Week 1: Introduction to Probability Theory

Set theory:

- Outcomes are the most basic possible results from an experiment
  - They are unique & mutually exclusive
  - Outcome = Heads or Tails
- Sample space is the set of all possible outcomes from an experiment
  - $S = \{(H,H),(H,T),(T,H),(T,T)\}$
- An event is collection of one or more outcomes sharing a common property of interest
  - It is a proper subset of the Sample space
  - $E\ (One\ head\ appears) = \{(H,T),(T,H)\}$
- Important combinations of Events:
  - Union (A u B) → Either A, B or Both occur
  - Intersection (A n B) → A and B occur simultaneously
  - Complement (A^c) → All outcomes NOT under Event A
  - Mutually exclusive → Events CANNOT occur simultaneously
  - Collectively exhaustive → At least one of the events MUST occur
- Algebraic rules:
  - Commutative/Associative → Can flip around
  - Distributive → Factorised in & out

Probability Theory:

- Probability is the long run frequency an outcome in the long run
  - Always between 0 and 1
  - $P(Event) = \frac{Number\ of\ ways\ Event\ can\ occur}{Total\ number\ of\ possible\ outcomes}$
- Conditional probability is the probability Event B occurs given that Event A has occurred
  - Event must be in both Event A and B
  - Event A must have occurred → Event A is the reduced sample space
  - $P(B|A) = \frac{P(BA)}{P(A)}$
- Independence is when the knowledge that event A occurs does not change the probability that event B occurs
  - $P(A\ n\ B) = P(A) * P(B)$
  - $P(A\ n\ B^c) = P(A) * P(B^c)$
  - Applicable for all possible combinations of events
- **Important rules:**
  - $P(A\ u\ B) = P(A) + P(B) - P(A\ n\ B)$
  - $P(A\ n\ B) = P(A|B) * P(B) = P(B|A) * P(A)$
  - $P(A^c) = 1 - P(A)$
  - $P(A) = P(A\ n\ B) + P(A\ n\ B^c)$
- Bayes Theorem is used to determine the probability of the initial event A given the subsequent event B has occurred
  - Putting the above rules together
  - $P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A) * P(A)}{P(B\ n\ A) + P(B\ n\ A^c)}$

Combinatorics (Methods of counting)

- Method of counting the total number of ways an event could occur
- Use Permutation and Combinations to count the number of ways for a single outcome
- Combine with the rule of product and rule of sum to count the number of ways for an Event (Collection of outcomes)

Rule of Product:	Rule of Sum:
• When an event E can be split into two ordered staged $E_1 \& E_2$ • $n\ ways\ to\ do\ E_1\ and\ m\ ways\ to\ do\ E_2$ • Total ways $= n * m * \dots k$	• When there are two non-overlapping events $E_1 \& E_2$ • $n\ ways\ to\ do\ E_1\ and\ m\ ways\ to\ do\ E_2$ • Total ways $= n + m + \cdots k$

Combination:

- Method to counting the number of ways an object can be chosen from a pool to form a group
- Objects MUST be of the SAME kind

Important caveats:

- **Forming ONE group of the SAME object**
    - Use the original combination formula
    - $nC_k = \frac{n!}{(n-k)! * k!}$
- **Forming ONE group made of DIFFERENT objects**
    - Use different combination formulas for each kind of object
    - $AC_a * BC_b * DC_d * \dots$
- **Forming MULTIPLE groups of the SAME/DIFFERENT object**
    - Need to change the total number of objects left in the pool to reflect the decreasing options for each subsequent group chosen
    - $n_1C_{k_1} * n_2C_{k_2} * n_3C_{k_3} * \dots (Where\ n_1 > n_2 > n_3)$
    - Same principle applies to multiple groups of different objects
- **Forming MULTIPLE groups with REPLACEMENT/REPITION ALLOWED**
    - NO need to change total number of objects left in the pool because no change
    - $(Group\ Size)^{Number\ of\ groups}$
- **Forming MULTIPLE groups of the SAME SIZE**
    - Need to divide by r! where r is the number of same sized groups
    - OR when choosing individual positions in a group (Don't care about order)
    - To avoid double counting cases: A-B and B-A are the same combination

Important problem-solving techniques:

- **Always apply restrictions first**
    - Assume given restriction is already in place
    - Must Exclude → Remove object from pool of choose-able items
    - Must Include → Assume that object is already chosen
- **Consider different cases to achieve event → Simplifies the problem**

Permutation:

- Method to count the number of ways to arrange distinct objects
- Objects must of the same kind but distinct

Important caveats:

- **Arranging objects in a straight line**
  - Consider number of possible options for each slots
  - $n!$
- **Arranging more objects than slots**
  - Choose which objects to get slotted in then arrange them
  - $nC_k * k!$
- **Arranging NON-distinct objects**
  - Need to divide by q! where q is the number of non-distinct objects being arranged
  - To avoid double counting cases: Non-distinct objects make no difference to arrangement

Problem-solving techniques:

- **Objects MUST be together**
  - Treat the objects as one unit & arrange them as per normal
  - Arrange them within their own unit as well
- **Objects MUST NOT be together**
  - Remove affected objects from the pool and arrange the rest
  - Consider the spaces/gaps next to each object $(n + 1)$
  - Choose which slots the affected objects will go into
  - Arrange the affected objects
- **Objects are alternating**
  - Special case of the previous scenario where now ALL gaps will be used (No need to choose)
  - Arrange the first set of objects
  - Arrange the second set of objects
  - Multiply by 2 to account for the front-back case (Additional gap)
- **Objects MUST be in between:**
  - Fix two ends points & consider the number of cases where the end points could be
  - Arrange the two end points
  - Arrange the objects between the end points
  - Arrange the objects outside of the end points
  - Multiply by the number of cases

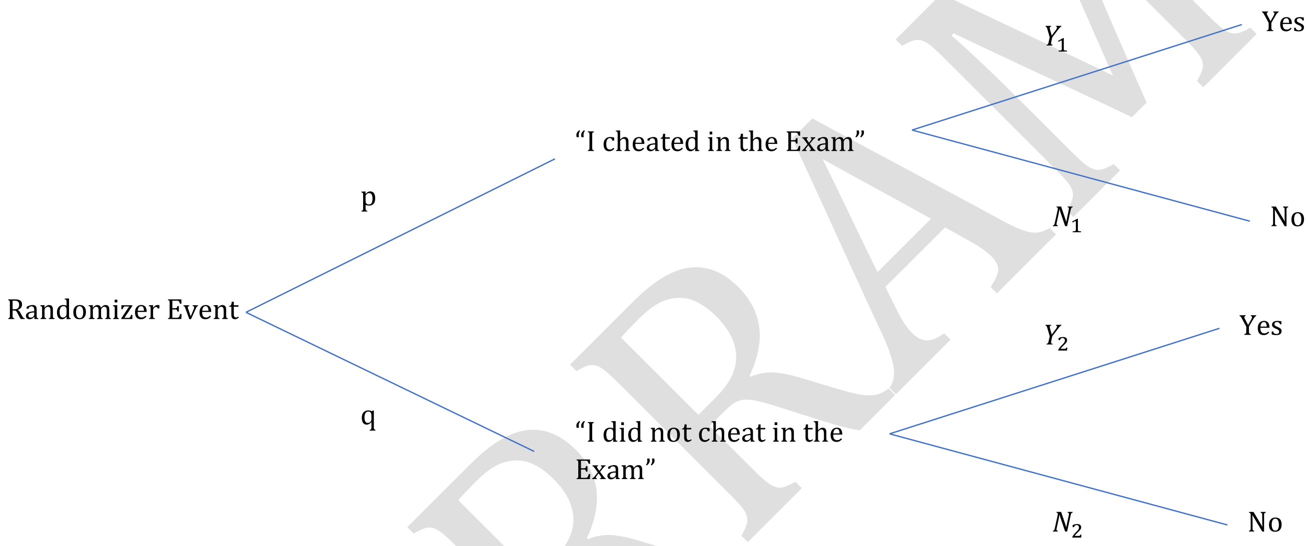
General problem-solving approaches:

- Splitting into different cases OR using the complementary event
- Visualising the problem: Venn Diagram, Tree Diagram & Table of Outcomes

Week 2: Practical Applications I

Randomized Response Trial

- People are not willing to answer sensitive questions truthfully → Biased results
- Can reduce the extent of this bias by assuring them of anonymity
- Randomised the question the answer depending on a random event
  - The two questions are mutually exclusive
    - Yes to type A question implies No to type B question
  - Researcher only receives arbitrary yes & no → No way to tell what your answer means as they don't know which question you answered
  - But able to determine the overall population's response



$$P(Cheated\ in\ the\ exam) = p * Y_1 + q * N_2$$

Hardy-Weinberg Equilibrium

- Proportion of Blood types among the population remain the same across generations
- Used as a means to determine the blood type of a potential offspring

Understanding blood types:

- Made up of Alleles: A, B and O
- Alleles can be either Dominant or Recessive → Blood type follows the Dominant one
- Each blood type is made up of TWO alleles:
  - Type A → AA or AO
  - Type B → BB or BO
  - Type AB → AB (Only one option because both A and B are dominant)
  - Type O → OO (Only one option because it is recessive)

Understanding the theorem:

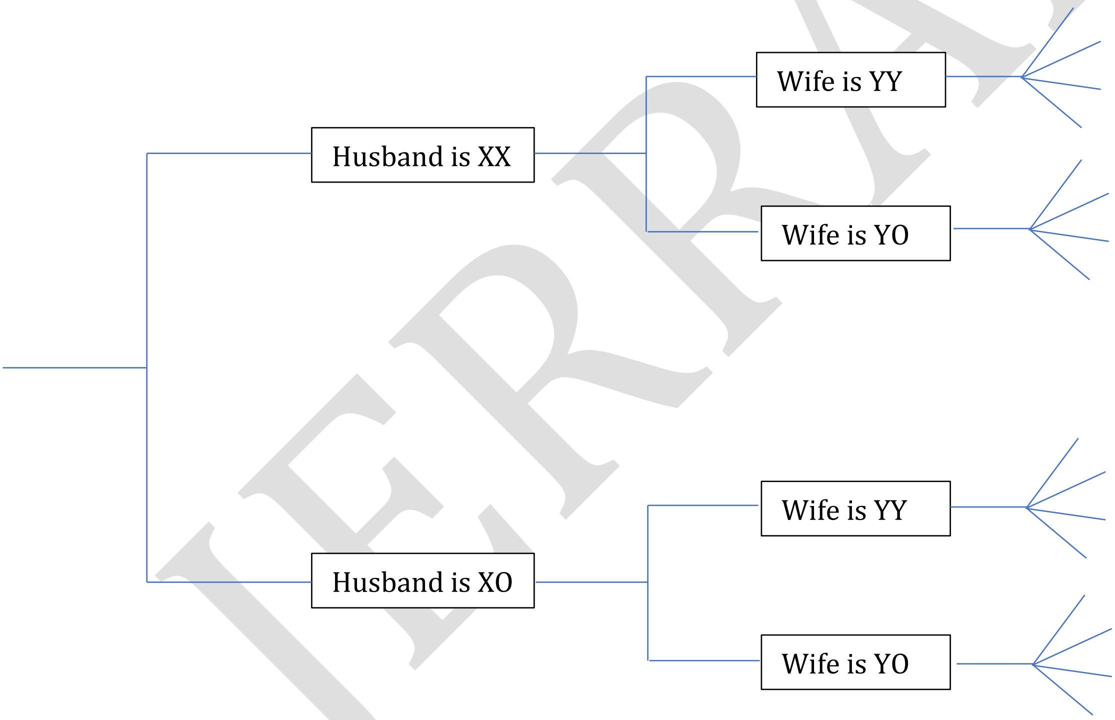
- $(A + B + O)^2 = 1$
  - A, B and O represent the frequency at which these alleles appear in the population
- $A^2 + 2AO + B^2 + 2BO + AB + O^2 = 1$
  - Proportion of people with Blood Type A → $A^2 + 2AO$
  - Proportion of people with Blood Type B → $B^2 + 2BO$
  - Proportion of people with Blood Type AB → AB
  - Proportion of people with Blood Type O → O^2
- Know which probability the question gives you → Allele or Blood type
- Solve for the frequency of each Allele

Determining the blood type of offspring:

- Child takes one allele from each parent
  - Always has 4 possible outcomes
  - Some overlapping so should account for that (Increased probability)
- Consider which combination of parent blood types will lead to the desired blood type
- Of these combinations consider the cases for which each parent is that blood type:
  - Case 1: Husband is XX, Wife is YY
  - Case 2: Husband is XX, Wife is YO
  - Case 3: Husband is XO, Wife is YY
  - Case 4: Husband is XO, Wife is YO
- Add together all the cases to get the probability is of a certain blood type
- Consider a complicated probability tree to best visualise the phenomenon

Complicated probability tree

- Can be split into two in the middle for easier viewing
- If the blood type of one parent is known then we can ignore the other parts of the probability tree
    - If nothing is stated → Calculate the probability of blood type based on Alleles ($A^2, 2AO\ etc$)
    - Most of the time it will be given that they are of a certain blood type but unsure of which allele combination
    - Calculate the probability that their child is that blood type using the tree than use conditional probability formula (Best to use given probabilities)
- The child has an equal likelihood of being any allele combination in the last column
    - If there are overlapping combinations → Remove the additional branch and increase the probability of occurrence
    - If there are multiple children involved just multiply by this column again



Race betting

- Stakes → How much money a gambler is willing to bet they the racer wins
  - If the racer loses → Gambler loses the amount of bet
  - If the racer wins → Gambler receives back more money
    - $Payoff = Stakes * Payoff\ odds$
- Payoff odds → Multiplier for how much more money they get if they win
  - Depends on the relative confidence of the gamblers on who will win
  - Can tell the confidence by comparing the quantity of bets on each racer
  - $Payoff\ odds = \frac{Total\ REMAINING\ stakes\ for\ ALL\ racers}{Total\ stakes\ for\ THAT\ racer}$
- Betting companies ALWAYS takes a share of the total stakes to ensure they earn money
  - Usually upwards of 10%
  - Numerator for Payoff odds thus takes this into account → Remaining money
  - Denominator for Payoff odds uses original values for each racer
- Applications:
  - Determining how many % profit the company should take
  - Determining payoffs based on bets
  - Determining how to bet based on payoff etc

TOTO Lottery

- Each person buys a ticket with 6 numbers from either 0 -49
- SG pools will draw 6 winning numbers
- AND one bonus number not from the above 6 numbers
- Based on the number of matches people get with the above numbers they win different prizes (Table below)

$$P(Win\ particular\ prize) = \frac{6C_{Number\ of\ matches\ for\ prize} * 43C_{Number\ of\ non-matches\ for\ prize}}{49C_6}$$

MULTIPLIED BY DEPENDING ON THE CASE:

Bonus number matches

- Probability that one of your non-matching numbers matches the bonus number

$$* \frac{Number\ of\ non-matches\ for\ prize}{43}$$

Bonus number does NOT match

- Probability that NONE of your non-matching numbers matched the bonus number

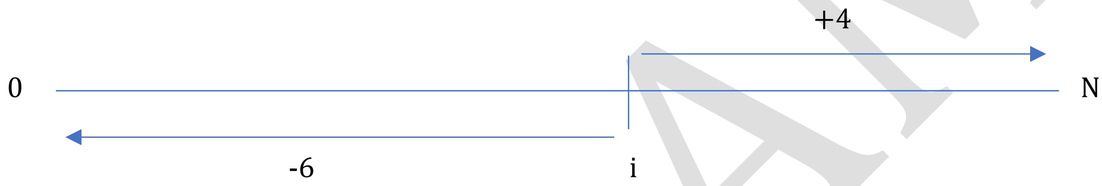
$$* \frac{43 - Number\ of\ non-matches\ for\ prize}{43}$$

Toto winning combinations:

Prize Group	Matches	Prize
1 (Jackpot)	6 numbers	38% of prize pool
2	5 numbers plus the additional number	8% of prize pool
3	5 numbers	5.5% of prize pool
4	4 numbers plus the additional number	3% of prize pool
5	4 numbers	$50 per winning combination
6	3 numbers plus the additional number	$25 per winning combination
7	3 numbers	$10 per winning combination

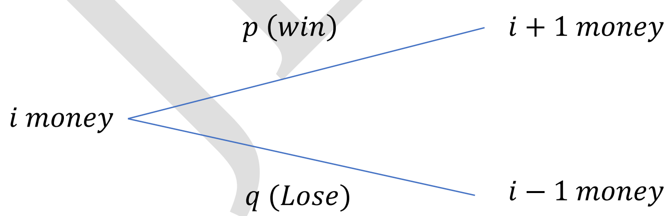
Gambler's Ruin Problem:

- Consider a simple two-player game
  - You start with i amount of money
  - Opponent has N-i amount of money
- Each round the same minigame is played
  - If you win (Probability p) → Your opponent gives you 1 dollar
  - If you lose (Probability q) → You give your opponent 1 dollar
  - Interested to find out who will win the overall game
- The overall game is won or lost when the other player reaches 0 (Bankrupt)
  - Variations where X amount of coins greater than the other
  - Can simulate these cases using the 0 bankrupt style as well
    - Player A wins if he gets 4 more coins than B
    - Player B wins if he gets 6 more coins than A
    - Consider player A:



- $N = 6 + 4 = 10$
- $i = 6 \;\&\; N - i = 4$
- ∴ Although A and B do not actually have 6 or 4 dollars, the probability that either one wins can be calculated via simulating this case of 6 & 4

Consider the probability tree for each minigame:



$Let\ P_i\ be\ the\ probability\ of\ winning\ the\ overall\ game\ given\ you\ start\ with\ i\ amount\ of\ money.$

Based on the above tree,

$P_i = p * P_{i+1} + q * P_{i-1}$

$Since\ p\ and\ q\ are\ mutually\ exclusive,$

$p + q = 1$

$\therefore p * P_i + q * P_i = P_i$ (Multiplying by P_i on both sides)

We equate $P_i = P_i$,

$p * P_i + q * P_i + p * P_{i+1} + q * P_{i-1}$

$P_i + \frac{q}{p} * P_i = P_{i+1} + \frac{q}{p} * P_{i-1}$

$P_{i+1} - P_i = \frac{q}{p}\ (P_i - P_{i-1})$

Since the overall game is only won when one party is bankrupt, we are interested to find out how this probability changes as our income changes.

We consider a recurrence relation,

$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) = \frac{q}{p}(P_1), since\ P_0 = 0$

$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \frac{q}{p} * \left(\frac{q}{p} * P_1\right) = \left(\frac{q}{p}\right)^2 * P_1$

$P_4 - P_3 = \frac{q}{p}(P_3 - P_2) = \frac{q}{p} * \left[\left(\frac{q}{p}\right)^2 * P_1\right] = \left(\frac{q}{p}\right)^3 * P_1$

$\vdots$

$P_N - P_{N-1} = \frac{q}{p}\ (P_{N-1} - P_{N-2}) = \frac{q}{p}\left[\left(\frac{q}{p}\right)^{N-2} * P_1\right] = \left(\frac{q}{p}\right)^{N-1} * P_1$

We observe that it forms a geometric series with a factor of $\left(\frac{q}{p}\right)$. We consider the sum of the series,

$$P_i = P_1 * \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)}$$

We know that $P_N = 1$,

$$P_N = P_1 * \frac{1-\left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)}$$

$$1 = P_1 * \frac{1-\left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)}$$

$$P_1 = \frac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^N}$$

Important Result:

$P_0 = 0$ because you can't win if you have no money to even play the game.

$P_N = 1$ because you instantly win if you have all the money.

We combine the above 2 equations,

$$P_i = P_1 * \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)}$$

$$P_i = \frac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^N} * \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)}$$

$$P_i = \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N}$$

Conclusions:

- Probability of winning dependent on two factors:
    - Fairness of the game $\left(\frac{q}{p}\right)$ → Less likely if the game is unfair
    - Relative starting incomes (i and N) → Less likely if opponent has more money
- In reality,
    - Having X more money than you ↔ You losing X more money
    - At what point do you give up on playing
    - In Casinos, most games favour the house AND the house has infinitely more money than you
    - By this theory, the probability of winning in a Casino is extremely low
    - Thus, all gamblers are destined to ruin, hence Gambler's Ruin

Matching Problem

- In a room of N people with their own unique hat
- Hats are mixed and everyone picks one hat
- Interested to determine the probability of n people finding their own hat
    - Number of ways to choose the n people
        - $\binom{N}{n} = \frac{N!}{n! * (N-n)!}$
    - Number of ways n people can find their own hat
        - Only one way to find your own hat
        - For n people → $1 * 1 * 1 \ldots (n\ times)$
    - Number of ways remaining people CAN'T find their own hats
        - $(N-n) * (N-n-1) * \ldots = (N-n)!$
    - Total number of ways people can find a hat $= N!$

$$\therefore P(n\ people\ get\ back\ their\ own\ hats) = \frac{N!}{n! * (N-n)!} * \frac{(N-n)!}{N!} = \frac{1}{n!}$$

Birthday Problem

<u>**Week 3: Discrete Random Variables**</u>

- Random Variables are a function of the outcome mapped onto the sample space instead of the actual outcome itself
    - Discrete Random Variables are variables that can only take on **countable** values
    - Continuous Random Variables are variables that can take on a range of values
- Distribution of a Random Variable is a function that describes all possible values a Random Variable could take & its associated probability
    - **Probability Mass Function (PMF)**
        - Table with all the possible values the random variable can take & its associated probability
        - Function that gives the probability that a random variable is exactly equal to some value
            - $P(X = a) = p(a)$
            - $\sum_{min}^{max} p(a) = 1$
    - **Cumulative Distribution Function (CDF)**
        - Piecewise Step Function (Sum of all previous probabilities up till that point)
        - Function that gives the probability that a random variable is less than or equal to some value
            - $F(a) = P(X \leq a) = \Sigma p(x)$

Moments of a discrete random variable

- The r-th moment of a random variable is the expected value of X^r
- **Expectation (First moment)**
    - Weighted average of all possible values that X can take based on its probability
    - $E(X) = \Sigma\big(x_i * p(x_i)\big)$
    - $E(aX + b) = aE(X) + b$
- **Variance (Second moment)**
    - Measure of how far apart the values of X are from its mean
    - $Var\ (X) = E(X^2) - [E(X)]^2$
    - $Var\ (aX + b) = a^2\ Var\ (X) + 0$
    - $SD\ (X) = \sqrt{Var\ (X)}$
        - SD is preferred as it is the same units as the random variable
        - If there is only one possible value or a constant the $Var\ (X) = SD = 0$

Commonly known Discrete Distributions

- Mostly based off the **Bernoulli Distribution**
- Single trial whose outcome can only be success (1) or Failure (0)
- k is ALWAYS the variable we are interested in finding
- $X \sim Bernoulli\ (p)$
  - p is the probability of success
  - $P(X = k) = p^k(1-p)^{1-k}$
  - $E(X) = p$
  - $Var\ (X) = p(1-p)$

Binomial Distribution

- Random variable measures the **number of successes** in n independent Bernoulli trials
- Sum of n independent Bernoulli variables
- $X \sim Bin(n, p)$
  - n is the number of trials & p is the probability of success
  - $P(X = k) = \binom{n}{k} p^k(1-p)^{n-k}$
  - $E(X) = np$
  - $Var\ (X) = np(1-p)$
- Sum of two independent Binomials with same parameter $\rightarrow X + Y \sim Bin\ (n+m, p)$
- If parameter p is different, the sum is NOT binomial

Geometric Distribution

- Random variable measures the **number of independent Bernoulli trials** required till the 1st success is obtained
- $X \sim Geom(p)$
  - p is the probability of success
  - $P(X = k) = (1-p)^{k-1}p$
  - $E(X) = \frac{1}{p}$
  - $Var(X) = \frac{1-p}{p^2}$

Negative Binomial Distribution

- Random variable measures the **number of independent Bernoulli trials** required till a total of r successes occur
- Sum of r independent Geometric variables
- $X \sim NB\ (r, p)$
  - r is the number of successes we want to observe & p is the probability of success
  - $P(X = k) = \binom{k-1}{r-1} p^r(1-p)^{k-r}$
  - $E(X) = \frac{r}{p}$
  - $Var\ (X) = \frac{r(1-p)}{p^2}$

Hypergeometric distribution

- Random variable measures the **number of successes in n NON-INDEPENDENT Bernoulli trials WITHOUT REPLACEMENT**
- $X \sim Hypergeometric\ (N, m, n)$
  - N is the total number of objects, n the number of objects sampled & m type of objects we are interested in
  - $P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$
  - $E(X) = \frac{nm}{N}$
  - $Var\ (X) = \left(\frac{N-n}{N-1}\right) n \left(\frac{m}{n}\right)\left(1 - \frac{m}{n}\right)$
- For the multivariate case, simply add more $\binom{m}{k}$ for each unique object

Poisson Distribution (Non-Bernoulli related)

- Random variable that measures the **number of events that occur over a given time-space interval**
- Assume that the events are random & are independent across non-overlapping time-space intervals
- $X \sim Poisson\ (\mu)$
  - μ is the expected number of events per unit time-space interval (Can scale)
  - $P(X = k) = \frac{e^{-\mu}\mu^{x}}{x!}$
  - $E(X) = \mu$
  - $Var\ (X) = \mu$
- Sum of two independent Poisson → $X + Y \sim Poisson\ (\mu_X + \mu_Y)$

Important caveats:

- Both the above distributions have an approximation relating to the binomial distribution
- Hypergeometric → Binomial when sampling within under 5% of the population
  - Dependency between trials are small & can be ignored
  - ∴ Both will produce the same answer
- Binomial → Poisson for large n and small p
  - $\mu = np$ remains constant
  - $As\ n \to \infty, then\ the\ PMF\ of\ the\ Binomial \to Poisson$

Week 4: Continuous Random Variable

- Random variable that can assume any of the values in some interval
  - Continuous scale → Assign 0 probability to any individual value
  - Does NOT mean that the individual value I occur
- Distribution of a Random Variable is a function that describes all the values a Random Variable could take & its associated probability
  - **Probability Density Function (PDF)**
    - Function that describes the distribution of variables NOT their probability
    - Area under an interval is the probability of RV taking a value within it
      - $P(a \leq X \leq b) = \int_a^b f(x)\, dx$
      - $\int_{min}^{max} f(x)\, dx = 1$
  - **Cumulative Distribution Function (CDF)**
    - Function that gives the probability that a random variable is less than or equal to some value
      - $P(X \leq b) = F(a) = \int_0^b f(x)$
      - $F'(a) = f(a)$

Moments of a continuous random variable (Similar to discrete)

- The r-th moment of a random variable is the expected value of X^r
- **Expectation (First moment)**
  - Weighted average of all possible values that X can take based on its probability
  - $E(X) = \int_a^b x * f(x)\ dx$
  - $E(aX + b) = aE(X) + b$
- **Variance (Second moment)**
  - Measure of how far apart the values of X are from its mean
  - $Var\ (X) = E(X^2) - [E(X)]^2$
  - $Var\ (aX + b) = a^2\ Var\ (X) + 0$
  - $SD\ (X) = \sqrt{Var\ (X)}$
    - SD is preferred as it is the same units as the random variable
    - If there is only one possible value or a constant the $Var\ (X) = SD = 0$

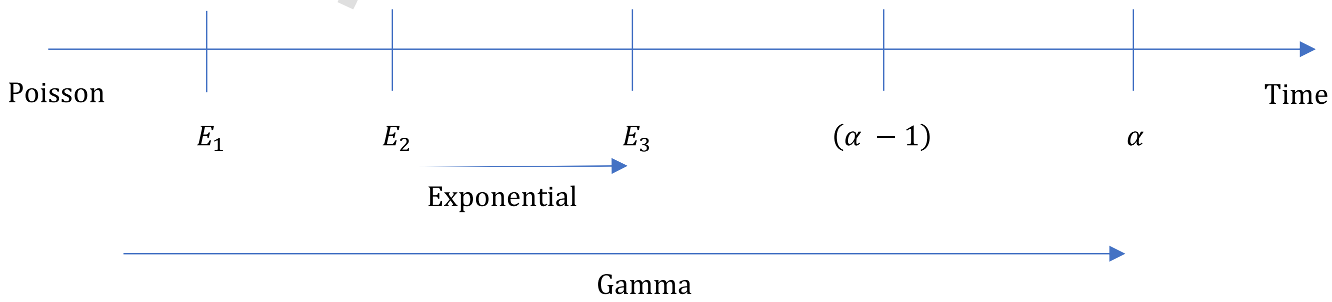
Commonly known distributions

Exponential Distribution (Poisson related)

- Random Variable that measures the waiting time for an event to occur in a Poisson process
- $X \sim \exp(\lambda)$
  - λ is the rate at which events occur $\rightarrow \lambda = \frac{1}{\mu}$
  - $f(x) = \begin{Bmatrix} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{Bmatrix}$
  - $F(x) = \begin{Bmatrix} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & otherwise \end{Bmatrix}$
  - $E(X) = \frac{1}{\lambda}$
  - $Var\ (X) = \frac{1}{\lambda^2}$
- **Memoryless property**
  - Probability of the waiting time between events does not change
  - Does not care how long you have waited in the past; Count from now
  - $P(X > a + b \mid x > b) = P(X > a)$

Gamma Distribution (Poisson related)

