Class Notes

STAT410 - Introduction to Probability Theory
Professor Jonathan Francis Fernandes
University of Maryland, College Park

Wenyu Chen

Summer 2022

Contents

1 | Axioms of Probabilities

Basic Definition of Probabilities

Definition 1.0.0.0.1
Experiment: *is repeatable task with well defined outcomes.*
Sample Space: *is a collection of all possible outcomes of the experiment.*
Event: *is a subset of the sample space.*

Example: Suppose we toss a coin three times, assume coins lands on H or T.
Experiment: tossing the coin three times, noting the outcomes.
Sample Space: {HHH, HHT,HTH, THH, HTT,THT,TTH, TTT }
Example of events: E_1 = getting all heads = $\{HHH\} \subseteq S$
E_2 = getting exactly one H and one T = $\{\} \subseteq S$
E_3 = getting at least two heads = $\{HHH, HHT, HTH, THH\} \subseteq S$

We want to assign a number to each event, which is a measure of the chance or probabilities that this event happened. Our goal is to understand the process of assigning probabilities to events.

Example: **Die Roll**
Experiment: Roll a six sided dice two times
Sample Space:$\{(1,1),(1,2)...(1,6),(2,1)...(2,6)...(6,6)\}$
Example of events:
E_1 = at least one six = $\{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6),(6,1),(6,2),(6,3),(6,4),(6,5)\}$
$|E_2| = 10$
E_2 = same numbers on both rolls = $\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$
$E_4 = \{(1,5),(3,4)\}$

Note:
Simple Event: is the event with exactly one outcomes. It is the cardinality of the sample space.
Number of events: Assume the sample space is finite of size n, the number of events is the cardinality of all possible outcomes of S = 2^n

Set in context
Let an event E be describes as a subset of the sample space S.
Let E, F two events be given.
$E \cap F$ is a new event that corresponds to outcomes in both E and F
$E \cup F$ is a outcome a new event include outcomes in E or F
E^c is a new event that include outcomes where E doesn't happen.

When we say an event E has happened, we means that the outcome ω of the experiment lies inside E
Example: Tossed a coin three times. In one run of the experiment, the result is HHT.
HHT $\in E_1$ = at least two heads, we said E_1 has happened.

Definition 1.0.0.0.2
A sigma algebra $\mathscr{B}$ for a set S is a collection of subset of S that satisfies:

1. $\emptyset \in \mathscr{B}$
2. *If $A \in \mathscr{B}$ then $A^c \in \mathscr{B}$ (closed under complement)*
3. *If $\{A_i\}_{i=1}^{\infty} \in \mathscr{B}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{B}$ (closed under countable unions)*

Note: Sigma algebra is a collection of events to which we want to assign the probabilities.
Example: Recall back to the dice roll example. Let B be the power set of its sample space. Then

note it is the sigma algebra of the sample space. The size will be $2^{|s|} = 236$
Example: : Let event E of S be given. Then its sigma algebra $B = \{\emptyset, E, E^c, S\}$

Axioms for a probability function

Definition 2.0.0.0.1
Suppose we are given the pair $(S, \mathscr{B})$, where S represents the sample space and $\mathscr{B}$ is a sigma algebra S. A ***probability function*** *P satisfies the following:*

$$P : \mathscr{B} \to \mathbb{R}$$
$$E \mapsto P(E)$$

1. *$P(A) \geq 0$ for all $A \in \mathscr{B}$*
2. *$P(S) = 1$*
3. *If $A_1, A_2, A_3, \ldots$ are mutually disjiont sets in $\mathscr{B}$ then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$*

Theorem 2.1
Properties of a probability function: *Suppose $P : \mathscr{B} \to \mathbb{R}$ is a probability function. Then*

1. *$P(\emptyset) = 0$*
2. *$P(A) \leq 1$ for all $A \in \mathscr{B}$*
3. *$P(A^c) = 1 - P(A)$*
4. *$P(A \cap B^c) = P(A) - P(A \cap B)$*
5. *$P(A \cup B) = P(A) + P(B) - P(A \cap B)$*

Theorem 2.2
Suppose the sample space S is countable. To define a probability function on $(S, \mathscr{B} = P(S))$, we do the following

1. *Find a seq $\{p_i\}_{i=1}^{\infty}$ such that (i) $0 \leq p_i \leq 1, \forall i$ and (ii) $\sum_{i=1}^{\infty} p_i = 1$*
2. *Define $P(\{s_i\}) = p_i$*
3. *For any $E \in \mathscr{B}, E = \{s_{i1}, s_{i2}, \ldots s_{ik}\}$*

$$P(E) = \sum_{j=1}^{k} P(S_{ij}) = \sum_{j=1}^{k} P_{ij}$$

Example: Probability Functions Examples
Coin Tosseo:
Experiment: tossing the coin three times, noting the outcomes.
Sample Space: {HHH, HHT,HTH, THH, HTT,THT,TTH, TTT }
To get a probability function, we will need to work with a sigma algebra. Suppose $\mathscr{B}_1 = \mathscr{P}(S)$, note that the cardinality of $\mathscr{B}_1$ is 256.
There are infinitely many ways to choose the sequence p_i. One way is to choose $p_i = \frac{1}{8}$ for all i. Then we assign the probability.

S_i	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$P(S_i)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

It is clear that the third properties is satisfied. For example, if $E = \{HHT, HTH, TTT\}$, then

$$\begin{aligned} P(E) &= P(HHT) + P(HTH) + P(TTT) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= \frac{3}{8} \end{aligned}$$

Another way is to set $p_1 = 1$ and rest to 0. Then, we got

S_i	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$P(S_i)$	1	0	0	0	0	0	0	0

It is clear that the third properties is satisfied. For example, if $E = \{HHT, HTH, TTT\}$, then

$$\begin{aligned} P(E) &= P(HHT) + P(HTH) + P(TTT) \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned}$$

Alternate way to assign probabilities is to use the information about the experiment, and need to construct tree diagrams. **Tree Diagram** is a graph that describes the flow of the outcomes of each steps in an experiment.

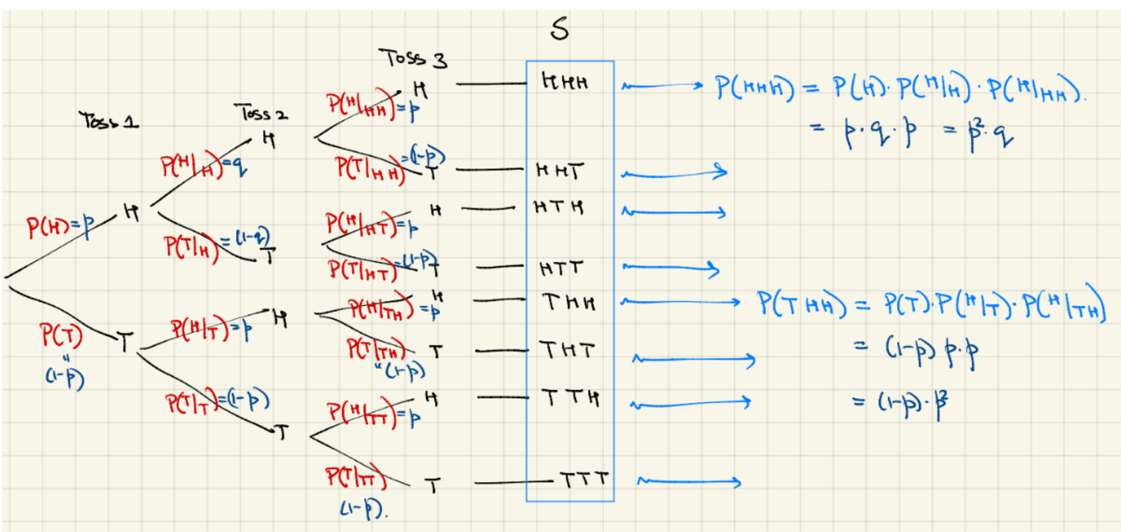
Definition 2.2.0.0.1
Conditional Probability *is* $P(A \mid B)$= *P(A happens given that B has already happened)*

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

And by ***Multiplication Principle****,*

$$\begin{aligned} P(A \cap B) &= P(A \mid B) \times P(B) \\ &= P(B \mid A) \times P(A) \end{aligned}$$

Experiment: Toss a fixed coin three times. We need to know that $P(H) = p \in (0, 1)$. Then we can got the tree diagram as



Then we can just assign p in (0,1). Then the rest can be assigned through the tree diagram.

Assumption: 1) Probability of events when all outcomes in the sample space are equally likely. 2) Sample space is finite.

Proposition 2.2.1
If every outcome in the sample space is equally likely, we can calculate the probability of $E \subseteq S$ *as follow:*

$$P(E) = \frac{n(E)}{n(S)}$$

where $n(E)$ *is the number of outcomes in* E*.*

Example: Suppose we toss a coin that $p(H) \in (0, 1)$ two times. The sample space is S = {HH, HT, TH, TT}. If the coin is fair, then $P(HH) = P(HT) = P(TH) = P(TT) = 0.25$

Theorem 2.3
Fundmental Theorem of Counting: *Suppose a task* T *can be performed as a sequence of subtasks:* $T_1, T_2, T_3, ..., T_k$*. And each* $n_1, ..., n_2, n_3, ..., n_k$ *is number of ways to perform* T_i
Then the total number ways to perform the task T *is*

$$n_1 \times n_2 \times n_3 \times \cdots \times n_k$$

Typically we will have to select k objects from n distinct objects.
Example: $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, we might be interested in knowing the total number of ways one can choose 4 digits from this 10 digits.

