E[X\|X+Y=m] \sim Hypergeometric\ (n,m,2n)$ Expectation of a Random Sum: - Suppose that $X_1, X_2, \cdots, X_k$ are independent and identically distributed with common mean μ - Suppose that N is a nonnegative integer value random variable independent of the X_k's ○ **Interpretation**: N denotes the number customers entering a department store during a period of time; X_k's amount spent by the kth customer; T total Revenue - To find the mean of $T = \sum_{k=1}^{n} X_k$ $\boldsymbol{E[T] = \mu E[N]}$

Moment Generating Functions:

Moment generating functions are functions that generate the moments of the random variable and they are always positive whereby $E[e^{tX}] \ge 0$

Differentiating n number of times and subbing in t = 0, will give the nth moment
$For\ n \ge 0,$

$$E(X^n) = M_x^{(n)}(0)$$

Where $M_x^{(n)}(0) = \frac{d^n}{dt^n}M_x(t)\Big|_{t=0}$

Discrete Random Variables:

$$M_x(t) = E[e^{tX}] = \sum_x e^{tX} p_x(x)$$

Continuous Random Variables:

$$M_x(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tX} f_x(x) dx$$

Properties of Moment Generating Functions (MGF):

1) Multiplicative Property

If X and Y are independent
$M_{X+Y}(t) = M_X(t)M_Y(t)$

2) Uniqueness Property
Let X and Y be random variables with their moment generating functions, $M_X(t)\ and\ M_Y(t)$
Suppose that there exists a h > 0 such that:

$M_x(t) = M_y(t)\ for\ all\ t \in (-h\ , h)$

Then X and Y have the same distribution (i.e. $F_x = F_y\ or\ f_x = f_y$). If they have the same MGF, they will have the same distribution

Distributions and MGFs:

Distribution	**Bernoulli Random Variable**	**Binomial Random Variable**	**Geometric Random Variable**	**Poisson Random Variable**
Parameters	X ~ Be(p)	X ~ Bin(n, p)	X ~ Geom(p)	X ~ $Poisson(\lambda)$
Moment Generating Function	$M_x(t) = 1 - p + pe^t$	$M_x(t) = \left(1 - p + pe^t\right)^n$	$M_x(t) = \dfrac{p\ e^t}{1 - (1 - p)e^t}$	$M_x(t) = e^{\lambda(e^t - 1)}$
Sum of 2 MGF		X ~ Bin(n, p) Y ~ Bin(m, p) $M_{X+Y}(t) = \left[1 - p + pe^t\right]^{n+m}$ ∴ X+Y is a Binomial Distribution by the uniqueness theorem since it is in the form of a Binomial Distribution $X + Y \sim Bin(n + m, p)$		$X \sim Poisson(\lambda_1)$ $Y \sim Poisson(\lambda_2)$ $M_{X+Y}(t) = e^{(\lambda_1 + \lambda_2)(e^t - 1)}$ ∴ X+Y is a Poisson Distribution by the uniqueness theorem $X + Y \sim Poisson(\lambda_1 + \lambda_2)$
Note			$p_x(x) = pq^{x-1}, for\ x = 1, 2 \cdots$	

Distribution	**Uniform Random Variable**	**Exponential Random Variable**	**Normal Random Variable**	**Chi-Squared Random Variable**
Parameters	X ~ U(α, β)	X ~ Exp(λ)	$X \sim N(\mu, \sigma^2)$	$X \sim \chi^2(n)$
Moment Generating Function	$M_x(t) = \dfrac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$	$M_x(t) = \dfrac{\lambda}{\lambda - t}, for\ t < \lambda$	$M_x(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$	$M_y(t) = (1 - 2t)^{-\frac{n}{2}}$
Sum of 2 MGF			$X \sim N(\mu_1, \sigma_1^2)$ $Y \sim N(\mu_2, \sigma_2^2)$ $M_{X+Y}(t) = e^{(\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2}$ ∴ X+Y is a Normal Distribution by the uniqueness theorem $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$	
Note			For Standard Normal Distribution: X ~ N(0, 1) $M_x(t) = e^{\frac{t^2}{2}}$	Each of the Normal Random Variables that are being squared and summed to make the Chi-Squared distribution are independent of one another

Joint Moment Generating Function:

For any n random variables $X_1, X_2, \cdots, X_n$:

$M_{X_1, \cdots, X_n}(t_1, \cdots, t_n) = E\left[e^{t_1 X_1 + \cdots + t_n X_n}\right]$

Individual Moment Generating Function from the Joint Moment Generating Function:
- Sub those variables that we want to t and the remaining to 0
- Can be used to find the moment generating function of 2 variables from n by just subbing the rest to be 0 and so on.