- Random variable that measures the waiting time till the $\alpha - th$ event to occur in a Poisson process
- Sum of α independent Exponential Variables
- $X \sim Gamma\ (\alpha, \lambda)$
  - α is the number of events we want to observe & λ is the rate at which events occur
  - $f(x) = \begin{Bmatrix} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & otherwise \end{Bmatrix}, where\ \Gamma(\alpha) = (\alpha - 1)!$
  - $E(X) = \frac{\alpha}{\lambda}$
  - $Var\ (X) = \frac{\alpha}{\lambda^2}$
- **Estimate $F(x)$ using Poisson Distribution**:
  - CDF is hard to evaluate $\rightarrow$ Rethink using Discrete Poisson distribution
  - Consider the probability of waiting time k for the $\alpha - th$ event to occur
  - Before time k, only $(\alpha - 1)$ events could have occurred $\rightarrow$ Poisson distribution
  - Find the expected number events for timespan $k \rightarrow \mu = \lambda * k$
  - $\therefore F(x) = P\big(W \leq (\alpha - 1)\big), where\ W \sim Poisson\ (\mu)$



<u>Prior Distributions:</u>

- Distribution of an unknown parameter/variable that would best express the known characteristics of the variable
- Consider what you can safely guess about the variable you seek to model then choose a distribution that best encapsulates it

Uniform distribution

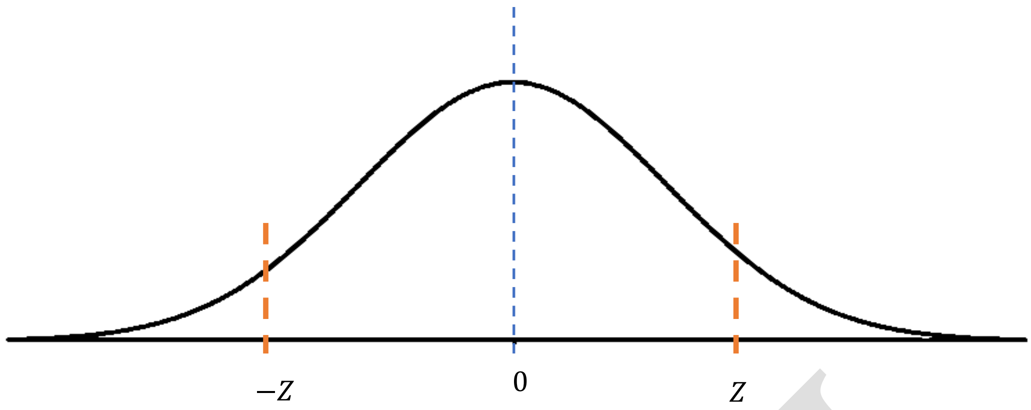
- Used to show the distribution of a variable that has to occur within two bounds and is equally likely to occur within the interval
- $X \sim Uni\ (a, b)$
    - $a\ and\ b$ are the upper and lower values the variable can take
    - $f(x) = \begin{Bmatrix} \frac{1}{b-a} & a \leq x \leq b \\ 0 & otherwise \end{Bmatrix}$
    - $F(x) = \begin{Bmatrix} \frac{x-a}{b-a} & a \leq x \leq b \\ 0 & x < a \\ 1 & x > b \end{Bmatrix}$
    - $E(X) = \frac{a+b}{2}$
    - $Var\ (X) = \frac{(b-a)^2}{12}$

Beta Distribution

- Used to show the distribution of a variable that is itself a probability
- $X \sim Beta(a, b)$
    - $f(x) = \begin{Bmatrix} \frac{\Gamma(a+b)}{\Gamma(a)*\Gamma(b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & Otherwise \end{Bmatrix}, where\ \Gamma(a) = (a-1)!$
    - $F(x)$ is too complicated to evaluate
    - $E(X) = \frac{a}{a+b}$
    - $Var\ (X) = \frac{ab}{(a+b)^2\ (a+b+1)}$

Normal Distribution

- Used to show the distribution of a variable with a known mean and variance.
- $X \sim N(\mu, \sigma^2)$
    - $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
    - $E(X) = \mu$
    - $Var\ (X) = \sigma^2$
- Estimate $F(x)$ using the Standard Normal Distribution $\Phi(Z)$
    - Consider the CDF $\rightarrow P(X \leq \alpha)$
    - Convert to the Standard Normal Distribution $\rightarrow Z = \frac{x-\mu}{\sigma}$
    - Refer to the standard normal table to obtain the probability $\rightarrow P\left(Z \leq \frac{\alpha-\mu}{\sigma}\right)$
- Sum of two independent normal $\rightarrow X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$



Reading the standard normal table:

- Standard normal table gives the CDF for POSITIVE variable Z
- Value on table gives the total area BEHIND the POSITIVE variable Z
    - $P(Z \leq A) = Refer\ to\ the\ table\ directly$
    - $P(Z \geq A) = 1 - P(Z \leq A)$
    - $P(Z \leq -A) = P(Z \geq A) = 1 - P(Z \leq A)$
    - $P(Z \geq -A) = P(Z \leq A) = Refer\ to\ the\ table\ directly$
- Other important interpretations:
    - Distribution stretches infinitely from $(-\infty . \infty)$
    - Distribution is symmetrical about μ
    - Mean = Mode = Median = μ

Finding the PDF of composite Variables $(\boldsymbol{Y = g(X)})$

- **Distribution Function Technique**
    - Consider the CDF of Y → $P(Y \leq y)$
    - Substitute in PDF of X → $P(g(X) \leq y)$
    - Make X the subject → $P\left(X \leq g^{-1}(y)\right)$
    - Use the CDF of X → $\int_0^{g^{-1}(y)} f(x)\, dx$
    - Differentiate with respect to Y → $f(y) = F'(X)$
- **Transformation Technique**
    - Check that the function is one to one → Substitute in range of X into Y
    - Make X the subject → Express X in terms of new variable
    - Differentiate with respect to Y → $\frac{dx}{dy}$
    - Use the formula → $f(y) = f_x(Substitute\ X\ with\ the\ new\ variable) * \frac{dx}{dy}$
    - $\therefore f(y) = \begin{Bmatrix} h(y) & a < y < b \\ 0 & otherwise \end{Bmatrix}$
    - Range can either be $<$ or $\leq$ → Follow what was originally given

Problem-solving:

- Find all unknown constants in the question → Use known identities to form equations
- Always simplify expressions (Avoid fractions) for easier integration & differentiation
- Composite variable → Think of cases regarding the base variable first & how it affects

Week 5: Joint Distribution

- Probability statements concerning two or more random variables
- Joint range is very important:
  - Each variable is defined on their own range
  - But the joint function only exists in the area in which the two ranges coincide
  - Need to identify the range for each variable where the joint function exists
- Marginal distributions
  - Distribution of each variable on its own within this joint range
  - Consider all the values the variable can take for all the values of the other variable

Discrete Joint Distributions:

- **Consider the Joint range:**
  - Imagine the experiment being carried out
  - What are all the possible outcomes?
  - For these outcomes, what are all the possible values of X and Y? (Range)
- **Joint Probability Mass Function**
  - Best described using a Contingency Table
  - Each individual $P(X_i, Y_i)$ should be evaluated individually using P&C or known distributions

	$\boldsymbol{Y_1}$	$\boldsymbol{Y_2}$	$\boldsymbol{Y_3}$	$\boldsymbol{P(Y)}$
$\boldsymbol{X_1}$	$P(X_1, Y_1)$	$P(X_1, Y_2)$	$P(X_1, Y_3)$	Sum of Row
$\boldsymbol{X_2}$	$P(X_2, Y_1)$	$P(X_2, Y_2)$	$P(X_2, Y_3)$	Sum of Row
$\boldsymbol{X_3}$	$P(X_3, Y_1)$	$P(X_3, Y_2)$	$P(X_3, Y_3)$	Sum of Row
$\boldsymbol{P(Y)}$	Sum of Column	Sum of Column	Sum of Column	

- **Marginal Distributions**
  - Obtained by summing the entire Row/Column for that variable
  - Can be presented on its own table as well for clarity

	$\boldsymbol{X_1 \text{ or } Y_1}$	$\boldsymbol{X_2 \text{ or } Y_2}$	$\boldsymbol{X_3 \text{ or } Y_3}$
$\boldsymbol{P(X) \text{ or } P(Y)}$	Sum of Row/Column 1	Sum of Row/Column 2	Sum of Row/Column 3

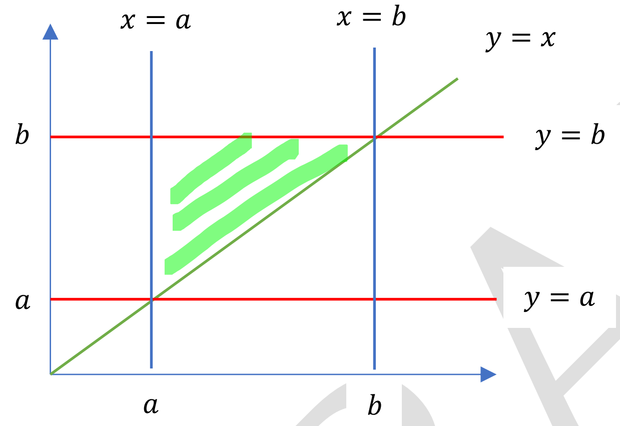
- **Conditional distributions**
  - Distribution of one variable given the value of the other
  - $P(X = x_k | Y = y) = \frac{P(X=x_k, Y=y)}{P(Y=y)}$
  - Usually only required to evaluate but can be presented in table as well

	$\boldsymbol{X_1 \| Y = y}$	$\boldsymbol{X_2 \| Y = y}$	$\boldsymbol{X_3 \| Y = y}$
$\boldsymbol{P(X \| Y = y)}$	$P(X = x_1 \| Y = y)$	$P(X = x_2 \| Y = y)$	$P(X = x_3 \| Y = y)$

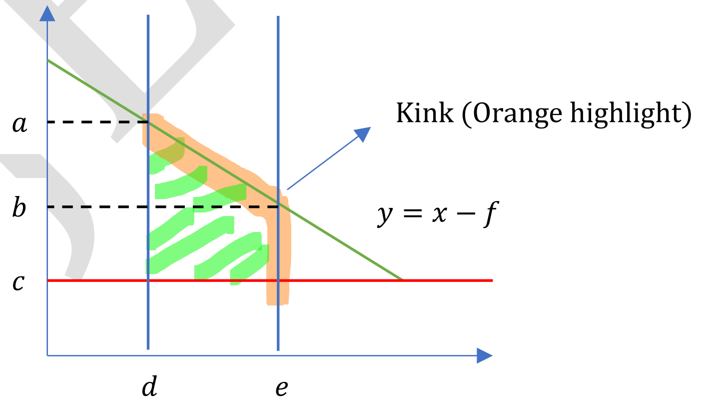
- **Independence**
  - Although two variables are jointly defined, they can be independent
  - Independent if $P(x, y) = p(x) * p(y)$ within their joint range
  - IF the given range already has the other variable in it → Must be considered together and hence NOT independent

Continuous Joint Distribution

- Consider the Joint range (Graphical Method)
  - Split the range into 2-3 different components:
    - Just X → $a < x < b$ $(Blue)$
    - Just Y → $a < y < b$ (Red)
    - Interaction → $x < y$ $(Green)$
  - Visualise using a graph
    - Replace all inequality signs with equal signs to know what to draw
    - Don't need to be drawn to scale but relative shape must be correct
    - Intersection points & positions should be considered carefully



  - Find the area where it is defined on
    - Consider the inequalities once again
    - Shade the area where the function is defined on
  - Conclude the range of each variable
    - Can be defined in two ways (Depending on which variable you start with)
    - If range reaches a function, use the equation of the function as the limit
    - Remember to make x and y the subject accordimgly
    - $Combination\ 1 = \{a < x < b\ and\ x < y < b\}$
    - $Combination\ 2 = \{a < y < b\ and\ a < x < y\}$
  - Important: Splitting the range

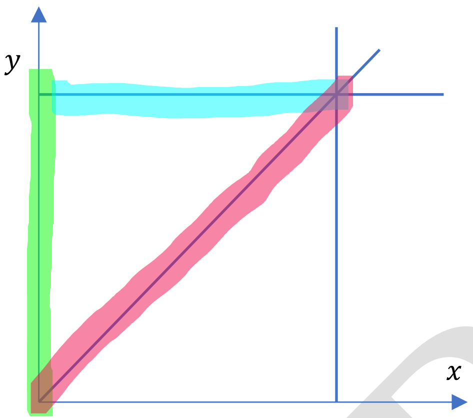


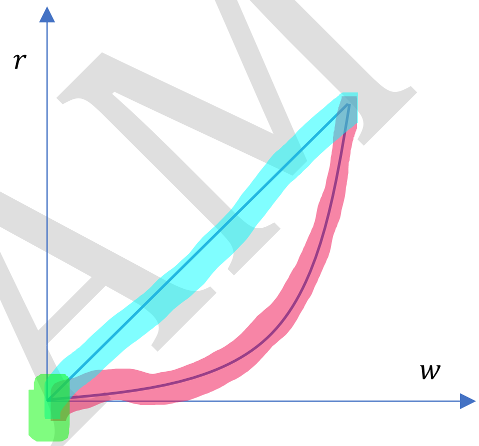
    - Kink in the defined area → Cannot define with one expression
    - Split into two areas without kinks (Think of normal basic shapes)
    - $Combination\ 1 = \begin{Bmatrix} d < x < e\ and\ c < y < b \\ d < x < e\ and\ b < y < x - f \end{Bmatrix}$
    - $Combination\ 2 = \begin{Bmatrix} c < y < b\ and\ d < x < e \\ b < y < a\ and\ d < x < y + f \end{Bmatrix}$

- **Joint Probability Density Function**
    - Similar to single variable case → Need to integrate to obtain probability
    - Since two variables need to perform DOUBLE integration
    - **Determine the range of the integration**
        - Consider the inequality $P(X < a, Y < b)$
        - Draw the inequality as functions → Replace with equal signs
        - Determine the range which all the inequalities are fulfilled
            - Consider intersection points
            - Consider if the area is kinked → Split into two
        - Two combinations available → Choose what is convenient
    - **Determine the order of the integration**
        - Double integration means the inner variable is integrated first
        - Always integrate the variable whose range contains algebra FIRST
        - This is so that the final product of the integration will be a whole number
- **Marginal Distribution**
    - $f(x) = \int_a^b f(x,y)\, dy \ where\ a < y < b$
    - Use the combination range where you fixed the variable you are finding (x) first
    - $f(x)$ is still a density function → Draw & Integrate to evaluate it
- **Conditional Distributions**
    - $f(x|y) = \frac{f(x,y)}{f(y)}\ for\ some\ range$
    - Draw the line $y = a\ or\ x = b$ on the graph
    - Determine the range of the other variable ON THIS LINE within the defined area
    - $f(x|y)$ is still a density function → Integrate to evaluate it
- **Independence**
    - Independent if $f(x,y) = f(x) * f(y)\ within\ their\ joint\ range$
    - IF the given range already has the other variable in it → Must be considered together and hence NOT independent
    - Useful for forming Joint PDFs → Assume independence and multiply together

Transformation technique for two variables:

- **Determine the range of the new variables**
  - Draw the range of the initial function > Determine the boundaries
  - Consider each boundary:
    - Should have one equation and ONE inequality
    - Substitute it into the new expressions
    - Convert all variables to the new ones where possible
    - Old boundaries become new boundaries (One equation one line)
    - Repeat for each of the old boundaries
    - Use graphical method to determine new range





***Example*: $R = XY$ and $W = X$**

- $\boldsymbol{Green - X = 0, 0 < Y < 1}$
  - $R = Y(0) = 0$
  - $W = 0$
  - $\therefore R = 0 \text{ and } W = 0\ (Origin)$
- $\boldsymbol{Blue - Y = 1, 0 < X < 1}$
  - $R = X(1) = X = W$
  - $W = X$
  - $\therefore R = W \text{ for } 0 < W < 1$
- $\boldsymbol{Red - Y = X, 0 < X < 1}$
  - $R = X(X) = X^2 = W^2$
  - $W = X$
  - $\therefore R = W^2 \text{ for } 0 < W < 1$

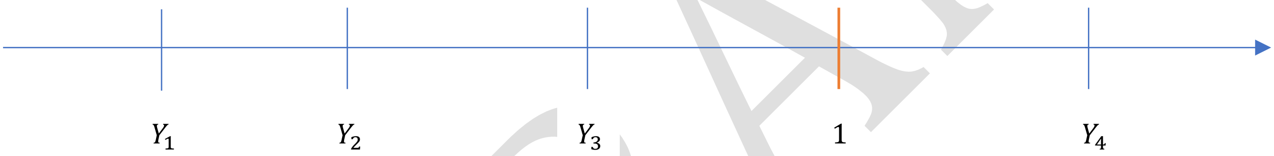
- **Consider the determinant of the Jacobian Matrix:**
  - Express x and y in terms of their new transformed variables
  - Consider the second order partial derivatives
  - $|J| = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial s} \end{vmatrix} = \left(\frac{\partial x}{\partial r} * \frac{\partial y}{\partial s}\right) - \left(\frac{\partial x}{\partial s} * \frac{\partial y}{\partial r}\right) \rightarrow Criss - cross\ method$
  - No need to follow order above but ensure each row is for a unique variable being differentiated and each column is with respect to a unique variable
- **Apply the formula:**
  - $f(r,s) = f_{x,y}(Substitute\ X\ and\ Y\ with\ the\ new\ variables) * |J|$
  - $\therefore f(r,s) = \begin{Bmatrix} h(r,s) & a < r < b \text{ and } c < s < d \\ 0 & otherwise \end{Bmatrix}$

Order statistics (Special transformation)

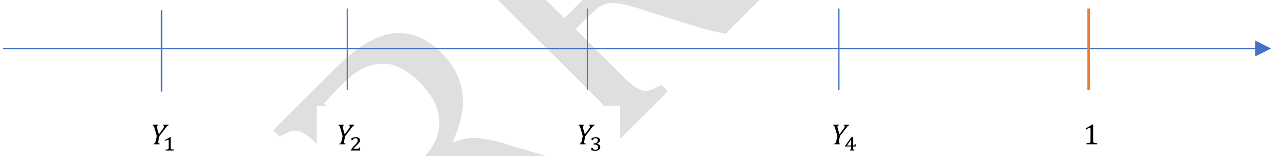
- Re-ordering the observed variables in ascending order
- Consider a sample $\{X_1, X_2, X_3 \dots X_n\}$
- Let $Y_1, Y_2 \dots Y_n$ be the random variable denoting the order statistics of the sample
  - Where Y_1 is the smallest of the sample $\{X_1, X_2, X_3 \dots X_n\}$
  - Where Y_n is the largest of the sample $\{X_1, X_2, X_3 \dots X_n\}$
- Problematic if the observations are the same size
  - Assume that X is a continuous variable
  - Probability that X is the same value is zero

Deriving the CDF and PDF of the Order Statistic RV

- Imagine there are 4 observations in the sample
- Order statistic $\rightarrow Y_1, Y_2 \dots Y_4$
- Consider $P(Y_3 \leq 1)$:
  - Case 1: $Y_1, Y_2, Y_3 \leq 1 \; and \; Y_4 > 1$



  - Case 2: $Y_1 . Y_2, Y_3 \; and \; Y_4 \leq 1$



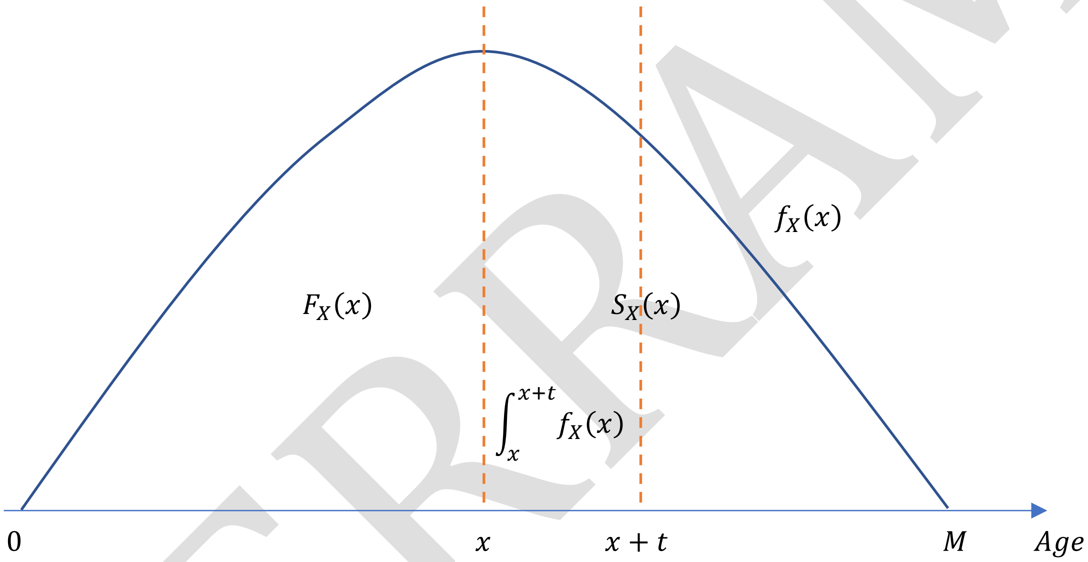
  - Let Z be the number of successes (≤ 1) out of 4 trials
    - $Z \sim Bin(n, p)$
    - n is the number of observations in the sample
    - p is the probability that $x_i \leq 1$ (Dependent on the distribution of X)
    - $\therefore P(Y_3 \leq 1) = P(Z = 3) + P(Z = 4)$
- **Consider the general CDF:**
  - $P(Y_k \leq y) = P(Z = k) + P(Z = k + 1) + \cdots P(Z = n)$
  - $G(Y_k) = \sum_k^n \binom{n}{k} p^k (1 - p)^{n-k}$
  - Where p is the probability that $X \leq y \rightarrow Dependent \; on \; CDF \; of \; X \rightarrow F(X)$
  - $\therefore G(Y_k) = \sum_k^n \binom{n}{k} * [F_X(y)]^k * [1 - F_X(y)]^{n-r}$
- **Consider the general PDF:**
  - Differentiate the CDF to obtain PDF (Complicated)
  - $g(y) = \frac{n!}{(k-1)!(n-k)!} * [F_X(y)]^{k-1} * [1 - F_X(y)]^{n-k} * f_X(y)$
  - Follow range of original function
  - Order statistic of a uniform distribution is a Beta distribution

Week 6: Practical Applications 2

Actuarial Survival Model

Unconditional Random Variable X

- Random Variable denoting the future lifetime (Time of death) a **new-born**
- Unconditional because a new-born has survived 0 years at time of calculation
- $f_X(x) \rightarrow \boldsymbol{Distribution}\ of\ lifetime\ (NOT\ probability)$
- $\int_x^{x+t} f_X(x) \rightarrow Death\ between\ time\ x\ and\ x+t$
- $F_X(x) \rightarrow Death\ by\ time\ x$
- $S_X(x) \rightarrow Survives\ by\ time\ x; Death\ after\ time\ x$
- Important points:
  - 0 is the time of birth of new-born
  - M is the maximum attainable age → Everyone will die by time M
  - $S_X(x) = 1 - F_X(x)$



Force of Mortality ($\boldsymbol{\mu_X}$)

- **Distribution** of the time of future lifetime given that it survives till time x (NOT probability)
- Variation of the above distribution given that we know it will survive till x

$$\mu_X(X) = \frac{f_X(x)}{S_X(x)}$$

$$\mu_X(X) = \frac{-S_X'(x)}{S_X(x)} \longrightarrow \quad f_X(x) = F_X'(x) = \frac{d}{dx}\left(1 - S_X(x)\right)$$

$$\therefore f_X(x) = -S_X'(x)$$

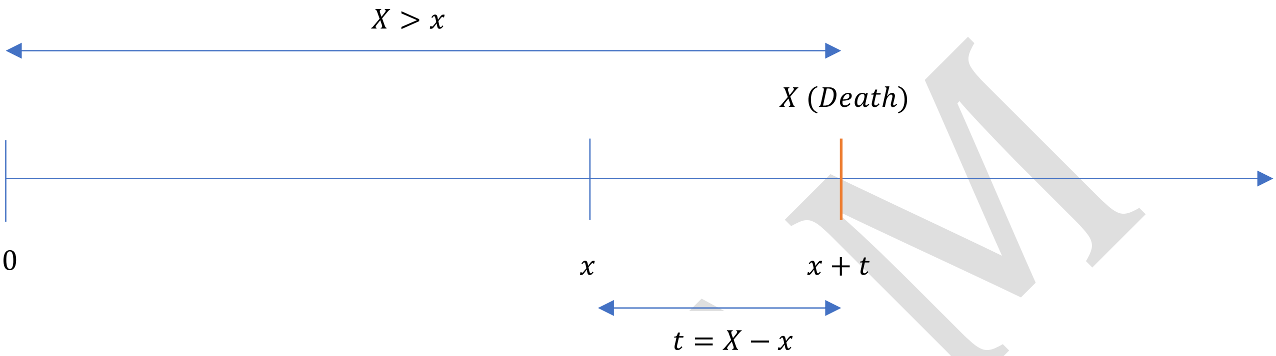
$$\mu_X(X) = -\frac{d}{dx}(\ln S_X(x))$$

$$\ln S_X(x) = -\int_0^x \mu_X(x)$$

$$\therefore S_X(x) = e^{-\int_0^x \mu_X(x)}$$

Conditional Random Variable T

- Random variable denoting the future lifetime (Time of death) of a person aged x
- Conditional because the person has survived x years at point of calculation
- $T = (X - x) \mid X > x$
- $f_T(t) \rightarrow \textbf{Distribution } of\ lifetime\ (NOT\ probability)$
- $F_T(t) \rightarrow Death\ within\ x\ and\ x + t\ years\ given\ survived\ till\ x$



Mathematical expressions:

$\boldsymbol{F_T(t)}$

$= P(T(x) \leq t)$

$= P(X - x \leq t \mid X > x)$

$= P(X \leq x + t \mid X > x)$

$= \dfrac{P(x \leq X \leq x + t)}{P(X > x)}$

$= \dfrac{P(X \leq x + t) - P(X \leq x)}{1 - P(X \leq x)}$

$= \dfrac{\boldsymbol{F_X(x + t) - F_X(x)}}{\boldsymbol{1 - F_X(x)}}$

$= \dfrac{F_X(x + t) - F_X(x)}{S(x)}$

$= \dfrac{1 - S_X(x + t) - (1 - S_X(x))}{S_X(x)}$

$= \dfrac{\boldsymbol{S_X(x) - S_X(x + t)}}{\boldsymbol{S_X(x)}}$

$= \boldsymbol{1 - \dfrac{S_X(x + t)}{S_X(x)}}$

$\boldsymbol{f_T(t)}$

$= \dfrac{d}{dt}\left(F_T(t)\right)$

$= \dfrac{d}{dt}\left(\boldsymbol{1 - \dfrac{S_X(x + t)}{S_X(x)}}\right)$

$= \dfrac{-S_X'(x + t)}{S_X(x)}$

$= \dfrac{\boldsymbol{f(x + t)}}{\boldsymbol{S_X(x)}}$

$Multiply\ by\ \dfrac{S_X(x + t)}{S_X(x + t)} = 1,$

$= \dfrac{f(x + t)}{S_X(x)} * \dfrac{S_X(x + t)}{S_X(x + t)}$

$= \dfrac{S_X(x + t)}{S_X(x)} * \dfrac{f(x + t)}{S_X(x + t)}$

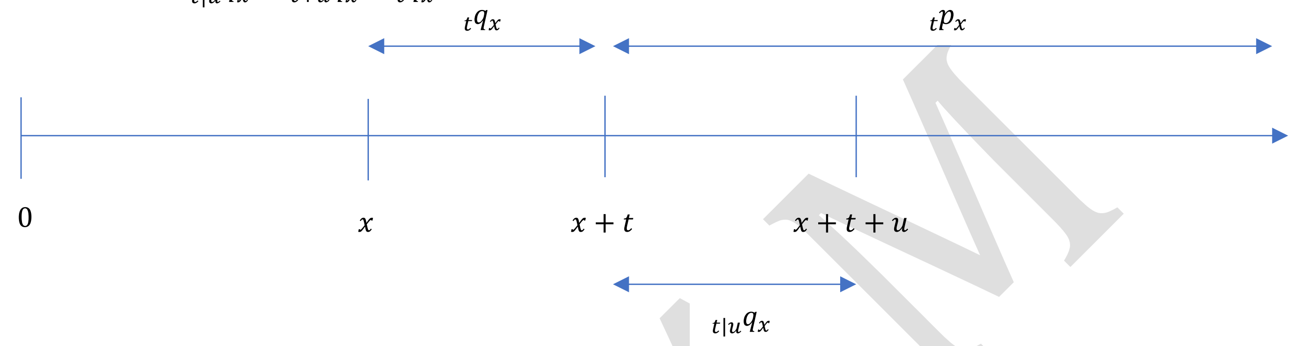
$= \boldsymbol{[1 - F_T(t)] * \mu_X(x + t)}$

$= \boldsymbol{{}_t p_x * \mu_X(x + t)}$

- Intuitive
- Survive till $x + t$ & the distribution of life from that point on

Special notation for conditional random variable T

- $_{t}q_{x} \rightarrow Death\ by\ time\ x+t$
- $_{t}p_{x} \rightarrow Survive\ till\ time\ x+t\ OR\ Death\ after\ time\ x+t$
- $_{t|u}q_{x} \rightarrow Death\ between\ time\ x+t\ \&\ x+t+u$
- Important points:
  - IF there is no subscript given → Assume the number is 1
  - $_{t}p_{x} = 1 - {}_{t}q_{x}$
  - $_{t|u}q_{x} = {}_{t+u}q_{x} - {}_{t}q_{x}$