	Without Replacement	With Replacement
Order Matters	(1,2,4,5) different from (1,5,2,4) (1,1,2,5) is not possible	(1,2,4,5) different from (1,5,2,4) (1,1,2,5) is possible
Order Does Not Matters	(1,2,3,4) is same as (4,3,2,1) (1,1,2,5) not possible	(1,2,3,4) is same as (4,3,2,1) (1,1,2,4) is possible

1. Without replacement and order matters
Use the fundamental theorem of counting, we divide T, which is select k digits from a set of n distinct objects divide into

$$T : T_1 \to T_2 \to T_3 \to \cdots \to T_k$$

where T_i is select ith object. Then, we will got

$$n \times (n-1) \times (n-2) \times (n-3) \times \cdots \times (n-k+1)$$

Then, we got

$$^nP_k = \frac{n!}{(n-k)!}$$

2. Without replacement and order does not matter
T = choose k objects from n distinct objects where order does not matter and without replacement.

$$T : T_1 \to T_2$$

T_1 is choose k objects where order matters and without replacement
T_2 is to get rid of all the times we have double counted.

\# ways to do $T_1 = {}^nP_k = \frac{n!}{(n-k)!}$
\# ways to do T_2 = number of arrangements of k objects = $k!$
Then we got

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

3. With replacement and order does not matter
T = Choose k objects where order does not matter and with replacement.

We keep track of how many times a given object repeats in the selection and the total number of objects in the selection is equal to k. To do so, we can have $n+k-1$ spots with n-1 walls (since it is with replacement) as following

Notice that the x^n wall is always in the last place, so we only need to consider a length of $n+k-1$. And the space left represents number of objects in this set. By placing the wall differently, we will get different combination of objects with a total number of k.

So we need to decide from $n+k-1$ spots to determine which are the $n-1$ walls. Notice that the order doesn't matter and we don't have replacement. Therefore, we got

$$^{n+k-1}C_k = {}^{n+k-1}C_{n-1}$$

Example: $\{1, 2, 3, 4\}$, $k = 10$, We are selecting 10 objects from $\{1, 2, 3, 4\}$ with replacement and order does not matter.
We only care about how many times each number shows up since order does not matter and the total objects in selection are 10. To achieve that, we setup 13 spots. Such that, the extra position is the walls.

4.With Replacement
Choose k objects where order does matter and with replacement in n different things is

$$n^k$$

2 | Conditional Probability, Independence, and Bayes' Theorem

Recall: Given two events A and B, the conditional probability is

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

as long as $P(B) > 0$

And **Multiplication Principle** is

$$\begin{aligned} P(A \cap B) &= P(A \mid B) \cdot P(B) \\ &= P(B \mid A) \cdot P(A) \end{aligned}$$

Definition 0.0.0.0.1
*We say two events A and B are **independent** if and only if*

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

which is equivalent to saying that fact B occured does not affect the probability of A happening.

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

and equivalent to saying that the fact A occurred does not affect the probability of B happening

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

Example: Now if A and B are independent can show that A and B^c are also independent.
Hint: note that $P(A \cap B^c)$ can be write in term of $P(A), P(B), P(A \cap B)$

Proof.

$$\begin{aligned} (A \cap B^c) &= P(A) + P(B^c) - P(A \cup B^c) \\ &= P(A) + P(B^c) - P(B^c \cup (A \cap B)) \\ &= P(A) + P(B^c) - (P(B^c) + (A \cap B)) \\ &= P(A) + P(B^c) - P(B^c) - (A \cap B)) \\ &= P(A) - (A \cap B)) \\ &= P(A) - P(A) \cdot P(B) \\ &= P(A)(1 - P(B)) \\ &= P(A) \cdot P(B^c) \end{aligned}$$

□

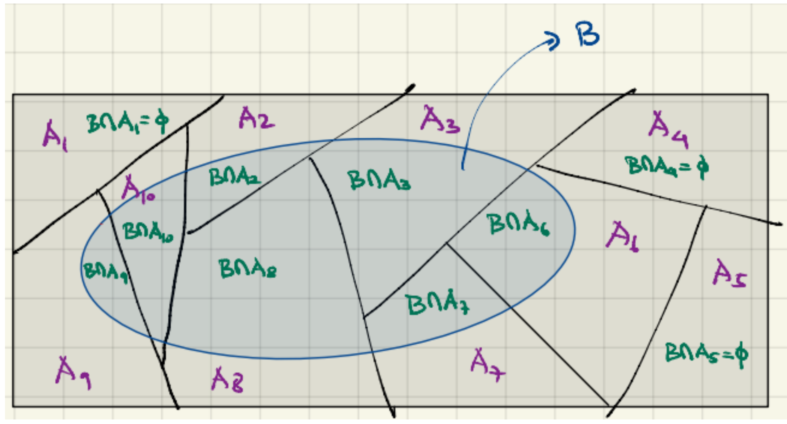
Theorem 0.1
The Law of Total Probability
If $A_1, A_2, A_3, ..., A_k$ is a partition of the sample space S, meaning

$$\text{(i) } A_i' s \text{ are mutually disjoint (ii) } \bigcup_{i=1}^{k} = S$$

Then for any event B in the sigma algebra associate with S

$$\begin{aligned} P(B) = P(B \cap S) = P\left(B \cap \left(\bigcup_{i=1}^{k} A_i\right)\right) \\ = \sum_{i=1}^{k} P(B \cap A_i) \\ = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_k) \end{aligned}$$

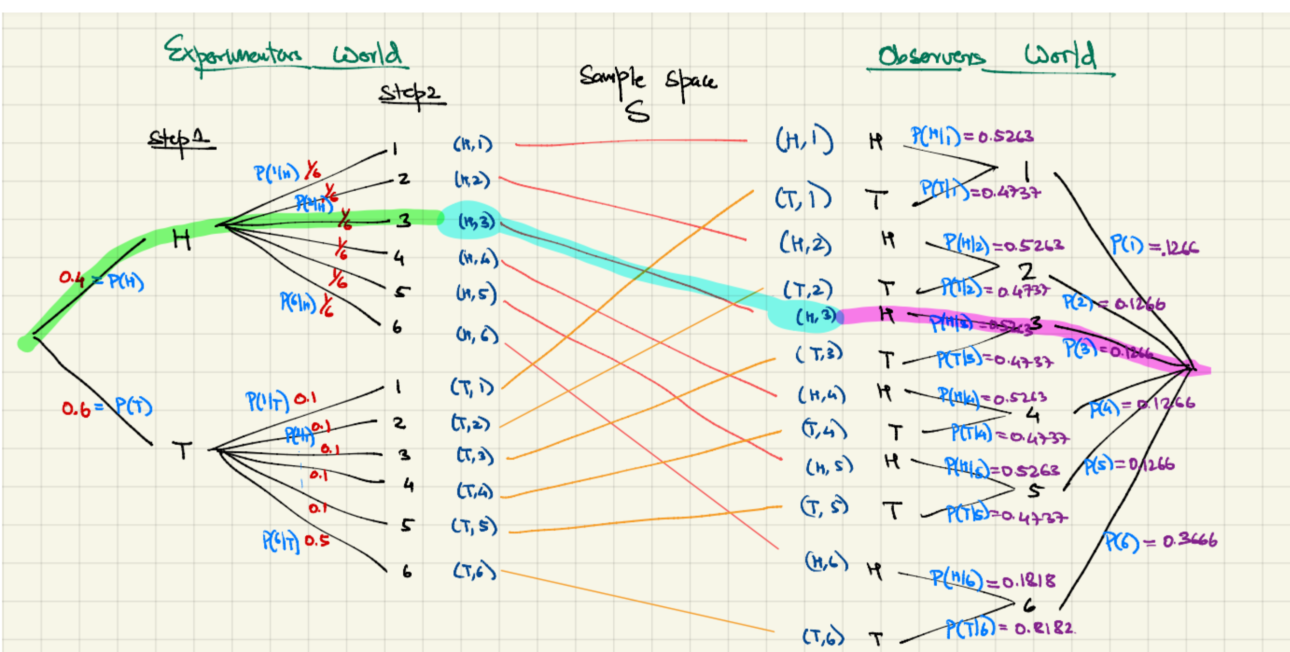
Visually,



Theorem 0.2
Bayes' Theorem:
If $A_1, A_2, A_3, ..., A_k$ is a partition of the sample space S, then for any event B in sigma algebra associated with S, and for any i,

$$
\begin{aligned}
P(A_i \mid B) &= \frac{P(A_i \cap B)}{P(B)} && \textit{by multiplication of principle} \\
&= \frac{P(A_i \cap B)}{\sum_{j=1}^{k} P(A_j \cap B)} && \textit{using the law of total probability} \\
&= \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{k} P(B \mid A_j) \cdot P(A_j)} && \textit{by multiplication of principle}
\end{aligned}
$$