$M_{X_i}(t) = E\left[e^{tX_i}\right] = M_{X_1, \cdots, X_n}(0, \cdots, 0, t, 0, \cdots, 0)$

Independence between the n random variables:
- Can be used to prove independence between random variables

$M_{X_1, \cdots, X_n}(t_1, \cdots, t_n) = M_{X_1}(t_1)M_{X_2}(t_2) \cdots M_{X_n}(t_n)$

Chapter 8

Thursday, 15 April 2021 6:58 PM

Limit Theorems:

Theorem	**Markov's Inequality**	**Chebyshev's Inequality**	**Weak Law of Large Numbers**	**Central Limit Theorem**
Description	Provides an upper bound for $P(X \geq a)$ The bigger the value, a, the smaller the bound	Provides a bound for the probability that a random variable X, lies outside the range $\mu \pm a$	States that the probability that the **sample mean, $\bar{X}$** deviating from the **mean, μ** by **ϵ** will tend to 0 if **n** tends to ∞	If we were to take **n number of samples** from the same random distribution, the distribution of the sample means will tend towards a normal distribution with the same **mean, μ** and a standard deviation (standard error) of $\boldsymbol{\sigma_x} = \frac{\sigma}{\sqrt{n}}$
Condition	Let X be a **nonnegative** random variable. For a > 0	Let X be a random variable with mean μ, then for a > 0	Let $X_1, X_2, \cdots$ be a sequence of independent and identical distributed random variables, with common mean, μ, Then for any $\epsilon > 0$	Let $X_1, X_2, \cdots$ be a sequence of independent and **identically distributed** random variables, each with **mean μ** and **variance σ^2**
Result	$P(X \geq a) \leq \frac{E[X]}{a}$	$P(\lvert X-\mu \rvert \geq a) \leq \frac{Var(X)}{a^2}$	$P\left(\left\lvert \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right\rvert \geq \epsilon\right) \to 0$ As $n \to \infty$	<u>Sum of the samples that we have taken</u> $X_1 + X_2 + \cdots + X_n \to N\left(\mu, \frac{\sigma^2}{n}\right)$ <u>Standardising:</u> <u>First Form:</u> $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$ $where\ \bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$ <u>Second Form:</u> $\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to N(0,1)$ $as\ n \to \infty,$ $n\mu = E(X_1 + X_2 + \cdots + X_n)$ $\sigma\sqrt{n} = \sqrt{Var(X_1 + \cdots + X_n)}$ $\lim_{n\to\infty} P(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ If we divide n for the top and bottom, we will get back the first form
Note	<u>Requirement:</u> Only requires the **mean, μ** of the distribution to find an upper bound	<u>Requirement:</u> Only requires the **mean, μ** and the **variance, $Var(X)$** of the distribution to find the upper bound <u>Validity:</u> Valid for all distributions of the random variable X, however the bound of the probability may not be close to the actual probability <u>Can Imply:</u> $Var(X) = 0 \to$ Random Variable X is constant. Since the probability that it deviates from the mean is 0 for all values of a	<u>Note:</u> No matter how big the value of n is, there is still a chance that $\lvert \bar{X} - \mu \rvert \geq \epsilon$ will occur	<u>Lemma to Prove:</u> Let $Z_1, Z_2, \cdots$ be a sequence of random variables with distribution function F_{Z_n} and moment generating function, M_{Z_n}. Let Z be a random variable having distribution function F_Z and moment generating function M_Z If $M_{Z_n}(t) \to M_Z(t)$ for all t $F_{Z_n}(x) \to F_Z(x)$ for all x at which $F_Z(x)$ is continuous <u>Normal Approximation:</u> Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables, each having $mean, \mu$ and variance, σ^2. Then for n (**$n \geq 30$**) large: $\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to N(0,1)$ <u>For $-\infty < a, b < \infty$:</u> $P\left(a < \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \approx \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{t^2}{2}} dt$ **Note**: For Discrete Random Variables, continuity correction will still need to be applied

Theorem	**Strong Law of Large Numbers**	**One-Sided Chebyshev's Inequality**	**Jensen's Inequality**	
Description	For **any number larger than n**, when the Weak Law of Large Numbers apply, any value of n will cause $\bar{X}$ to tend to μ	-	If we have a convex function of a random variable, the expectation of the function is larger or equals to the function of the expectation	
Condition	Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = E[X_i]$. Then with **probability 1**, we have the result	If X is a random variable with mean 0 and finite variance σ^2, then for any a > 0	If g(x) is a convex function	
Result	$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$ As $n \to \infty$	$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$	$E[g(X)] \geq g(E[X])$	
Note	<u>Difference between Weak and Large Laws of large numbers:</u> The weak law states that for any specified large number n^*, $(X_1 + X_2 + \cdots + X_n)/(n^*)$ is likely to be near μ However, it does not state that the sample mean is bound to stay near μ for ALL values of n larger than n^*	The bound obtained by Markov inequality is weaker than the one obtained using one sided Chebyshev's inequality but we use it since it easier as well and is used to prove Chebyshev's inequality This provides a tighter and more accurate bound	<u>Conditions for a function to be convex:</u> It has a concave shape to the graph (E.g. $g(x) = x^2$) One of the 3 conditions 1) A function g(x) is convex if for all $0 \leq p \leq 1$ and all $x_1, x_2 \in R_x$ $g(px_1 + (1-p)x_2) \leq pg(x_1) + (1-p)g(x_2)$ 2) A differentiable function of one variable is convex on an interval if and only if $g(x) \geq g(y) + g'(y)(x-y)$	

	The large law states that for any positive $\mid \sum_{i=1}^{n} \left(\frac{X_i}{n} - \mu\right) \mid \geq \epsilon$ all the time if n is larger than n^*	as compared to Markov's inequality	For all x and y in the interval (i.e. its graph lies above all of its tangents) 3) A twice differentiable function of one variable is convex over an interval if and only if its second derivative is non-negative there	