Mathematical expressions:

$$_{t}q_{x} = F_T(t) = \frac{S_X(x) - S_X(x+t)}{S_X(x)}$$

$$_{t}p_{x} = 1 - F_T(t) = \frac{S_X(x+t)}{S_X(x)}$$

$$_{t|u}q_{x} = F_T(t+u) - F_T(t) = \frac{S_X(x+t) - S_X(x+t+u)}{S_X(x)}$$

$$If\ multiply\ by\ \frac{S_X(x+t)}{S_X(x+t)} = 1,$$

$$_{t|u}q_{x} = \frac{S_X(x+t) - S_X(x+t+u)}{S_X(x)} * \frac{S_X(x+t)}{S_X(x+t)}$$

$$_{t|u}q_{x} = \frac{S_X(x+t)}{S_X(x)} * \frac{S_X(x+t) - S_X(x+t+u)}{S_X(x+t)}$$

$$_{t|u}q_{x} = {}_{t}p_{x} * {}_{u}q_{x+t}$$

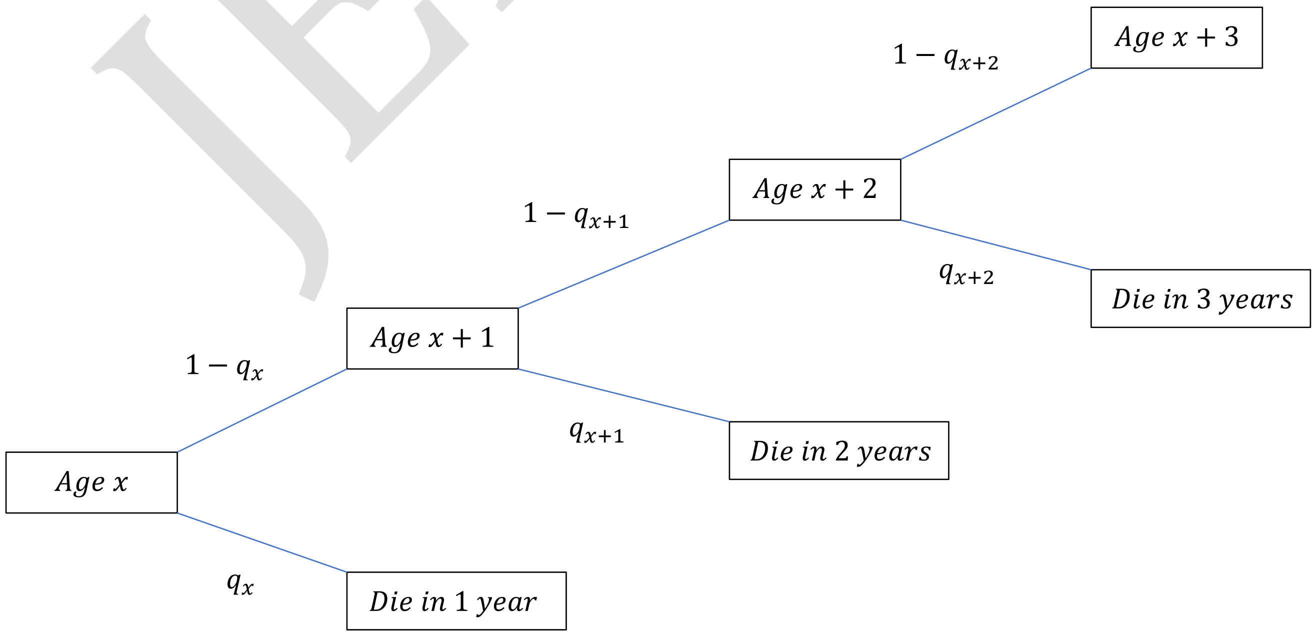
- Intuitive
- Survive till $x+t$ and die within next u years

Life table (L_x)

- Table showing the number of people alive at x years
  - $l_x \rightarrow Number\ of\ people\ alive\ at\ x\ year$
  - $l_{x+1} \rightarrow Number\ of\ people\ still\ alive\ after\ 1\ year$
  - $l_{x+2} \rightarrow Number\ of\ people\ still\ alive\ after\ 2\ year$
  - $\vdots$
- Number of people can be modelled using Binomial Distribution
  - $l_x \sim Bin\ (Number\ of\ people, Rate\ of\ survival)$
  - $l_x \sim Bin\left(l_0, S_X(x)\right)$
- Using life table to solve for important values:
  - Unconditional random variable X:
    - $S_X(x) = \frac{l_x}{l_0}$
    - $F_X(x) = 1 - \frac{l_x}{l_0}$
  - Conditional random variable T:
    - ${}_tq_x = \frac{\frac{l_x}{l_0} - \frac{l_{x+t}}{l_0}}{\frac{l_x}{l_0}} = \frac{l_x - l_{x+t}}{l_x}$
    - ${}_tp_x = \frac{\frac{l_{x+t}}{l_0}}{\frac{l_x}{l_0}} = \frac{l_{x+t}}{l_x}$
    - ${}_{t|u}q_x = \frac{\frac{l_{x+t}}{l_0} - \frac{l_{x+t+u}}{l_0}}{\frac{l_x}{l_0}} = \frac{l_{x+t} - l_{x+t+u}}{l_x}$

What if given insufficient information?

- Not given the number of people to calculate values for specific ages
- Only given ${}_1q_x$ for each x but need to calculate ${}_2q_x, {}_3q_x\ etc$
- Manually calculate the cases of when the person could die
- Consider a probability tree:



Multiple Life Functions

- Considering the life of a group of people rather than just one
- Joint Life $\left(T_{X_1:X_2\ldots}\right)$
  - Considering when the first person is going to die
  - $P\left(\min\left(T_{X_1}, T_{X_2} \ldots\right)\right)$
- Last Survivor $\left(T_{\overline{X_1:X_2\ldots}}\right)$
  - Considering when the last person is going to die
  - $P\left(\max\left(T_{X_1}, T_{X_2} \ldots\right)\right)$

Consider special case when there are only TWO people in the group

- Use logic to resolve each statement
- Consider the possible cases for each of the statements

	$\mathbf{min}\left(\boldsymbol{T_{X_1}}, \boldsymbol{T_{X_2}}\right)$	$\mathbf{max}\left(\boldsymbol{T_{X_1}}, \boldsymbol{T_{X_2}}\right)$
$P(min/max\ \geq t)$	○ Minimum will always be greater than t ○ $P\left(T_{X_1}, T_{X_2} \geq t\right)$ **“and”**	○ Could be smaller or greater ○ Consider cases: ○ $P\left(T_{X_1} \geq t\right), P\left(T_{X_2} \leq t\right)$ ○ $P\left(T_{X_1} \leq t\right), P\left(T_{X_2} \geq t\right)$ ○ $P\left(T_{X_1} and\ T_{X_2} \geq t\right)$ **“and/or”**
$P(min/max\ \leq t)$	○ Could be smaller or greater ○ Consider cases: ○ $P\left(T_{X_1} \geq t\right), P\left(T_{X_2} \leq t\right)$ ○ $P\left(T_{X_1} \leq t\right), P\left(T_{X_2} \geq t\right)$ ○ $P\left(T_{X_1} and\ T_{X_2} \leq t\right)$ **“and/or”**	○ Maximum will always be smaller than t ○ $P\left(T_{X_1}, T_{X_2} \leq t\right)$ **“and”**

Not included: Expectation

Week 7&9: Practical Applications 3

Generating continuous random variables

- Method of randomly generating continuous variables using CDF
- Some random x will result in some random $F(x)$ between 0 and 1
- ∴ If inverse a random number between 0 and 1 → Will result in random x
  - How to get a random number between 0 and 1? → Use uniform distribution
  - Randomly generate numbers from the uniform distribution using Calc/Excel/R
  - Find the CDF of X → Make X the subject (Inverse function)
  - Substitute in the randomly generated numbers between 0 and 1 to get a randomly generated X

Capture – Recapture:

- Method for **estimating** the population size
  - Collect a sample and mark all units captured then release them into wild
  - Collect a sample & note of how many marked units are recaptured
  - Proportion of animals marked should remain the same:
    - $\frac{Marked}{Population} = \frac{Recaptured}{Second\ sample} \rightarrow \frac{m}{\hat{N}} = \frac{c}{n}$
    - Solve for $\hat{N}$
- Key Assumption:
  - Population is closed (Remains constant)
  - All units have equal probability of being captured
  - All units captured are independent of one another
  - Markings will remain on initially captured units

Maximum Likelihood Method (MLM)

- Method for estimating a parameter θ in the distribution that best explains the dataset
- What is the θ that is most likely to produce these set of results?
- **Likelihood function $\left(\boldsymbol{L}(\boldsymbol{\theta})\right)$:**
    - Function that describes the probability of observing the dataset
    - No specific results → $\prod PDF/PMF$
    - Specific data→ Multiply the probability (PDF/PMF) of each result together
    - Use calculus to maximize the function with respect to θ
- **Log-Likelihood function** $\left(\boldsymbol{l}(\boldsymbol{\theta})\right)$:
    - Natural Logarithm is monotonic → $\max l(\theta) = \max L(\theta)$
    - Original likelihood function is complicated and hard to differentiate
    - Log-likelihood simplifies complicated functions:
        - Multiplication becomes addition
        - Powers can be brought down
        - Log differentiation is simpler
    - Maximize the log-likelihood function instead → Solve for θ
    - $\therefore$ **$\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$**
- **What if $\boldsymbol{\theta}$ is included in the range?**
    - Likelihood function depends on the range $x \leq\geq \theta$
    - Likelihood is maximized when all $x_i \leq\geq \theta$
        - $x \geq \theta \to \theta = \min x \; (x_{\min})$
        - $x \leq \theta \to \theta = \max x \; (x_{max})$
        - Use order statistics to determine min/max of the sample

Method of Moments (MOM)

- Alternative (Inferior) method for estimating parameter θ
- Based on the Law of Large Numbers → Sample mean converges to population expectation
- Equate sample mean to population moment
    - $\overline{X^k} = E\left(X^k\right)$
    - $\frac{\sum X^k}{n} = \sum x^k * PMF = \int x^k * PDF$
    - Solve for θ
    - $\therefore$ **$\widehat{\boldsymbol{\theta}}$ is the method of moments estimator of $\boldsymbol{\theta}$**

Comparison between the two:

- **Advantages & Disadvantages of MLM:**
    - Asymptotically unbiased & efficient
    - Invariance property → Also the MLE of any function of the MLE
    - Complicated mathematics → Requires computing
- **Advantages & Disadvantages of MOM:**
    - Simple to use → No heavy calculation required
    - Requires a large sample → If not inaccurate & unbiased

Log-normal distribution: Brownian Stock Price Model

- Log-normal → Logarithm of the variable is normally distributed
  - $\mathrm{Y} = \ln X$ is normal
  - $X = e^{Y}$ is log-normal
  - More generally → $X = a * e^{by}$
- Brownian Stock Price Model
  - $S_t = S_0 * e^{rt}$
    - $S_t \rightarrow Future\ stock\ price\ at\ time\ t$
    - $S_0 \rightarrow Current\ stock\ price$
    - $r \rightarrow Rate\ of\ return\ of\ the\ stock$
    - $t \rightarrow\ Time\ elasped$
  - $E(S_t) = S_0 * e^{\mu t}$
  - $Var\ (S_t) = S_0^2 * e^{2\mu t} * \left(e^{\sigma^2 t} - 1\right)$
    - $\mu \rightarrow Expected\ return\ per\ time\ period$
    - $\sigma \rightarrow Volatility\ per\ time\ period$
  - r is normally distributed
    - $\frac{S_t}{S_0} = e^{rt}$
    - $\ln \frac{S_t}{S_0} = rt \sim Normal$
    - $\ln S_t - \ln S_0\ \sim N\left(\left(\mu - \frac{\sigma^2}{2}\right) t, \sigma^2 t\right)$
    - $\ln S_t \sim N\left(\ln S_0 + \left(\mu - \frac{\sigma^2}{2}\right) t, \sigma^2 t\right)$
    - Convert probability problems surrounding S_t to $\ln S_t$
    - Use normal distribution to solve

Week 10: Expectation & Moments

Expectation

- **Single Variable** → $E[g(x)] = \sum g(x) * p(x) = \int g(x) * f(x)\, dx$
- **Joint Variable** → $E[g(x,y)] = \sum g(x,y) * p(x,y) = \int \int g(x,y) * f(x.y)\, dxdy$
- **Properties** → $E(aX \pm bXY + c) = aE(X) \pm bE(XY) + c$
- **Terminology:**
    - Product moment about origin (0,0)
        - $E((X-0)^r(Y-0)^s) = E(X^rY^s) = \mu'_{r,s}$
    - Product moment about mean (μ_x, μ_y)
        - $E\left((X-\mu_x)^r\left(Y-\mu_y\right)^s\right) = \mu_{r,s}$

Covariance $(\boldsymbol{\sigma_{X,Y}})$

- Measure of the LINEAR relationship between two variables
- Product of the first moment about mean
    - $\mu_{1,1} = \mu'_{1,1} - \mu_x * \mu_y$
    - $Cov\,(X,Y) = E(XY) - E(X) * E(Y)$
        - If independent → $E(XY) = E(X) * E(Y)$ → $\sigma_{x,y} = 0$
        - If $\sigma_{x,y} = 0$ → Does NOT mean independent → Could be NON-linear RS

Joint Variance $(\boldsymbol{\sigma^2_{X+Y}})$

$Var\,(aX \pm bY)$

$= E[(aX \pm bY)^2] - [E(aX \pm bY)]^2$

$= E(a^2X^2 \pm 2abXY + b^2Y^2) - [aE(X) + bE(Y)]^2$

$= a^2E(X^2) \pm 2abE(XY) + b^2E(Y^2) - \left[[aE(X)]^2 \pm 2abE(X)E(Y) + \left[b\left(E(Y)\right)^2\right]\right]$

$= a^2E(X^2) \pm 2abE(XY) + b^2E(Y^2) - a^2[E(X)]^2 \mp 2abE(X)E(Y) - b^2[E(Y)]^2$

$= a^2[E(X^2) - [E(X)]^2] + b^2[E(Y^2) - [E(Y)]^2] \pm 2ab[E(XY) - E(X)E(Y)]$

$= a^2\, Var\,(X) + b^2\, Var\,(Y) \pm 2ab\, Cov\,(X,Y)$

Application:

- Variance of the sum of variables is the sum of the individual variances plus two times the covariance between each variable
- $Var\,(X_1 + X_2 + X_3 + \cdots X_n) = \sum Var\,(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=1+1}^{n} Cov\,\left(X_i, X_j\right)$

Conditional Moments

- **Expectation** → $E(X|Y) = \sum x * p(x|y) = \int x * f(x|y)\, dx$
  - $p(x|y)$ and $f(x|y)$ can be calculated methods taught in week 5
- **Variance** → $Var\ (X|Y) = E(X^2|Y) - [E(X|Y)]^2$
- Same properties as unconditional ones