Example: Assume there is an experiment
Step 1: Toss a coin with P(H)=0.4
Step 2: If H in step 1, roll a fair die.
If T in step 1, we roll an unfair die with $P(1) = P(2) = \cdots = P(5) = 0.1$ and $P(6) = 0.5$



If an observer see a 6, then

$$
\begin{aligned}
P(H \mid 6) &= \frac{P(H,6)}{P(6)} = \frac{P(H,6)}{P(H,6) + P(T,6)} \\
&= \frac{0.4 \times \frac{1}{6}}{0.4 \times \frac{1}{6} + 0.6 * 0.5} \\
&= 0.1818
\end{aligned}
$$

3 | Random Variables

Introduction to Random Variable

Definition 1.0.0.0.1
*A **random variable** is a function defined on a sample space S such that*

$$X : S \to \mathbb{R}$$
$$\omega \mapsto X(\omega) \in \mathbb{R}$$

Definition 1.0.0.0.2
Given X, we consider the set

$$\mathcal{X} := \textit{values of the random variable X.}$$

For $x \in \mathbb{R}$, we define

$$\begin{aligned}\{X = x\} &:= \{\omega \in S \mid X(w) = x\} \\ &= X^{-1}(x)\end{aligned}$$

Also,

$$\{X \leq x\} := \{\omega \in S \mid X(\omega) \leq x\}$$

and for $a, b \in \mathbb{R}$

$$\{a \leq x \leq b\}$$

and for $a, b \in \mathbb{R}$

$$\begin{aligned}\{a \leq X \leq b\} &:= \{\omega \in S \mid a \leq X(\omega) \leq b\} \\ &= X^{-1}([a, b])\end{aligned}$$

Example: **Experiment:** Toss a voin three times
Sample Space: {TTT, TTH, THT, HTT, THH, HTH, HHT, HHH}
Now given an outcome in S, want to attach a number. Then we define a random variable X

$$\begin{aligned}X &: S \to \mathbb{R} \\ \omega &\mapsto X(\omega) = \text{ number of H in } \omega \\ \mathcal{X} &= \{0, 1, 2, 3\}\end{aligned}$$

Then

$$\begin{aligned}\{x = 0\} &= \{TTT\} \\ \{x = 1\} &= \{HTT, THT, TTH\} \\ \{x = 2\} &= \{HHT, HTH, THH\} \\ \{x = 3\} &= \{HHH\}\end{aligned}$$

Then, we can define a probability function for the pair $(S, \mathscr{B})$ where the sigma algebra is generated by $< \{x = 0\}, \{x = 1\}, \{x = 2\}, \{x = 3\} >$. Here is the distribution table for P defined for the pair $(S, \mathscr{B})$

x	0	1	2	3
$P(X = \mathscr{X})$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Also, the distribution table for $(S, \mathscr{B}(\mathscr{P}(S)))$ is

ω	TTT	TTH	THT	HTT	THH	HTH	HHT	HHH
$P(\omega)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

The advantages of using first distribution table is 1) has less columns than the second distribution table, which makes visualizing and analyzing substantial easier. 2) It combines information that is relevant to the question we interested in. 3) X is a new sample space subset of $\mathscr{R}$ so we can use the properties of $\mathscr{R}$

Definition 1.0.0.0.3
Induced Probability Function:
Suppose we have a probability function P defined on $(S, \mathscr{P}(S))$. Then, if X is a random variable with values x, we can define

$$P_X(\{X = x\}) := \sum_{\omega \in \{X=x\}} P(\omega)$$

And for any subset $E \subseteq x$

$$\begin{aligned} P_X(E) &= P_X\left(\bigcup_{x \in E} \{X = x\}\right) \\ &= \sum_{x \in E} P(\{X = x\}) \end{aligned}$$

Definition 1.0.0.0.4
The Cummulative Distribution Function of X
Given a random variable $X : S \to \mathbb{R}$, the cumulative distribution function, is defined as :

$$F_X : \mathbb{R} \to \mathbb{R}$$
$$F_X(x) := P(X \leq x)$$

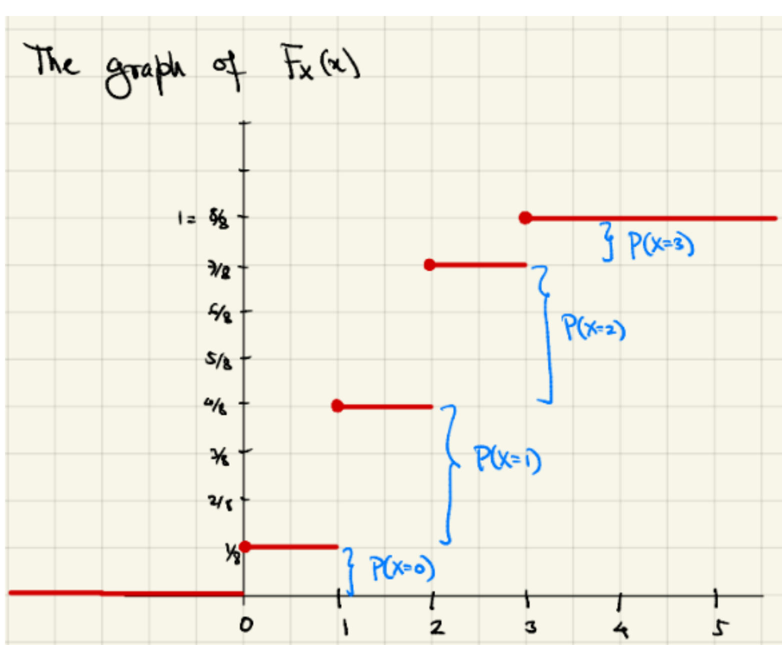
Definition 1.0.0.0.5
We say two random variables X and Y are identically distributed if they have the same cumulative distribution function, i.e

$$F_X(u) = F_y(u) \qquad \forall u \in \mathbb{R}$$

Example: Back to the previous example with X is the number of heads in the toss, we can get

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8} & x \in [0,1) \\ \frac{4}{8} & x \in [1,2) \\ \frac{7}{8} & x \in [1,2) \\ 1 & x \in [3,\infty) \end{cases}$$

And the graph of $F_X(x)$



Note: We say X is discrete if F_X is a step function.
We say X is continuous if F_X is continuous function.

Theorem 1.1
Classification of Cumulative Distribution Function For Random Variables
The $F(x)$ is a cumulative distribution function of a random variable if and only if the following condition hold

1. $\lim_{x\to\infty} F(x) = 1$ *and* $\lim_{x\to-\infty} F(x) = 0$ *is right continuous*
2. $F(x)$ *is a non-decreasing function*
3. $F(x)$ *is right continuous*

Note: Random Variable X can be discrete, continuous, or neither discrete nor continuous.

Expected Values, Variance, Moment Generating Function

Definition 2.0.0.0.1
Expected Value *of a random variable X is longterm average value X will take if experiment is performed repeatedly.*

Definition 2.0.0.0.2
Variance *is expected squared deviation of the values of X from its expected value.*
If the variance small, it will be more confident.

Theorem 2.1
*Suppose X is a **discrete** random variable, meaning the cumulative distribution function $F_X(x)$ is a step function.*
*The **Probability Mass Function** of X is*

$$p_X(x) := P(X = x) \qquad x \in \mathbb{R}$$

*The **Expected Values** of X is*

$$M_X := E(X) := \sum_{x\in X} x \cdot p_X(x)$$

if $h : \mathbb{R} \to \mathbb{R}$ is any function, then

$$E(h(X)) = \sum_{x\in X} h(x)p_X(x)$$

*The **Variance** of X is*

$$\begin{aligned} V(X) := \sigma_X^2 &= E((X - M_X)^2) \\ &= \sum_{x\in X} (x - M_X)^2 p_X(x) \end{aligned}$$

*The **Moment Generating Function** of X is*

$$M_X(t) = E(e^{tx}) := \sum_{x\in X} e^{tx} \cdot p_X(x)$$

Theorem 2.2
*Suppose X is a **continuous** random variable, meaning the cummulative distribution function $F_X(x)$ is continuous.*
*The **Probability Density Function** for X, a function $f_X : \mathbb{R} \to \mathbb{R}$ satisfies the following*

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)dt \qquad \forall\ x \in \mathbb{R}$$

then, for any $a, b \in \mathbb{R}$

$$P(a \le X \le b) = \int_a^b f_X(t)dt$$

Expected Value *of X is*

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

Variance *of X is*

$$\begin{aligned}\sigma_X^2 &:= V(X) := E((X - M_X)^2)\\ &= \int_{-\infty}^{\infty} (x - M_X)^2 f_X(x)dx\end{aligned}$$

Moment Generating Function of X *is*

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)dx$$

Theorem 2.3
Classification of Probability Density Function and Probability Mass Function
A function p(x) (or f(x)) is a pmf (or pdf) of a random variable X if and only if

1. $p(x) \geq 0$ *for all x (or* $f(x) \geq$ *for all x)*
2. $\sum_{x \in X} p_x(x) = 1$ *(or* $\int_{-\infty}^{\infty} f(x)dx = 1$*)*

Theorem 2.4
Suppose X is a random variable, the ***kth moment of X*** *is the expected value of* X^k*, i.e*

$$kth \text{ moment of } X = E(X^k)$$

And if X is a random variable with moment generation function $M_X(t)$*, then*

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

Theorem 2.5
Variance Formula:

$$\begin{aligned}V(X) &= E((X - M_x)^2)\\ &= E(X^2 - 2M_X X + M_X^2)\\ &= E(X^2) - E(2M_X X) + E(M_X^2)\\ &= E(X^2) - 2M_X E(X) + M_X^2 E(1)\\ &= E(X^2) - 2M_X M_X + M_X^2\\ &= E(X^2) - M_X^2 = E(X^2) - E(X)^2\end{aligned}$$

Example: **Experiment:** Toss a coin until a H appears where $P(H) = p \in (0, 1)$
$S = \{H, TH, TTH, TTTH, ...\}$ is infinite.

So we want to first find the cumulative distribution function of X first. Notice that for $x \in [k, k+1)$,

$$\begin{aligned}F_X(x) &= P(X = 1) + P(X = 2) + \cdots + P(X = k)\\ &= P(H) + P(TH) + P(TTH) + \cdots + P(T_{k-1}H)\\ &= p + (1-p)p + (1-p)^2 p + \cdots + (1-p)^{k-1} p\\ &= \sum_{i=1}^{k} (1-p)^{i-1} \cdot p\end{aligned}$$

It is a step function, so X is discrete. We will need to find the probability mass function of X, which is

$$p_X(x) = P(X = x)$$

Note that $p_X(x) = 0$ if $x < 1$.

Then, we need can check if

$$p_X(x) = \begin{cases}(1-p)^{x-1} \cdot p & x = 1, 2, 3, \cdots\\ 0 & \text{otherwise}\end{cases}$$

satisfies conditions for being a probability mass function.

Proof. It is clear that $p_X(x) \geq 0$ is true. So we need to prove $\sum_{x=1}^{\infty}(1-p)^{x-1} = 1 \quad (p \in (0,1))$

$$\begin{aligned}
\sum_{x=1}^{\infty}(1-p)^{x-1} &= p\sum_{x=1}^{\infty}(1-p)^{x-1} \\
&= p \cdot \left(\lim_{k\to\infty}\sum_{x=1}^{k}(1-p)^{x-1}\right) \\
&= p \cdot \left(\lim_{k\to\infty}\sum_{x=0}^{k-1}(1-p)^{x}\right) \\
&= p \cdot \left(\lim_{k\to\infty}\left(\frac{1-(1-p)^{k-1+1}}{1-(1-p)}\right)\right) \quad \text{By Gemetric Sum } \sum_{i=0}^{n} r^i = \frac{1-r^{n+1}}{1-r}, \quad r \neq 1 \\
&= p \cdot \left(\lim_{k\to\infty}\left(\frac{1-(1-p)^{k}}{p}\right)\right) \\
&= \lim_{k\to\infty}\left(1-(1-p)^k\right) \\
&= 1 - \lim_{k\to\infty}(1-p)^k \\
&= 1 - 0 = 1
\end{aligned}$$

So $p_X(x)$ is a probability mass function of a discrete random variable! $\square$

4 | Discrete Random Variables

Suppose X is a discrete random variable and x is a set of values of X. Then x can either be (i) a finite set (ii) a countably infinite set.

Uniform Discrete Distribution

Definition 1.0.0.0.1
*We say a random variable X with parameter N has the **uniform discrete distribution** if and only if*

$$X = \{1, 2, 3, \cdots, N\}$$
$$p_X(x) = \frac{1}{N}, \forall x \in X$$

First, let's calculate the **E(x)**

$$\begin{aligned}
E(X) &= \sum_{x \in X} x \cdot p_X(x) \\
&= \sum_{i=1}^{N} i \cdot \frac{1}{N} \\
&= \frac{1}{N} \sum_{i=1}^{N} i \\
&= \frac{1}{N} \cdot \frac{N(N+1)}{2} \\
&= \frac{N+1}{2}
\end{aligned}$$

Then, we calculate the **V(x)**, note that $V(X) = E(X^2) - E(X)^2$, so we want to calculate the $E(X^2)$

$$\begin{aligned}
E(X^2) &= \sum_{x \in X} x^2 \cdot p_X(x) \\
&= \sum_{i=1}^{N} i^2 \cdot \frac{1}{N} \\
&= \frac{1}{N} \sum_{i=1}^{N} i^2 \\
&= \frac{1}{N} \cdot \frac{N(N+1)(2N+1)}{6} \\
&= \frac{(N+1)(2N+1)}{6}
\end{aligned}$$

Then,

$$\begin{aligned}
V(X) &= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 \\
&= \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} \\
&= \frac{2(N+1)(2N+1) - 3(N+1)^2}{12} \\
&= \frac{(N+1)(2(2N+1) - 3(N+1))}{12} \\
&= \frac{(N+1)(4N+2-3N-3))}{12} \\
&= \frac{(N+1)(N-1)}{12}
\end{aligned}$$

Now we can calculate the **moment generating function**

$$\begin{aligned}
M_X(t) &= \sum_{x \in X} e^{tx} \cdot p_X(x) \\
&= \sum_{k=1}^{N} e^{tk} \cdot p_X(x) \\
&= \sum_{k=1}^{N} e^{tk} \cdot \frac{1}{N} \\
&= \frac{1}{N} \sum_{k=1}^{N} (e^t)^k \\
&= \frac{1}{N} \cdot \left(\frac{1-(e^t)^{N+1}}{1-e^t} - 1 \right)
\end{aligned}$$

Example: **Uniform Discrete Distribution Code in R**

```
uniform_discrete <- function(N){
  uni_values <- 1:N
  sample(x= uni_values,
         size = 1,
         replace = TRUE)
}
```

Binomial Distribution

Definition 2.0.0.0.1
We say a random variable X with parameters

$$\begin{aligned}
n &\to \textit{Sample Size} \\
p &\to \textit{Probability of getting a success}
\end{aligned}$$

*has the **Binomial Distribution** if*

$$\begin{aligned}
X &= \{0, 1, 2, \cdots, N\} \\
\forall k \in X, p_X(k) &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}
\end{aligned}$$

Theorem 2.1
Binomial Theorem

$$\begin{aligned}
(x+y)^n &= \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\
&= \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}
\end{aligned}$$

Note: The experiment is performing n-independent with exactly two outcomes: success, failure. And p(success)$\in (0, 1)$

First, we want to calculate the $p_X(k)$ Recall $X = \{0, 1, 2, 3, \cdots, n\}$, $P(X = k) = P($ there are exactly k success in n independent trials.
$|\{X = k\}$ = number of ways to select n objects from the set $\{S, F\}$. Think of in n boxes, we choose k boxes that will contain the S. This is $\binom{n}{k}$ ways.
The number of all subsets of a set containing n elements is 2^n
Note: $\bigcup_{i=0}^{k} |\{X = k\}| = 2^n$

Note that each way has $(p)^k(1-p)^{n-k}$ probability, so the total probability and the probability mass function is

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

Then, we can check if P_X is a probability mass function. Since $\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \geq 0$ all the time, so we need check $\sum_{x \in X} p_X(x) = 1$

$$\begin{aligned}
\sum_{x \in X} p_X(x) &= \sum_{k=0}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= (p + (1-p))^n && \text{By Binomial Theorem} \\
&= 1^n \\
&= 1
\end{aligned}$$

Now, we can calculate the **Expected Value** of X∼Binom(n, p)

$$\begin{aligned}
E(X) &= \sum_{x \in X} x \cdot p_X(x) = \sum_{k=0}^{n} k \cdot p_x(k) \\
&= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \cdot \frac{n!}{(n-k)!k!} p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \cdot p \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} \cdot (1-p)^{n-k} \\
&= n \cdot p \cdot \sum_{k=1}^{n} \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} \cdot (1-p)^{n-k} \\
&= n \cdot p \cdot \sum_{r=0}^{n-1} \frac{(n-1)!}{(n-r-1)!(r)!} \cdot p^r \cdot (1-p)^{n-r-1} && \text{set } r = (k-1) \\
&= n \cdot p \cdot \sum_{r=0}^{n-1} \frac{(n-1)!}{((n-1)-r)!(r)!} \cdot p^r \cdot (1-p)^{(n-1)-r} \\
&= n \cdot p \cdot \sum_{r=0}^{n-1} \binom{n-1}{r} \cdot p^r \cdot (1-p)^{(n-1)-r} \\
&= np
\end{aligned}$$

Note: $\sum_{r=0}^{n-1} \binom{n-1}{r} \cdot p^r \cdot (1-p)^{(n-1)-r}$ is a probability mass function of Binom$(n-1, p)$, which is the sum of all probability from Binom$(n-1, p)$. By the definition of a probability mass function, it is 1.