Unconditional & Conditional Expectation

- **Law of total probability**
  - $P(A) = P(AB_1) + P(AB_2) + \cdots$
  - $P(A) = P(A|B_1) * P(B_1) + P(A|B_2) * P(B_2) + \cdots$
- **Law of total expectation:**
  - $A * P_A(A) = A * P_A(A|B_1) * P(B_1) + A * P_A(A|B_2) * P(B_2) + \cdots$
  - $E_A(A) = E_A(A|B_1) * P(B_1) + E_A(A|B_1) * P(B_2) + \cdots$
  - $\boldsymbol{E_A(A) = E_B[E_A(A|B)]}$ → **Law of iterated expectations**

Unconditional & Conditional Expectation

- Consider the expectation of the conditional variance:
  - $E\big(Var\ (X|Y)\big) = E[E(X^2|Y)] - E[E(X|Y)^2]$
  - $E\big(Var\ (X|Y)\big) = E(X^2) - E[E(X|Y)^2]$
- Consider the Variance of the conditional Expectation:
  - $Var\ \big(E(X|Y)\big) = E[E(X|Y)^2] - \big[E[E(X|Y)]\big]^2$
  - $Var\ \big(E(X|Y)\big) = E[E(X|Y)^2] - [E(X)]^2$
- Adding these two terms together:
  - $E(X^2) - E[E(X|Y)^2] + E[E(X|Y)^2] - [E(X)]^2 = E\big(Var\ (X|Y)\big) + Var\ \big(E(X|Y)\big)$
  - $E(X^2) - [E(X)]^2 = E\big(Var\ (X|Y)\big) + Var\ \big(E(X|Y)\big)$
  - $\boldsymbol{Var\ (X) = E\big(Var\ (X|Y)\big) + Var\ \big(E(X|Y)\big)}$

Actual Calculation:

- Question will typically give a conditional distribution:
  - Distributed $X_1\ when\ Y = y_1, X_2\ when\ Y = y_2$ ...
  - $\therefore\ Distributed\ X|Y$
- Use the formula for the known distributions Expectation and Variance and substitute into the above unconditional formulas
- Manipulate equation till have something that can be calculated from the given
- **IMPORTANT** → Sometimes some parameters given may be a known distribution as well

Moment Generating Function (MGF)

- Method that allows us for calculating higher moments of complicated distributions
  - $M_X(t) = E(e^{tx}) = \sum e^{tx} * p(x) = \int e^{tx} * f(x)$
  - The r-th derivative of the MGF evaluated at 0 is the value of the r-th moment of the variable
- Why would we want to know higher moments?
  - First moment → Location (Expectation)
  - Second moment → Dispersion (Variance)
  - Third moment → Skewness
  - Fourth moment → Flatness (Kurtosis)

Proof of MGF:

$$E(e^{tx}) = E\left(1 + t + \frac{t^2}{2!}x^2 + \cdots \frac{t^n}{n!}x^n\right)$$

$$E(e^{tx}) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots \frac{t^n}{n!}E(X^n)$$

$$\frac{d}{dt}|_{t=0} = E(X)$$

$$\frac{d^2}{dt^2}|_{t=0} = E(X^2)$$

$$\vdots$$

$$\frac{d^n}{dt^n}|_{t=0} = E(X^n)$$

- Works because each successive differentiation cancels out the factorial denominator and cancels out all variable t from moments smaller than or equal to n
- Substituting $t = 0$ helps to remove all moments above n as they still have variables

Joint Moment Generating Functions (MGFs)

- Same principle but using partial differentiation instead
  - $M_{X,Y}(t_1, t_2) = E(e^{tx+ty})$
  - $E(X^r) = \frac{\partial^r}{\partial x^r}|_{t=0}$ and $E(Y^s) = \frac{\partial^s}{\partial y^s}|_{t=0}$
  - $E(X^r Y^s) = \frac{\partial^{r+s}}{\partial x^r \partial y^s}|_{t=0}$
- Independence:
  - $If\ M_W(t) = M_X(t) * M_Y(t),$
  - $W = X + Y$, where X and Y are independent variables

Commonly Known MGFs:

Bernoulli Distribution

- $X \sim Bernoulli\ (p)$
- $p(x) = p^x(1-p)^{1-x}$

$M_X(t)$
$= \sum e^{tx} * p(x)$

$M_X(t) = \sum_{x=0}^{1} e^{tx}\, p^x(1-p)^{1-x}$

$M_X(t) = \sum_{x=0}^{1} (pe^t)^x(1-p)^{1-x}$

$M_X(t) = (pe^t)^0(1-p)^1 + (pe^t)^1(1-p)^0$
$= (1-p) + pe^t$

$\therefore M_X(t) = (1-p) + pe^t$

Binomial Distribution

- $Y \sim Binomial\ (n, p)$
- $p(y) = \binom{n}{y} p^y(1-p)^{n-y}$
- Binomial is a sum of n independent Bernoulli trials
- $Y = X_1 + X_2 + \cdots X_n$

$M_Y(t)$
$= M_{X_1}(t) * M_{X_2}(t) * \ldots M_{X_n}(t)$
$= \prod_1^n (1-p) + pe^t$
$= \left(pe^t + (1-p)\right)^n$

$\therefore M_Y(t) = \left(pe^t + (1-p)\right)^n$

Geometric Distribution

- $X \sim Geometric\ (p)$
- $p(x) = (1-p)^{x-1}p$

$M_X(t)$
$= \sum e^{tx} * p(x)$
$= \sum_{x=1}^{\infty} e^{tx}\,(1-p)^{x-1}p$
$= \sum_{x=1}^{\infty} e^{tx}\,(1-p)^x(1-p)^{-1} * p$
$= \sum_{x=1}^{\infty} \left(e^t(1-p)\right)^x * \frac{p}{1-p}$
$= \frac{p}{1-p}\sum_{x=1}^{\infty} \left(e^t(1-p)\right)^x$
$= \frac{p}{1-p}\left[\left(e^t(1-p)\right)^1 + \left(e^t(1-p)\right)^2 + \cdots\right]$
$= \frac{p}{1-p}\left(e^t(1-p)\right)\left[1 + \left(e^t(1-p)\right)^1 + \cdots\right]$
$= pe^t\left[1 + \left(e^t(1-p)\right)^1 + \left(e^t(1-p)\right)^2 + \cdots\right]$
$= pe^t * \frac{1}{1-\left(e^t(1-p)\right)}$
$= \frac{p}{1-(1-p)e^t}$

$\therefore M_X(t) = \frac{p}{1-(1-p)e^t}$

Negative Binomial

- $Y \sim negbin\ (r, p)$
- $p(y) = \binom{y-1}{r-1} p^r(1-p)^{y-r}$
- Negative Binomial is a sum of r independent Geometric variables
- $Y = X_1 + X_2 + \cdots X_r$

$M_Y(t)$
$= M_{X_1}(t) * M_{X_2}(t) * \ldots M_{X_n}(t)$
$= \prod_1^n \frac{p}{1-(1-p)e^t}$
$= \left(\frac{p}{1-(1-p)e^t}\right)^r$

$\therefore M_Y(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$

Poisson Distribution

- $X \sim Poisson\ (\mu)$
- $p(x) = \frac{e^{-\mu} * \mu^x}{x!}$

$M_X(t)$

$$= \sum e^{tx} * p(x)$$

$$= \sum_{x=0}^{\infty} e^{tx} * \frac{e^{-\mu} * \mu^x}{x!}$$

$$= e^{-\mu} \sum_{x=0}^{\infty} \frac{e^{tx} * \mu^x}{x!}$$

$$= e^{-\mu} \sum_{x=0}^{\infty} \frac{(e^t \mu)^x}{x!}$$

$$= e^{-\mu} \left(1 + \frac{(e^t \mu)^1}{1!} + \frac{(e^t \mu)^2}{2!} + \cdots \infty\right)$$

$$= e^{-\mu}\left(e^{e^t \mu}\right) \to Taylor\ series$$

$$= e^{e^t \mu - \mu}$$

$$= e^{\mu(e^t - 1)}$$

$\therefore M_X(t) = e^{\mu(e^t - 1)}$

Uniform Distribution

- $Y \sim Uni\ (a, b)$
- $f(y) = \frac{1}{b-a}$

$M_Y(t)$

$$= \int_a^b e^{ty} * f(y)\, dy$$

$$= \int_a^b e^{ty} * \frac{1}{b-a}\, dy$$

$$= \frac{1}{b-a} * \int_a^b e^{ty}\, dy$$

$$= \frac{1}{b-a} * \frac{1}{t} * \left[e^{ty}\right]_a^b$$

$$= \frac{1}{b-a} * \frac{1}{t} * \left(e^{tb} - e^{ta}\right)$$

$$= \frac{e^{tb} - e^{ta}}{t(b-a)}$$

$\therefore M_Y(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$

Exponential Distribution

- $X \sim Exp\ (\lambda)$
- $f(x) = \lambda e^{-\lambda x}$

$M_X(t)$

$$= \int_0^{\infty} e^{tx} * f(x)\, dy$$

$$= \int_0^{\infty} e^{tx} * \lambda e^{-\lambda x}\, dy$$

$$= \lambda \int_0^{\infty} e^{(t-\lambda)x}\, dy$$

$$= \lambda \left[\frac{e^{(t-\lambda)x}}{t-\lambda}\right]_0^{\infty}$$

$$= \lambda \left[0 - \frac{1}{t-\lambda}\right]$$

$$= \frac{\lambda}{\lambda - t}$$

$\therefore M_X(t) = \frac{\lambda}{\lambda - t}$

Gamma Distribution

- $Y \sim Gamma\ (\alpha, \lambda)$
- $f(y) = \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha - 1}}{\Gamma(\alpha)}$
- Gamma is a sum of α independent Exponential Variables
- $Y = X_1 + X_2 + \cdots X_n$

$M_Y(t)$

$$= M_{X_1}(t) * M_{X_2}(t) * \ldots M_{X_n}(t)$$

$$= \prod_1^n \frac{\lambda}{\lambda - t}$$

$$= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$$

$\therefore M_Y(t) = \left(\frac{\lambda}{\lambda - t}\right)^{\alpha}$

Mixed distributions

- Distributions that are BOTH discrete and continuous for certain range of values
- How to know if something is mixed?
    - Does the variable look continuous?
    - Is there some situation where the variable takes a specific value with $P > 0$
- Examples:
    - Queue time → Time is continuous
        - But there are times when there is no one in the queue → Variable takes 0 with probability greater than zero
    - Rainfall → Volume is continuous
        - But there are times when it does not rain → Variable takes 0 with probability greater than zero

Method to solve:

- Typically given a CDF for the whole variable
    - Split the range of the variables
    - Identify where is discrete and continuous
    - $a \leq x < b \rightarrow x = a \; and \; a < x < b$
- Find a probability function of the whole variable
    - PDF and PMF for the respective sections
    - Differentiate to obtain the PDF → Integrate with limits to obtain the probability
    - Work backwards from CDF to obtain Probability of the discrete outcomes
    - Combine PDF and PMFs to find the probability function
- Application
    - Use the normal expectation and variance formulas to calculate from here on out

Week 11: Limit Theorems

Markov Inequality

- For all non-negative X
- Probability has an upper bound

$$P(x \geq a) \leq \frac{E(X)}{a}$$

Since X is positive,

Restricting to a smaller range would reduce the magnitude of the expression, hence the inequality to 'Smaller'.

Proof:

$$E(X) = \int_0^\infty x * f(x)\,dx \geq \int_a^\infty x * f(x)\,dx \geq \int_a^\infty a * f(x)\,dx$$

$$E(X) \geq \int_a^\infty a * f(x)\,dx$$

$$E(X) \geq a * \int_a^\infty f(x)\,dx$$

$$E(X) \geq a * P(x \geq a)$$

$$P(x \geq a) \leq \frac{E(X)}{a}\ (Proven)$$

Chebyshev's Inequality

- For all non-negative X
- Probability that X differs from μ is fixed ('Confidence Interval')

$$P(|X - \mu| > k) = \frac{Var\ (X)}{k}$$

Proof:

$$P(X \geq a) = \frac{E(X)}{a}$$

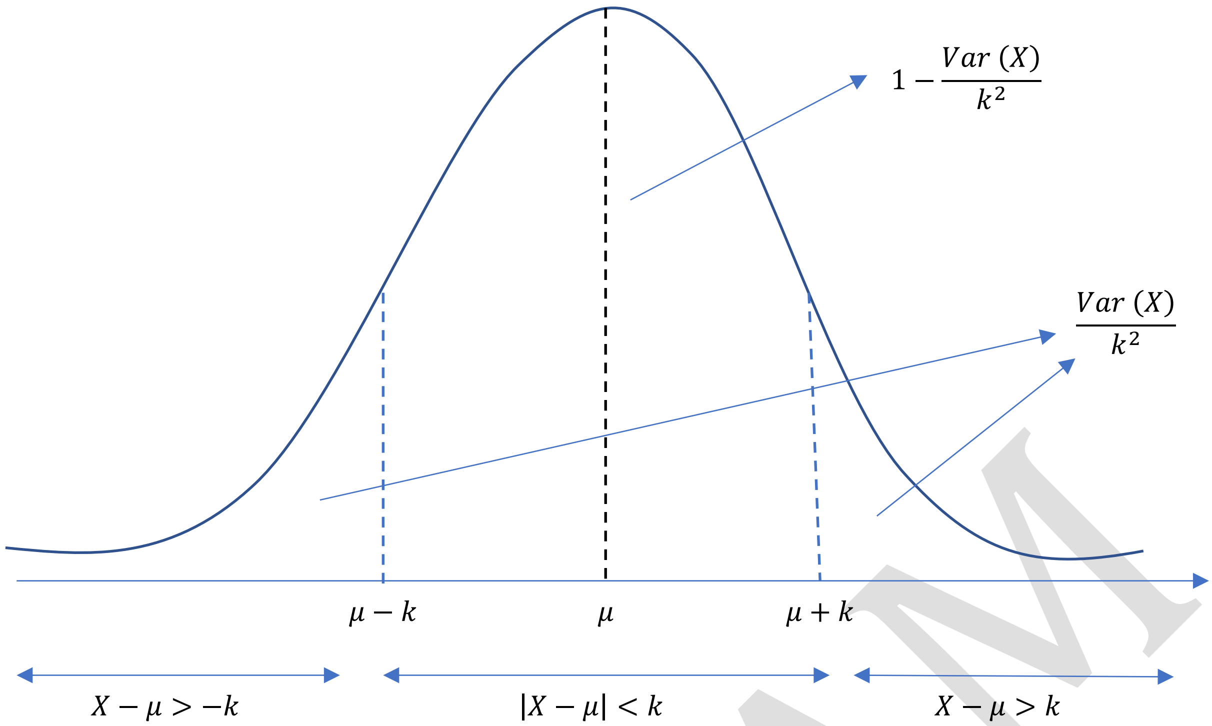
$$P(|X - \mu| > k) = \frac{E(X - \mu)}{k}$$

$$P(|X - \mu|^2 > k^2) = \frac{E(|X - \mu|^2)}{k^2} = \frac{Var\ (X)}{k^2}$$

Since referring to the same area,

$$P(|X - \mu| > k) = P(|X - \mu|^2 > k^2)$$

$$P(|X - \mu| > k) = \frac{Var\ (X)}{k^2}\ (Proven)$$



Application: Consider Standard Deviations

$$P(|X - \mu| > k\sigma) = \frac{Var\ (X)}{k^2\sigma^2} = \frac{\sigma^2}{k^2}$$

$$P(|X - \mu| > k\sigma) = \frac{1}{k^2}$$

$$P(|X - \mu| \leq k\sigma) = 1 - \frac{1}{k^2}$$

- At least $1 - \frac{1}{k^2}$ of the distribution lies within k standard deviations from the mean
- General theorem that works for ALL distributions
- But more specific distribution analysis is more rigorous:
  - Normal distribution:
    - 1 SD from mean → 68%
    - 2 SD from mean → 95%
    - 3 SD from mean → 99.7%
  - Chebyshev:
    - 2 SD → 75%
    - 3 SD → 88.89%

Law of Large Numbers

- For a population that is Independently and Identically Distributed
- Average of a sample converges to the true mean of the population

Let X be a random variable, where $X_1, X_2 \dots X_n$ are IID.

$$\bar{X} = \frac{X_1 + X_2 + \cdots X_n}{n} = \frac{\sum X}{n}$$

Since $\bar{X}$ is a sum of random variables, $\bar{X}$ itself is a random variable with a distribution.

$$E(\bar{X}) = E\left(\frac{\sum X}{n}\right) = \frac{1}{n}E\left(\sum X\right) = \frac{1}{n} * \left(nE(X)\right) = E(X)$$

$$Var\ (\bar{X}) = Var\ \left(\frac{\sum X}{n}\right) = \frac{1}{n^2}Var\ \left(\sum X\right) = \frac{1}{n^2}\left(n\ Var\ (X)\right) = \frac{Var\ (X)}{n}$$

By Chebyshev's inequality,

$$P(|\bar{X} - \mu| > k) = \frac{Var\ (\bar{X})}{k^2}$$

$$P(|\bar{X} - \mu| > k) = \frac{Var\ (X)}{nk^2}$$

$$As\ n \to \infty, \frac{Var\ (X)}{nk^2} \to 0,$$

$$P(|\bar{X} - \mu| > k) \to 0$$

- Probability that the sample mean differs from the true mean by some value tends to zero
- ∴ Sample mean is not very far off (Equal) to the true mean in the limit

Central Limit Theorem

- For a population that is Independently and Identically Distributed
- Distribution of a sum of a sufficiently large number of variables is approximately normal

Important points:

CLT is NOT just for the sample mean → Can be for any sum of variables

Variation: Sum of TWO DIFFERENT random variables:

- Use CLT twice on each random variable → Two normal distributions
- Use the sum of two independent normal distributions
- $S_X \pm S_Y \sim N(\mu_X \pm \mu_Y, \sigma_X^2 + \sigma_Y^2)$

Proof:

We normalize $\bar{X}$,

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

We consider the MGF of Z:

$$M_Z(t) = E\left(e^{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}t}\right)$$

$$M_Z(t) = E\left(e^{\frac{\sqrt{n}\bar{X}}{\sigma}t} * e^{-\frac{\sqrt{n}\mu t}{\sigma}}\right)$$

$$M_Z(t) = e^{-\frac{\sqrt{n}\mu t}{\sigma}} * E\left(e^{\frac{\sqrt{n}\bar{X}}{\sigma}t}\right)$$

$$M_Z(t) = e^{-\frac{\sqrt{n}\mu t}{\sigma}} * M_{\bar{X}}\left(\frac{\sqrt{n}}{\sigma}t\right)$$

$$M_Z(t) = e^{-\frac{\sqrt{n}\mu t}{\sigma}} * M_{n\bar{X}}\left(\frac{1}{\sigma\sqrt{n}}t\right)$$

$$M_Z(t) = e^{-\frac{\sqrt{n}\mu t}{\sigma}} * \left[M_{\bar{X}}\left(\frac{1}{\sigma\sqrt{n}}\right)\right]^n$$

$$\ln M_Z(t) = \ln\left(e^{-\frac{\sqrt{n}\mu t}{\sigma}} * \left[M_{\bar{X}}\left(\frac{1}{\sigma\sqrt{n}}\right)\right]^n\right)$$

$$\ln M_Z(t) = -\frac{\sqrt{n}\mu t}{\sigma} + n \ln M_{\bar{X}}\left(\frac{1}{\sigma\sqrt{n}}\right)$$

$$\ln M_Z(t) = -\frac{\sqrt{n}\mu t}{\sigma} + \left(-\frac{\sqrt{n}\mu}{\sigma} + \frac{\sqrt{n}\mu}{\sigma}\right)t - \left(\frac{\mu_2' - \mu_1}{2\sigma^2}\right)t^2 + \left(\frac{\mu_3'}{6\sigma^3\sqrt{n}} - \frac{\mu_1'\mu_2'}{2\sigma^3\sqrt{n}} + \frac{\mu_1^3}{3\sigma^3\sqrt{n}}\right)t^3 + \cdots$$

$$\lim_{n\to\infty} \ln M_Z(t) \to \frac{t^2}{2}$$

$$\lim_{n\to\infty} M_Z(t) = e^{\frac{t^2}{2}}\ (MGF\ of\ a\ standard\ normal\ distribution)$$

∴ $\bar{X}$ tends to a normal distribution as n tends to infinity.

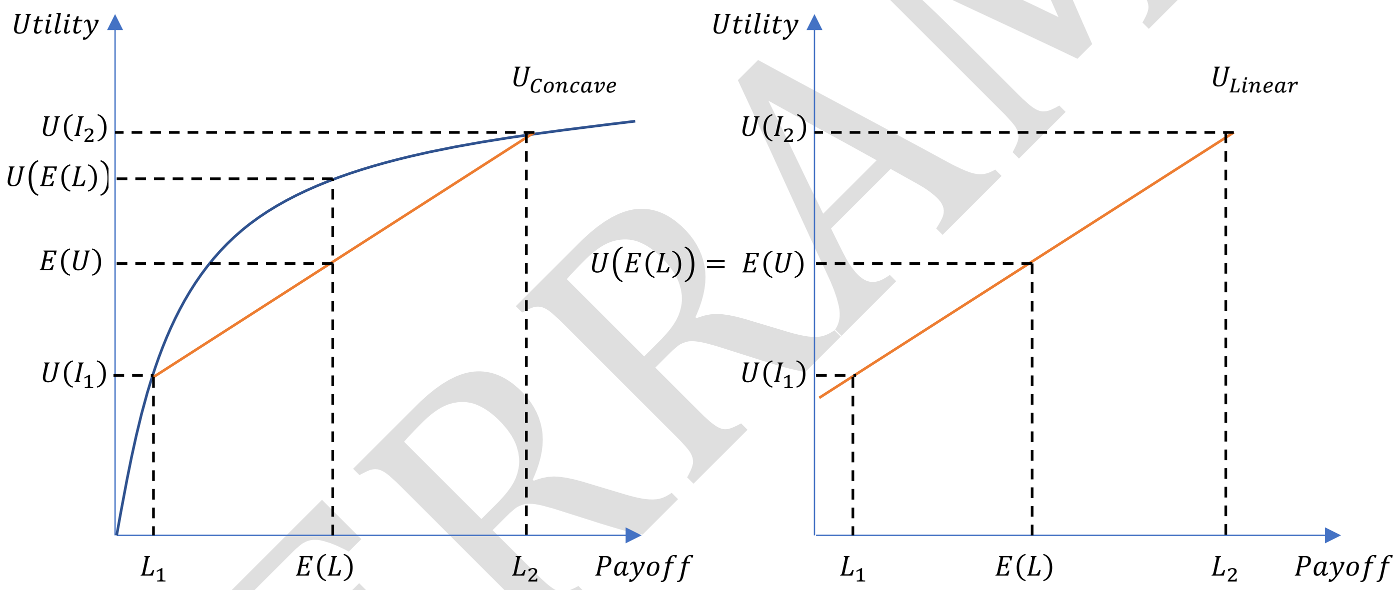
<u>Insurance & Utility functions (Micro 2 related)</u>

- Any rational human being would choose the option that maximizes their utility
- Compare between:
    - Without Insurance → Loss is a random variable
    - With Insurance → Pay premium but receive coverage equal to the loss
    - Either way → The loss is covered
    - How to determine which one provides more utility?

Janssen's Inequality

For any concave utility function,

$E[U(X)] \leq U[E(X)]$



Most individuals are Risk Averse:

- Concave Utility functions
- ∴ They gain more Utility from having Insurance policies rather than paying the Loss

Note for calculations:

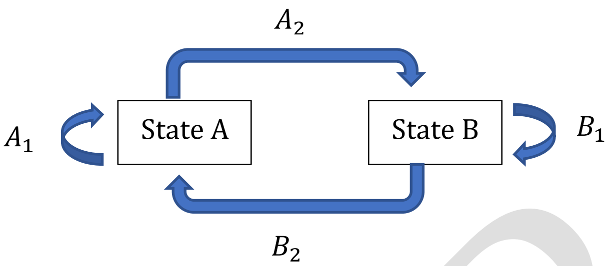
- Cannot gain utility from Loss → Convert to utility from remaining wealth instead
- $u(Income), u(Income - Loss)$ or $u(Income - premium)$
- Without insurance → Expected Utility
    - $E(U) = p * u_1 + (1-p) * u_2$
- With insurance → Utility of Expected Loss
    - $u(EV) = u(p * l_1 + (1-p) * l_2)$

Stochastic Processes: Finite Markov Chains

- Stochastic Process → Value of variable changes over time in an uncertain way
- Markov Chain → Stochastic process where ONLY the present value matters

Consider a Markov Chain:

- Each chain has multiple discrete outcomes → States
- Variable may retain their state or move to another state each time period → Transitions
- Visualization of process:
  - May have more states and states may jump or skip states to another
  - Transitions (Arrows from a state) should add up to 1 as they are probabilities
  - Final state → Absorptive state → Will not change anymore



Position Matrix

- Matrix representing the state
- 1 if it is, 0 if it is not

$$p = \begin{array}{cc} [1 & 0] \\ A & B \end{array}$$

Current State

Transition Matrix

- Matrix representing the arrows
- Each row should add up to one

$$P = \begin{array}{c|cc} & A & B \\ \hline A & A_1 & A_2 \\ B & B_1 & B_2 \end{array}$$

Future State

Current State

Markov Chain:

$p^{(1)} = p^{(0)} * P$

$p^{(2)} = p^{(1)} * P = p^{(0)} * P^2$

$p^{(3)} = p^{(2)} * P = p^{(0)} * P^3$

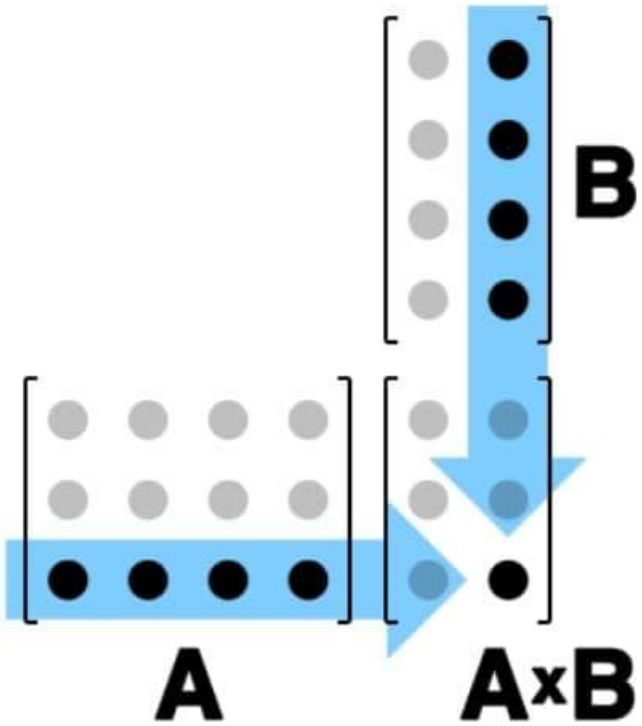
$\vdots$

$p^{(n)} = p^{(n-1)} * P = p^{(0)} * P^n$

Pro Tip: Always multiply position matrix first to simplify calculation process.

Recap: How to multiply Matrices?

- Each matrix has a Row & Column $(m * n)$
- Can only multiply matrices when their opposite Row & Columns match
    - $(2 * 3) * (3 * 4) = (2 * 4)$
    - Common middle number gets cancelled out
    - Sum of the dot product to find each new value in the matrix (Repeat for all new positions)



Week 12: Practical Applications 4 (Insurance)

General Idea of Insurance:

- People are always at risk of financial losses (Accident, Health etc)
- People can reduce this risk by buying insurance
  - People pay the insurance company **Pure Premium**
  - Insurance company reimburses losses (All or partially)

Losses are Random Variables

- **Depends on two factors:**
  - Number of losses → Frequency
  - Amount per loss → Severity
- **Parameters:**
  - Expectation → Expected Claims (μ)
  - Standard Deviation → Risk → σ
  - Coefficient of Variation → $CV = \frac{\sigma}{\mu}$
    - General indicator of risk
    - Can be used to compare across options

Why does an insurance company work?

- Insurance company aggregates losses:

$$S_n = X_1 + X_2 + \cdots X_n$$

$$E(S_n) = n\mu$$

$$Var\ (S_n) = n\sigma^2$$

$$CV = \frac{\sqrt{n\sigma^2}}{n\mu} = \frac{\sqrt{n}\sigma}{n\mu} = \frac{\sigma}{\sqrt{n}\mu}$$

- Insurance company faces smaller risk compared to any individual $\frac{\sigma}{\sqrt{n}\mu} < \frac{\sigma}{\mu}$
- As $n \to \infty. \frac{\sigma}{\sqrt{n}\mu} \to 0$ *and* S_n *becomes normal*
- Risk to the insurance becomes closer to zero the more people it has

How does an insurance company make profits?

- Increasing the premium of the policy beyond just the expected loss reduces the utility gained from the insurance policy
- Consumers will choose the insurance policy so long as they gain more than or equal utility from paying the expected loss rather than the expected utility from the loss
- Can continue to charge extra up till the point where the consumer is indifferent
- This is known as the **Risk Premium**
- **Gross Premium = Pure Premium + Risk Premium**
- To the insurance company this could be profit, administrative fees etc

Caveats of insurance policies:

- Deductibles → Part of the Loss that the individual must pay themselves (Flat amount)
- Co-Insurance → Part of the remaining loss that is split between you and insurer (%)
- Benefit Limits → Maximum claim payout (Flat amount or %)

Form a piecewise function on the insurance payout:

$$Payout = \left\{\begin{array}{cc} 0 & 0 < X \leq Deductible \\ X - deductible & Deductible < X < Benefit\ Limit + Deductible \\ Benefit\ Limit & Benefit\ Limit + Deductible < X < Maximum\ Loss \end{array}\right\}$$

- Probability of payout is the probability of X being in that range
- Expectation and Variance of payout is using the payout function as functions of X

Why do insurers have these?