Then, let's calculate the **Variance**

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \cdot p_X(k) = \sum_{k=0}^{n} k^2 \cdot \binom{n}{k} p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \cdot \frac{n!}{(n-k)!k!} p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \cdot \frac{n!}{(n-k)!(k-1)!} p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \cdot n \cdot p \frac{(n-1)!}{(n-k)!(k-1)!} p^{k-1} \cdot (1-p)^{n-k} \\
&= n \cdot p \sum_{k=1}^{n} k \cdot \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{n-k} \\
&= n \cdot p \sum_{k=1}^{n} (r+1) \cdot \binom{n-1}{r} \cdot p^r \cdot (1-p)^{(n-1)-r} && \text{Set } r = k-1 \\
&= n \cdot p \left(\sum_{r=0}^{n-1} r \binom{n-1}{r} p^r \cdot (1-p)^{(n-1)-r} + \sum_{r=0}^{n-1} \binom{n-1}{r} p^r \cdot (1-p)^{(n-1)-r} \right) \\
&= n \cdot p((n-1)p + 1) \\
&= n(n-1)p^2 + np
\end{aligned}
$$

Then the **variance** is

$$
\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= np((n-1)p + 1 - np) \\
&= np(np - p + 1 - np) \\
&= np \cdot (1-p)
\end{aligned}
$$

Then, we can calculate the **Moment Generating Function**

$$
\begin{aligned}
M_X(t) &= E(e^{tx}) \\
&= \sum_{k=0}^{n} e^{tk} \cdot p_X(k) \\
&= \sum_{k=0}^{n} e^{tk} \cdot \binom{n}{k} p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=0}^{n} e^{tk} \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} \cdot (e^t p)^k \cdot (1-p)^{n-k} \\
&= (pe^t + (1-p))^n && \text{By Binomial Theorem}
\end{aligned}
$$

Using the moment generating function, we can find the $E(X)$. But first we need to find

$$
\begin{aligned}
\frac{d}{dt}(M_X(t)) &= \frac{d}{dt}((pe^t + (1-p))^n) \\
&= n \cdot ((pe^t + (1-p))^{n-1} \cdot pe^t && \text{Chain Rule}
\end{aligned}
$$

Then

$$
\begin{aligned}
E(X) &= \frac{d}{dt}(M_X(t))\big|_{t=0} \\
&= n \cdot ((pe^t + (1-p))^{n-1} \cdot pe^t\big|_{t=0} \\
&= n \cdot ((p + (1-p))^{n-1} \cdot p \\
&= np
\end{aligned}
$$

Special Case: Bernoulli Distribution
When $n = 1$, $X = \{0, 1\}, p \in (0, 1)$, then

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

Hyper Geometric Distribution

Definition 3.0.0.0.1
We say a random variable X with parameters

$$\begin{aligned} N &= \textit{population size} \\ M &= \textit{number successes in the population} \\ n &= \textit{sample size} \end{aligned}$$

*is a **hypergeometric distribution** if and only if*

$$\begin{aligned} X &= \{0, 1, 2, \cdots, min(n, M)\} \\ p_X(k) &= \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} \end{aligned}$$

In this setting, X is the number of successes in a sample of size n, sampled from a population of size N with M success and sampling without replacement.
$|\{X = k\}|$ = number elements in this set, the number of ways to sample out a sample with exactly k success.
This is same as first find the exactly k success in M successes. And select $(n-k)$ spots with Failure, then fill the rest to failure.

$$\binom{M}{k}\binom{N-M}{n-k}$$

And the total number of ways to n objects from N without replacements is $\binom{N}{n}$, so we got

$$\begin{aligned} p_X(k) = P(X = k) &= \frac{|\{X = k\}|}{|S|} \\ &= \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}} \end{aligned}$$

The **Expected Value** is

$$E(X) = n \cdot \left(\frac{M}{N}\right)$$

The **Variance** is

$$V(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

Example: **Experiment:** Suppose a bag has N balls. M balls are Blue, $(N - M)$ balls are Red. We draw n balls from this bag.
If we are sampling with replacement, then it is a **binomial distribution** with parameters

$$\begin{aligned} n &\rightarrow \text{Sample Size} \\ p = \frac{M}{N} &\rightarrow \text{Probability of getting a success} \end{aligned}$$

If we are sampling without replacement, then it is a **Hypergeometric Distribution** with parameters

$$\begin{aligned} N &\rightarrow \text{Population} \\ M &\rightarrow \text{Number success in the population} \\ n &\rightarrow \text{Sample Size} \end{aligned}$$

Geometric Distribution

Definition 4.0.0.0.1
We say X is a geometric distribution if

$$X = \{1, 2, 3, 4, \cdots\}$$
$$p_X(x) = p(1-p)^{x-1}, x \in X$$

Where $X =$ *number of trials until a success.*

We can find the moment generating function of X

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \sum_{x\in X} e^{tx} \cdot p_X(x) \\
&= \sum_{k=1}^{\infty} e^{tk} \cdot (1-p)^{k-1} \cdot p \\
&= p \cdot \sum_{k=1}^{\infty} e^{t(k-1)+t} \cdot (1-p)^{k-1} \\
&= p \cdot \sum_{k=1}^{\infty} e^{t} \cdot e^{t(k-1)} \cdot (1-p)^{k-1} && a^x \times a^y = a^{x+y} \\
&= p \cdot e^t \cdot \sum_{k=1}^{\infty} e^{t(k-1)} \cdot (1-p)^{k-1} \\
&= p \cdot e^t \cdot \sum_{k=1}^{\infty} (e^t(1-p))^{k-1} && (ab)^x = a^x b^x \\
&= p \cdot e^t \cdot \sum_{n=0}^{\infty} (e^t(1-p))^{n} && n = k-1 \\
&= p \cdot e^t \cdot \left(\frac{1}{1-e^t(1-p)}\right) && \sum_{n=0}^{\infty} a^n = \frac{1}{1-a} \text{ if } |a| < 1, \text{ else diverge} \\
&= \frac{p \cdot e^t}{1-e^t(1-p)}
\end{aligned}$$

Therefore,

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

By theorem **2.4** in chapter 3, we can find the expected value of X using the moment generating function of X.

$$\begin{aligned}
\frac{d}{dt}M_X(t) &= \frac{d}{dt}\left(\frac{p \cdot e^t}{1-e^t(1-p)}\right) \\
&= \frac{pe^t \cdot (1-e^t(1-p)) - pe^t \cdot (-e^t(1-p))}{(1-e^t(1-p))^2} && \text{quotient rule and difference rule}
\end{aligned}$$

We need to evaluating at $t = 0$

$$\begin{aligned}
M_X &= E(X) \\
&= \left(\frac{d}{dt}M_X(t)\right)\Big|_{t=0} \\
&= \frac{p(1-(1-p)) - pe^t \cdot (-(1-p))}{(1-(1-p))^2} \\
&= \frac{p^2 + p - p^2}{p^2} \\
&= \frac{p}{p^2} = \frac{1}{p}
\end{aligned}$$

We can also calculate the $E(X)$ without using moment generation function. Refers back to **theorem 2.1** in chapter 3, we will get the same result as above

$$
\begin{aligned}
E(X) &= \sum_{x \in X} x \cdot p_X(x) \\
&= \sum_{k=1}^{\infty} k \cdot p_X(x) \\
&= \sum_{k=1}^{\infty} k \cdot (1-p)^k p \\
&= p \cdot \sum_{k=1}^{\infty} k(1-p)^{k-1} \\
&= p \cdot \sum_{k=0}^{\infty} k(q)^{k-1} && \text{set } q = 1-p \\
&= p \cdot \frac{d}{dq} \sum_{k=0}^{\infty} (q)^k && \frac{d}{dx}(x^n) = nx^{n-1} \\
&= p \cdot \frac{d}{dq} \left(\frac{1}{1-q} \right) = p \cdot \frac{d}{dq} \left((1-q)^{-1} \right) \\
&= p \cdot -(1-q)^{-2} \cdot -1 && \text{Chain Rule } \frac{d}{dx}[f(g(x))] = f'(g(x))g'(x) \\
&= p \cdot \frac{1}{(1-(1-p))^2} \\
&= \frac{p}{p^2} = \frac{1}{p}
\end{aligned}
$$

Using **Theorem 2.5** in chapter 3, we can calculate the the Variance $V(x)$, first we need to calculate $E(x^2)$

$$
\begin{aligned}
\frac{d}{dt}\frac{d}{dt} M_X(t) &= \frac{d}{dt} \frac{pe^t \cdot (1 - e^t(1-p)) - pe^t \cdot (-e^t(1-p))}{(1-e^t(1-p))^2} \\
&= \frac{d}{dt} \frac{pe^t - pe^t \cdot e^t(1-p)) - pe^t \cdot -e^t(1-p))}{(1-e^t(1-p))^2} \\
&= \frac{d}{dt} \frac{pe^t - pe^t \cdot e^t(1-p)) + pe^t \cdot e^t(1-p))}{(1-e^t(1-p))^2} \\
&= \frac{d}{dt} \frac{pe^t}{(1-e^t(1-p))^2} \\
&= \frac{pe^t(1-e^t(1-p))^2 - pe^t \cdot 2(1-e^t(1-p)) \cdot (-e^t(1-p))}{(1-e^t(1-p))^4}
\end{aligned}
$$