- To prevent Moral Hazards/Adverse selection results (Micro 2)
- To reduce the number of small claims → Lowers expected claim and risk (Safer)
- To help keep insurer safely within their financial capabilities

<u>**Week 13: Practical Application 5 (Investment)**</u>

Modern Portfolio Theory

- Owning just one kind of asset is VERY risky
- If the asset goes poorly then the portfolio suffers
- Need to diversify portfolio in order to reduce risk
- Question is how to allocate funds between two risky assets?

Consider the share of your total income for investing:

- 1 = 100% of income
- Share A to allocate on asset X
- Share (1-A) to allocate on asset Y
- $P = AX + (1 - A)Y$

$$E(P) = E(AX + (1 - A)Y)$$

$$E(P) = A * E(X) + (1 - A) * E(Y)$$

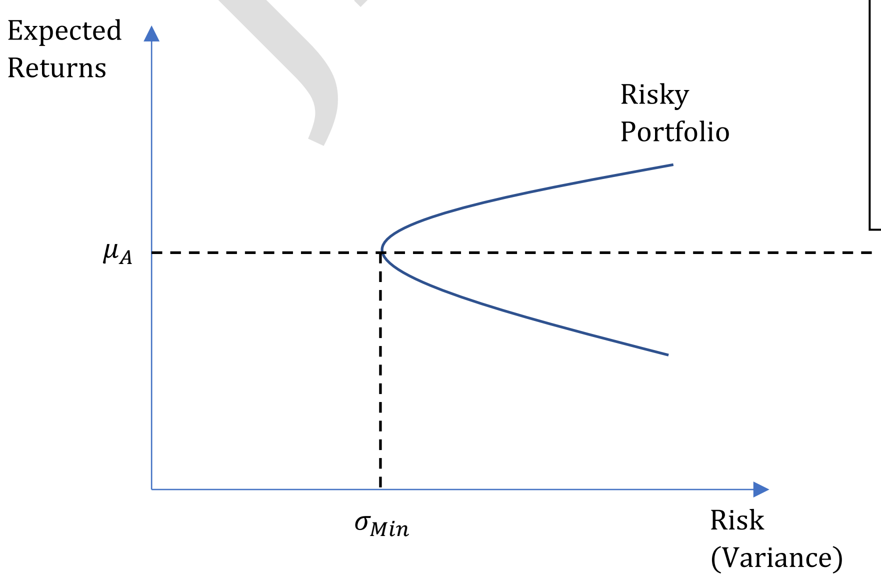
$$Var\ (P) = Var\ (AX + (1 - A)Y)$$

$$Var\ (P) = A^2 * Var\ (X) + (1 - A)^2 * Var\ (Y) + 2A(1 - A) * Cov\ (X, Y)$$

$$SD\ (P) = \sqrt{A^2 * Var\ (X) + (1 - A)^2 * Var\ (Y) + 2A(1 - A) * Cov\ (X, Y)}$$

Solving for A depends on what we want to achieve:

- Most people want to minimize risk in their portfolio
- Use minimization on the expression for Variance
- $\frac{dVar\ (P)}{dA} = 0$
- Solve for A and hence 1-A to determine how to allocate



Efficient Frontier:

- Has the same risk but higher expected return
- It is rational to only invest at a point in this area

What about deciding income allocation between Risky and Risk-Free Assets?

- Risk free assets have fixed guaranteed interest rates (Market rate)
- EG. Government Bonds

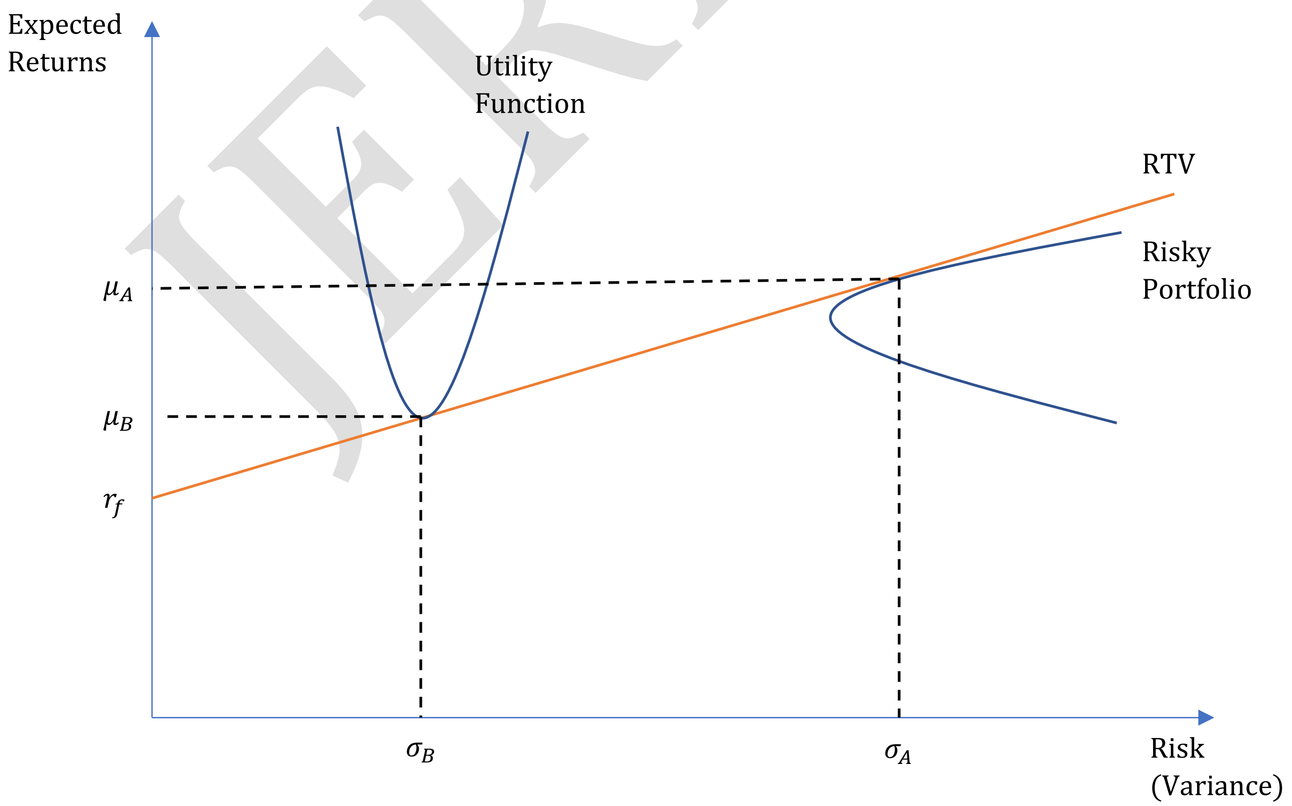
Comparison between your portfolio returns and the market rate:

- $1 = 100\%$ of income
- Share B to allocate on asset B
- Share (1-B) to allocate on asset B
- $RTV = \frac{E(P-r_f)}{SD(P-r_f)} = \frac{E(P)-r_f}{SD(P)}$
    - Want to maximize the numerator (Returns) while Minimizing the denominator (Risk)
    - Substitute in the expressions for E(P) and SD(P) to solve for the optimal share between risky assets
    - Manipulate to form an expression for A (Formula sheet)

$$A = \frac{[E(X) - r_f]\sigma_Y^2 - [E(Y) - r_f]\sigma_{XY}}{[E(X) - r_f]\sigma_Y^2 + [E(Y) - r_f]\sigma_X^2 - [E(X) - r_f + E(Y) - r_f]\sigma_{XY}}$$

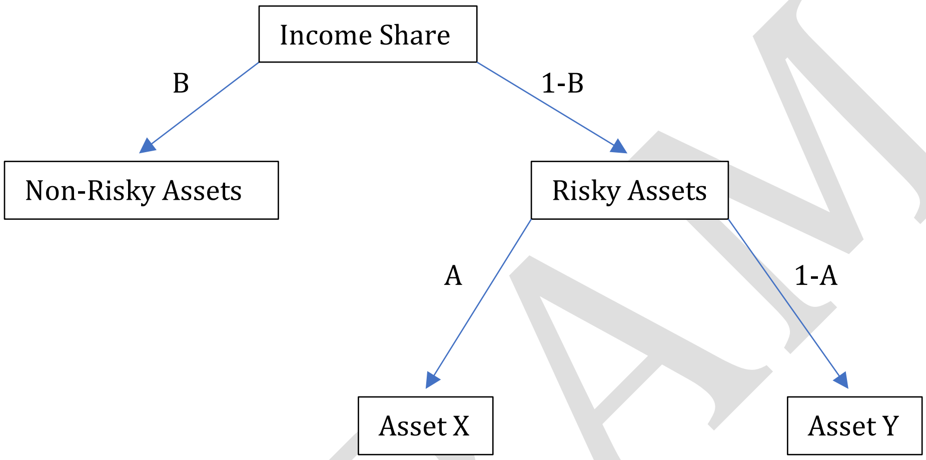
$$\mu_A = A * \mu_X + (1 - A)\mu_Y$$

$$\sigma_A^2 = A^2\,\sigma_X^2 + (1 - A)^2\sigma_Y^2 + 2A(1 - A)\sigma_{XY}$$



Explanation:

- Each unique set of Risk-Free and Risky portfolio combinations will lead to a different Sharpe Ratio
- Depending on the Risk Preference (Utility) of the individual, they may choose to invest a different proportion of their income to Risky or Risk-Free assets
- No matter how much they allocate between risk (B), the allocation between risky assets (A) remain the same for the same portfolio because you would always aim to minimize the risk regardless of your risk preference



How to determine B?

- **Calculated the expected return from a joint portfolio of risky and non-risky**
  - $\mu_B = B * r_f + (1 - B)\mu_A$
  - μ_A can be obtained via a formula using RTV
  - μ_B should be identified using the graph's intersection with the utility function

European Option Pricing Model

- Call option → Option to buy stock at strike price K at maturity
- Put option → Option to sell stock at strike price K at maturity
- Can be used as a means of Speculation for profit or Hedging a position
- Known as a **Derivative Security** (Asset whose value depends on another asset)
- Option prices depend on how the underlying stock is going to change

How much should options be charged for?

- Depends on how we expect the underlying stock price changes
- We know that stock prices are random and fluctuate greatly
- How can we model stocks and then options? → Stochastic Process

Option 1: Brownian Motion

- Every unit of time one standard normal variable is drawn that decides how it changes
- $\Delta z = \epsilon\sqrt{\Delta t}$
- $\therefore \Delta z \sim N(0, T)$
  - We know do not expect stock prices to be zero
  - Stock prices cannot also vary into negative region
  - NOT a good way to model

Option 2: Generalized Brownian Motion

- Consider a general form with Draft Rate a and Variance b^2
- $\Delta x = a\ dt + b\ dz$
- $\therefore dx \sim N(aT. b^2 T)$
  - Both a and b are absolute values → Not possible to remain constant throughout since they depend on the current stock price
  - NOT a good way to model

Option 3: Geometric Brownian Motion (Ito's Process)

- Adjusts a and b to become a function of both the current value and time
- $dx = a(x,t)\ dt + b(x,t)\ dz$
- $dS = \mu S\ dt + \sigma S\ dz$
- $\frac{dS}{S} = \mu\ dt + \sigma\ dz$
- $\therefore \frac{dS}{S} \sim N(\mu T, \sigma^2 T)$
  - Percentage change in stock price follows a Brownian motion
  - Since it is percentage change, μ & σ are likely to be constants throughout (Valid)

Main question: How to model options?

- Derivative securities are a function of stock prices and time → $G(S, t)$
- Through Stochastic Calculus, we obtain Ito's Lemma:

$$dG = \left(\frac{\partial G}{\partial S}\mu S + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial S^2}\sigma^2 s^2\right)\, dt + \frac{\partial G}{\partial S}\sigma S\, dz$$

- Let G = …
- Find the derivatives and substitute in to see if it also follows a Brownian Motion
  - If coefficients of dt and dz are constant
    - "Follow a geometric Brownian motion with mean and volatility…"
  - If coefficients of dt and dz are NOT constant
    - "Does not follow a geometric Brownian motion as its mean/volatility are in terms of…"

Special case: Proof for the Log-Normal stock price model

Let $G = \ln S$

$$\frac{\partial G}{\partial S} = \frac{1}{S}$$

$$\frac{\partial^2 G}{\partial S^2} = -\frac{1}{S^2}$$

$$\frac{\partial G}{\partial t} = 0$$

$$dG = \left(\frac{1}{S}\mu S + 0 + \frac{1}{2}\left(-\frac{1}{S^2}\right)\sigma^2 S^2\right) dt + \frac{1}{S}\sigma S\, dz$$

$$dG = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dz$$

$$dG = \ln S_T - \ln S_0$$

$$\therefore \ln S_T - \ln S_0 = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma\, dz$$

$$\ln S_T - \ln S_0 \sim N\left(\left(\mu - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right)$$

Riskless portfolio:

- Both dS and dG both have a common random dz term in their expressions
- Remove dz → Remove risk in portfolio

$$P = -G + \frac{\partial G}{\partial S} S$$

$$\Delta P = -\Delta G + \frac{\partial G}{\partial S}\Delta S$$

$\vdots$ $(Manipulation)$

$$\Delta P = -\left(\frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial S^2}\sigma^2 s^2\right)\Delta t$$

- No more dz in the end → Risk free
- Positive/Negative G or S means to buy or sell that asset respectively
- Net cashflow is the amount of capital required to invest in this risk-free portfolio

All risk-free portfolios must earn the same interest rate (r)

- Market forces will force the interest rate of any two risk-free portfolios to be the same
- Otherwise Arbitrage will occur:
- People buying and selling the products in the same time period to earn quick profit
- Risk free portfolio must earn interest rate r

$$\Delta P = P * \Delta t * r$$

$$-\left(\frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial S^2}\sigma^2 s^2\right)\Delta t = P * \Delta t * r$$

$\vdots$ $(Manipulation)$

$$\frac{\partial G}{\partial t} + rS\frac{\partial G}{\partial S} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 G}{\partial S^2} = rG$$

This is known as the Black-Scholes Equation.

Theoretical price of a European options

Solving the BS equation,

$$c = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

$$p = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1)$$

$$d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$

Important notes:

- Variable μ does not appear in any of the equations → Price does not matter on risk preference
- As S_0 becomes very large, we can almost be sure that you will exercise the call option and earn profit equal to $S_0 - Ke^{-rT}$ (Maximum willingness to pay for option)
- e^{-rT} is the discount rate for what the strike price value would be in the current time period