Then, we evaluating at $t = 0$

$$
\begin{aligned}
E(X^2) &= \left(\frac{d}{dt}\frac{d}{dt} M_X(t) \right) \Bigg|_{t=0} \\
&= \frac{p(1-(1-p))^2 - p \cdot 2(1-(1-p)) \cdot (-(1-p))}{(1-(1-p))^4} \\
&= \frac{p \cdot p^2 - p \cdot 2p \cdot (-1+p))}{(p^4} \\
&= \frac{p^2(p - 2 \cdot (-1+p))}{p^4} \\
&= \frac{(p - 2 \cdot (-1+p))}{p^2} \\
&= \frac{(p + 2 \cdot (1-p))}{p^2}
\end{aligned}
$$

Refers back to **Theorem 2.5** in chapter 3

$$
\begin{aligned}
V(X) &= E(X^2) - (E(x))^2 \\
&= \frac{(p + 2 \cdot (1-p))}{p^2} - \frac{1}{p^2} \\
&= \frac{p + 2 - 2p - 1}{p^2} \\
&= \frac{p + 1 - 2p}{p^2} = \frac{1-p}{p^2}
\end{aligned}
$$

The Negative Binomial Distribution

From now, the examples can take infinitely many values. In particular, the distributions with infinite values are calculating possibilities of events where one is waiting for something to happen.

Definition 5.0.0.0.1
We say X with parameters

$$p = \textit{Probability of a success}$$
$$r = \textit{Number of successes we are waiting for}$$

has the ***Negative Binomial Distribution*** *if*

$$\begin{aligned} X &= \{0, 1, 2, 3, \cdots\} \\ p_X(k) &= \binom{k+r-1}{r-1} p^r \cdot (1-p)^k \\ &= \binom{k+r-1}{k} p^r \cdot (1-p)^k \end{aligned}$$

Theorem 5.1
Sum of Negative Binomial Series:

$$(1-w)^{-r} = \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} w^k$$

To understand Negative Binomial Distribution and how we got the probability mass function, we could consider the following experiment.

Experiment: Keep tossing a coin independently, fix $r \in \mathbb{R}$, $p(H) \in (0, 1)$
$X =$ number of tails until exactly r heads have appeared
OR: Perform a trial whose outcomes are successes independently, $p(S) \in (0, 1)$
$X =$ number of failure until exactly r successes have appeared

We want to calculate the probability mass function of X. Note that the values of $\{X = x\} = \{0, 1, 2, 3, \cdots, x\}$

For example, If $r = 2$
$\{X = 0\} = \{$ All outcomes with 0 failures until 2 successes$\} = \{SSS\}$
$\{X = 0\} = \{$ All outcomes with 1 failures until 2 successes$\} = \{FSS, SFS\}$

Observe: All outcomes in the set $\{X = k\}$ is equally likely, so we first, We find the probability of a single outcome in $\{X = k\}$

$$\{X = k\} = \{\text{All outcomes with exactly k failures before the } r^{th} \text{ success}\}$$

Then, we consider a single event ω in the set $\{X = k\}$

$$\begin{aligned} \omega &= FFF \cdots F_k SSS \cdots S_r \\ \text{Since } P(S) &= p, \ P(F) = (1-p) \\ P(\omega) &= p^k p^r \end{aligned}$$

Then, we want to count the $|\{X = k\}|$.
Since we need $k + r$ trials for any outcomes in $\{X = k\}$, we can think of having $k + r$ slots, whereas the last slot $k + r$ will always be S. Then the remaining $(r - 1)$ successes can happen in any of the remaining $(k + r - 1)$ slots.
Therefore we only need to choose $(r - 1)$ slots out of $(k + r - 1)$ to put the success, which is a total of

$$\binom{k+r-1}{r-1}$$

or, if we choose the failure seats, we will get

$$\binom{k+r-1}{k}$$

Therefore, we find that

$$|\{X = k\}| = \binom{k+r-1}{r-1} = \binom{k+r-1}{k} = \frac{(k+r-1)!}{(r-1)!k!}$$

So the probability mass function is

$$p_X(k) = \binom{k+r-1}{r-1} p^r \cdot (1-p)^k$$

Now, suppose $X \sim$NegBinom(p, r), we want to calculate the **Expected Value** E(X)

$$\begin{aligned}
E(X) &= \sum_{x \in X} p_X(x) \\
&= \sum_{k=0}^{\infty} k \cdot p_X(k) \\
&= \sum_{k=0}^{\infty} k \cdot \binom{k+r-1}{k} \cdot p^r \cdot (1-p)^k \\
&= \sum_{k=0}^{\infty} k \cdot \frac{(k+r-1)!}{(r-1)!k!} \cdot p^r \cdot (1-p)^k \\
&= \sum_{k=1}^{\infty} \frac{(k+r-1)!}{(r-1)!(k-1)!} \cdot p^r \cdot (1-p)^k \\
&= \sum_{j=0}^{\infty} \frac{(j+1+r-1)!}{(r-1)!(j)!} \cdot p^r \cdot (1-p)^{j+1} && \text{Set} j = k-1 \\
&= \sum_{j=0}^{\infty} r \cdot \frac{(j+(r+1)-1)!}{(r)!(j)!} \cdot \frac{p^{r+1}}{p} \cdot (1-p)^j \cdot (1-p) \\
&= \frac{r(1-p)}{p} \cdot \sum_{j=0}^{\infty} \frac{(j+(r+1)-1)!}{(r+1)-1)!(j)!} \cdot p^{r+1} \cdot (1-p)^j \\
&= \frac{r(1-p)}{p} \cdot \sum_{j=0}^{\infty} \binom{(j+(r+1)-1)}{j} \cdot p^{r+1} \cdot (1-p)^j \\
&= \frac{r(1-p)}{p}
\end{aligned}$$

Note that $\sum_{j=0}^{\infty} \binom{(j+(r+1)-1)}{j} \cdot p^{r+1} \cdot (1-p)^j$ is the sum of all the possibilities associated to NegBinom$(r+1, p)$

Similarly, we can find the **Variance** $V(X)$ to get

$$V(x) = \frac{r(1-p)}{p^2}$$

We can also find the **Moment Generating Function** is

$$\begin{aligned}
M_x(t) &= E(e^{tx}) \\
&= \sum_{k=0}^{\infty} e^{tk} \cdot p_X(k) \\
&= \sum_{k=0}^{\infty} e^{tk} \cdot \binom{k+r-1}{r-1} \cdot p^r \cdot (1-p)^k \\
&= p^r \cdot \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} \cdot (e^t(1-p))^k \\
&= p^r \cdot (1-(1-p)e^t)^{-r} && \text{Above is Negative Binomial Series} \\
&= \left(\frac{p \cdot e^t}{1-(1-p)e^t}\right)^r
\end{aligned}$$

Therefore, we got

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

Poission Distribution

Note:

$$
\begin{aligned}
e^{\lambda} &= 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \\
&= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}
\end{aligned}
$$

$$\therefore 1 = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \left(\frac{e^{-\lambda} \lambda^k}{k!} \right)$$

So we can use to define a probability mass function since it satisfies the requirement of probability mass function.

Definition 6.0.0.0.1
We say X with parameters

$$\lambda \to rate$$

*has the **Poission Distrubution** if*

$$
\begin{aligned}
X &= \{0, 1, 2, 3, \cdots\} \\
p_X(k) &= e^{-\lambda} \frac{\lambda^k}{k!}
\end{aligned}
$$

We can calculate the **Expected Value**

$$
\begin{aligned}
E(x) &= \sum_{k=0}^{\infty} k \cdot p_X(k) \\
&= \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{(k-1)!} \\
&= \sum_{k=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{(k-1)!} \\
&= \sum_{j=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{j+1}}{(j)!} && \text{Set} j = k - 1 \\
&= \sum_{j=0}^{\infty} e^{-\lambda} \cdot \lambda \cdot \frac{\lambda^j}{(j)!} \\
&= \lambda \cdot \sum_{j=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^j}{(j)!} \\
&= \lambda
\end{aligned}
$$

To calculate the **Variance** $V(x)$, we first need to find $E(X^2)$

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 \cdot p_X(k) = \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\
&= \lambda \sum_{k=1}^{\infty} k e^{-\lambda} \cdot \frac{-\lambda^k}{(k-1)!} \\
&= \lambda \sum_{j=0}^{\infty} (j+1) e^{-\lambda} \cdot \frac{-\lambda^k}{(j)!} && \text{Set} j = k - 1 \\
&= \lambda \sum_{j=0}^{\infty} j \cdot e^{-\lambda} \cdot \frac{-\lambda^k}{(j)!} + \lambda \sum_{j=0}^{\infty} e^{-\lambda} \cdot \frac{-\lambda^k}{(j)!} \\
&= \lambda \cdot \lambda + \lambda \\
&= \lambda^2 + \lambda
\end{aligned}
$$

Then, we can get

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

Finally, we can calculate the **Moment Generating Function** $M_X(t)$

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} e^{tk} \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(e^t \cdot \lambda)^k}{k!} \\ &= e^{-\lambda} \cdot e^{e^t \cdot \lambda} && \text{Note } \sum_{x=0}^{\infty} \frac{a^x}{x!} = e^a \\ &= e^{\lambda(e^t - 1)} \end{aligned}$$

5 | Continuous Random Variable

Recall that

Definition 0.0.0.0.1
Suppose X is a random variable with cumulative density function F_X, i.e

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

*We say X is a **continuous random variable** if and only if F_x is a continuous function.*

And if there exists a $f_X : \mathbb{R} \to \mathbb{R}$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \qquad \forall x \in \mathbb{R}$$

*We call f_X a **probability density function for X***

Note: Given a random variable X, there can be multiple probability density functions associated to X

Definition 0.0.0.0.2
We say two random variables, X, Y with cumulative density function F_X, F_Y respectively are identically distributed if

$$F_X(u) = F_Y(u) \qquad \forall u \in \mathbb{R}$$

Theorem 0.1
Given X a continuous random variable with cumulative density function $F_X(x)$ and suppose a probability density function $f_x(x)$ exists. Then

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x \cdot f_X(x)dx \\
V(X) &= \int_{-\infty}^{\infty} (x - M_X)^2 \cdot f_X(x)dx \\
M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)dx
\end{aligned}$$

if the integrals exists.

Theorem 0.2
Suppose X is a random variable, and $a, b \in \mathbb{R}$. Let $Y = aX + b$. Then

1. $E(Y) = aE(X) + b$
2. $V(Y) = a^2V(X)$
3. $M_Y(t) = e^{bt} \cdot M_X(at)$

Proof. Assume for simplicity that X is continuous with probability density function $f_X(x)$
1. $E(Y) = aE(X) + b$

$$\begin{aligned}
E(Y) = E(aX + b) &= \int_{-\infty}^{\infty} (aX + b)f_X(x)dx \\
&= \int_{-\infty}^{\infty} axf_x(x)dx + \int_{-\infty}^{\infty} bf_x(x)dx \\
&= a\int_{-\infty}^{\infty} xf_x(x)dx + b\int_{-\infty}^{\infty} f_x(x)dx \\
&= aE(x) + b
\end{aligned}$$

2. $V(Y) = a^2 V(X)$

$$\begin{aligned} V(Y) &= V(ax+b) = E((Y-E(Y))^2) \\ Y - E(Y) &= aX + b - (aE(X)+b) = aX - aE(X) = a(X-E(X)) \\ V(Y) &= E((a(X-E(X)))^2) \\ &= E(a^2(X-E(X))^2) \\ &= a^2 E((X-E(X))^2) \\ &= a^2 V(X) \end{aligned}$$

3.$M_Y(t) = e^{bt} \cdot M_X(at)$

$$\begin{aligned} M_Y(t) &= E(e^{Yt}) \\ &= E(e^{(ax+b)t}) \\ &= E(e^{axt} \cdot e^{bt}) \\ &= e^{bt} \cdot E(e^{X(at)}) \\ &= e^{bt} \cdot M_X(at) \end{aligned}$$

$\square$

Uniform Continuous Distribution

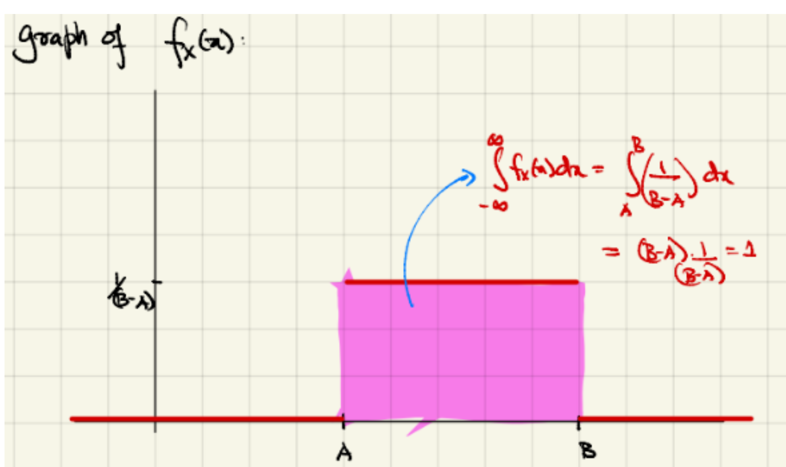
Definition 1.0.0.0.1
Uniform Continuous Distribution *has parameters*

$$A, B \rightarrow \textit{end points of an interval}$$

with probability density function

$$f_X(x; A, B) = \begin{cases} \frac{1}{B-A} & x \in [A, B] \\ 0 & otherwise \end{cases}$$

Here is the graph of $f_X(x)$



Now, we need to see that

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

if $x < A$, then $F_X(x) = 0$

if $x \in [A, B]$, then

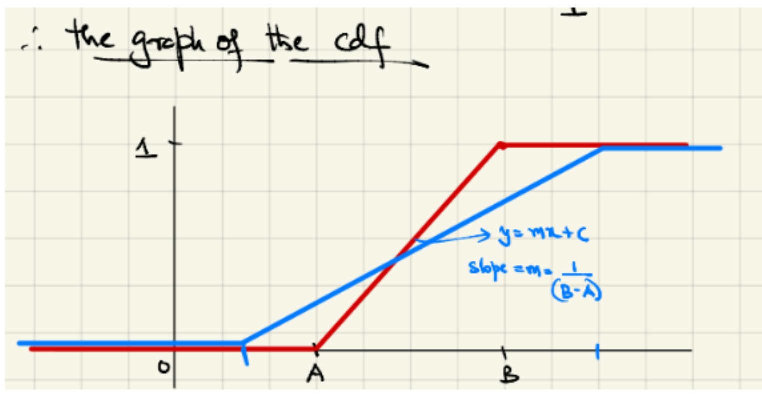
$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)dt = \int_A^x \frac{1}{(B-A)} dt \\ &= \frac{t}{(B-A)}\Big|_A^x = \frac{x-A}{B-A} \\ &= \frac{1}{B-A}x - \frac{A}{(B-A)} \end{aligned}$$

note that $\frac{1}{B-A}$ is the slope and $-\frac{A}{(B-A)}$ is the y-intercept.

if $x > B$,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(x)dx \\ &= \int_A^B \frac{1}{(B-A)} + \int_B^x 0dx \\ &= 1 \end{aligned}$$

The graph of the cumulative density function



Then, we can calculate the **Expected Value** $E(X)$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f_X(x)dx = \int_A^B x \cdot \frac{1}{B-A}dx \\ &= \left(\frac{1}{B-A}\right) \cdot \frac{x^2}{2}\bigg|_A^B = \left(\frac{1}{B-A}\right)\left(\frac{B^2-A^2}{2}\right) \\ &= \frac{1}{(B-A)} \frac{(B-A)(B+A)}{2} \\ &= \frac{B+A}{2} \end{aligned}$$

Then, to calculate the Variance, we first calculate the $E(X^2)$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \cdot f_X(x)dx \\ &= \int_A^B x^2 \cdot \frac{1}{B-A}dx \\ &= \left(\frac{1}{B-A}\right) \cdot \frac{x^3}{3}\bigg|_A^B \\ &= \left(\frac{1}{B-A}\right)\left(\frac{B^3-A^3}{3}\right) \\ &= \left(\frac{1}{B-A}\right)\frac{(B-A)(B^2+AB+A^2)}{3} \\ &= \frac{(B^2+AB+A^2)}{3} \end{aligned}$$

Then, we calculate the **Variance** $V(X)$

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= \frac{(B^2+AB+A^2)}{3} - \left(\frac{B+A}{2}\right)^2 \\ &= \frac{(B^2+AB+A^2)}{3} - \frac{B^2+2AB+A^2}{4} \\ &= \frac{4(B^2+AB+A^2)-3(B^2+2AB+A^2)}{12} \\ &= \frac{4B^2+4AB+4A^2-3B^2-3AB-3A^2}{12} \\ &= \frac{B^2-2AB+A^2}{12} = \frac{(B-A)^2}{12} \end{aligned}$$

Now, lets calculate the **Moment Generating Function** $M_X(t)$

$$\begin{aligned}
M_X(t) &= E(e^{tx}) \\
&= \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)dx \\
&= \int_A^B e^{tx} \cdot \frac{1}{(B-A)}dx \\
&= \frac{1}{(B-A)} \cdot \frac{e^{tx}}{t}\bigg|_A^B \\
&= \frac{1}{(B-A)} \cdot \frac{e^{Bt}-e^{At}}{t} \\
&= \frac{e^{Bt}-e^{At}}{t(B-A)}
\end{aligned}$$

Standard Normal Distribution

Definition 2.0.0.0.1
We say Z has the standard normal distribution if the probability density function of Z is

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}, z \in (-\infty, \infty)$$

Lets check that f_Z is a probability density function.

Proof.
i) $f_Z(z) \geq 0, \forall z \in \mathbb{R}$

$$\begin{aligned}
e^{\frac{-z^2}{2}} &\geq 0 \\
\frac{1}{\sqrt{2\pi}} &\geq 0 \\
e^{\frac{-z^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} &\geq 0
\end{aligned}$$

ii) $\int_{-\infty}^{\infty} f_Z(z)dz = 1$

We want to show that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}} dz = 1$. To do so, let's try the following

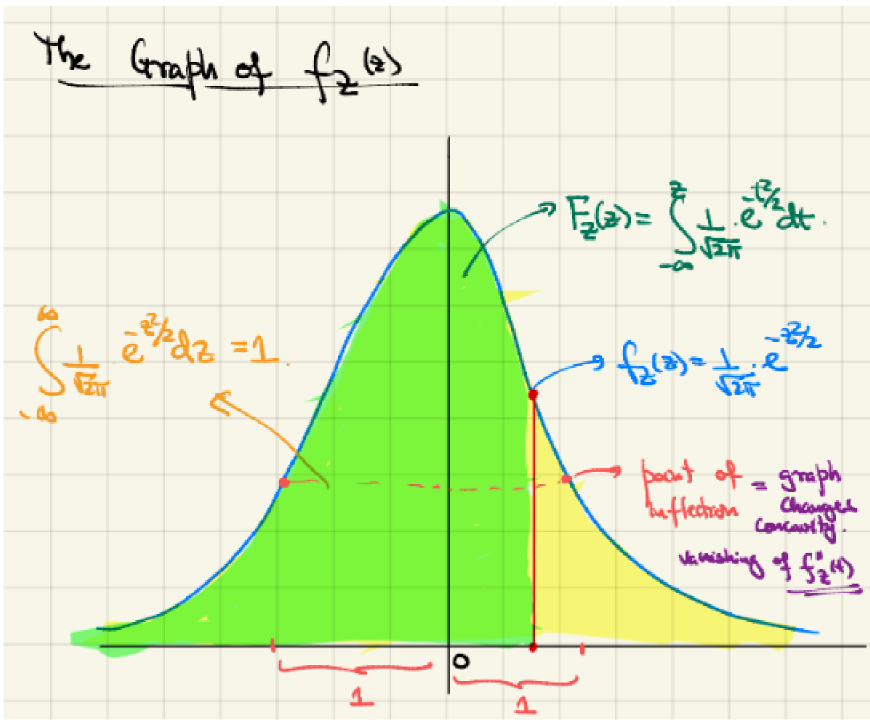
$$\begin{aligned}
\left(\int_{\infty}^{\infty} e^{\frac{-z^2}{2}} dz\right)^2 &= \left(\int_{\infty}^{\infty} e^{\frac{-z^2}{2}} dz\right)\left(\int_{\infty}^{\infty} e^{\frac{-z^2}{2}} dz\right) \\
&= \left(\int_{\infty}^{\infty} e^{\frac{-z^2}{2}} dz\right)\left(\int_{\infty}^{\infty} e^{\frac{-y^2}{2}} dy\right) \\
&= \int_{\infty}^{\infty}\int_{\infty}^{\infty} e^{\frac{-z^2}{2}} \cdot e^{\frac{-y^2}{2}} dzdy \\
&= \int_{\infty}^{\infty}\int_{\infty}^{\infty} e^{\frac{-(z^2+y^2)}{2}} dzdy \\
&\text{Then, we can change it to polar coordinate} \\
&= \int_0^{2\pi}\int_0^{\infty} e^{\frac{-r^2}{2}} rdrd\theta \qquad r^2 = r^2+y^2 \qquad dzdy = rdrd\theta \qquad 0 \leq \theta \leq 2\pi \\
&= \int_0^{2\pi}\int_0^{\infty} e^{\frac{-s}{2}} dsd\theta \qquad s = \frac{r^2}{2} \qquad ds = \frac{2r}{2}dr = rdr \\
&= \int_0^{2\pi} \left(\lim_{c\to\infty} \int_0^c e^{\frac{-s}{2}} ds\right) d\theta \\
&= \int_0^{2\pi} \left(\lim_{c\to\infty} \frac{e^{-s}}{-1}\bigg|_0^c\right) d\theta \\
&= \int_0^{2\pi} \left(\lim_{c\to\infty} \left(1-e^{-c}\right)\right) d\theta \\
&= \int_0^{2\pi} 1d\theta \\
&= 2\pi
\end{aligned}$$

So, we know that

$$\left(\int_{\infty}^{\infty} e^{\frac{-z^2}{2}dz}\right)^2 = 2\pi$$
$$\int_{\infty}^{\infty} e^{\frac{-z^2}{2}dz} = \sqrt{2\pi}$$
$$\int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}dz} = 1$$

Therefore, it is a probability density function. □

Here is visualizing the $f_Z(z)$



Also, the cumulative distribution function of the standard normal distribution is

$$F_Z(z) = \int_{-\infty}^{z} f_Z(t)dt = \int_{-\infty}^{z} \int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt$$

Then, we can calculate the **Expected Value** $E(Z)$.

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} z \cdot f_Z(z) dz \\ &= \int_{-\infty}^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \\ &= \int_{-\infty}^{0} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz + \int_{0}^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \end{aligned}$$

Then, let's first calculate the $\int_0^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$.

$$\begin{aligned} \int_{0}^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} &= \lim_{c\to\infty} \frac{1}{\sqrt{2\pi}} \int_{0}^{C} z \cdot e^{\frac{-z^2}{2}} dz \\ &= \lim_{c\to\infty} \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{-c^2}{2}} e^{u} - du \qquad \text{Set } u = \frac{-z^2}{2} \implies du = -zdz \quad -du = zdz \\ & \qquad z = 0 \implies u = 0 \qquad z = c \implies u = -\frac{c^2}{2} \\ &= \lim_{c\to\infty} \frac{1}{\sqrt{2\pi}} \left(-e^{u}\Big|_0^{\frac{-c^2}{2}}\right) \\ &= \lim_{c\to\infty} \frac{-1}{\sqrt{2\pi}} \left(-e^{\frac{-c^2}{2}} + e^{0}\right) \\ &= \lim_{c\to\infty} \frac{1}{\sqrt{2\pi}} \left(1 - \frac{1}{e^{\frac{c^2}{2}}}\right) \\ &= \frac{1}{\sqrt{2\pi}} \lim_{c\to\infty} \left(1 - \frac{1}{e^{\frac{c^2}{2}}}\right) = \frac{1}{\sqrt{2\pi}} \end{aligned}$$

Then, let's calculate the $\int_{-\infty}^{0} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$

$$
\begin{aligned}
\int_{-\infty}^{0} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} &= \lim_{c \to -\infty} \frac{1}{\sqrt{2\pi}} \int_{c}^{0} z \cdot e^{\frac{-z^2}{2}} dz \\
&= \lim_{c \to -\infty} \frac{1}{\sqrt{2\pi}} \int_{0}^{\frac{-c^2}{2}} e^u - du \qquad \text{Set } u = \frac{-z^2}{2} \implies du = -zdz \quad -du = zdz \\
& \qquad z = 0 \implies u = 0 \qquad z = c \implies u = -\frac{c^2}{2} \\
&= \lim_{c \to -\infty} \frac{1}{\sqrt{2\pi}} \left(-e^u \Big|_{\frac{-c^2}{2}}^{0} \right) \\
&= \lim_{c \to -\infty} \frac{1}{\sqrt{2\pi}} \left(-e^{\frac{-c^2}{2}} + e^0 \right) \\
&= \lim_{c \to -\infty} \frac{1}{\sqrt{2\pi}} \left(\frac{1}{e^{\frac{c^2}{2}}} - 1 \right) \\
&= \frac{1}{\sqrt{2\pi}} \lim_{c \to -\infty} (0 - 1) = \frac{1}{-\sqrt{2\pi}}
\end{aligned}
$$

Then,

$$
\begin{aligned}
E(Z) &= \int_{-\infty}^{0} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz + \int_{0}^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \\
&= \frac{1}{\sqrt{2\pi}} + \frac{1}{-\sqrt{2\pi}} = 0
\end{aligned}
$$

Then, we can calculate the **Variance**. Since $E(Z) = 0$, then $E(Z)^2 = 0$.

$$
\begin{aligned}
V(Z) &= E(Z^2) - E(Z)^2 \\
&= E(Z^2) \\
&= \int_{-\infty}^{\infty} z^2 \cdot f_Z(z) dz \\
&= \int_{-\infty}^{\infty} z^2 \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \cdot e^{\frac{-z^2}{2}} dz \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} z^2 \cdot e^{\frac{-z^2}{2}} dz + \int_{0}^{\infty} z^2 \cdot e^{\frac{-z^2}{2}} dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{0} z^2 \cdot e^{\frac{-z^2}{2}} dz + \int_{0}^{\infty} z^2 \cdot e^{\frac{-z^2}{2}} dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(uv|_{-\infty}^{0} - \int_{-\infty}^{0} vdu + uv|_{0}^{\infty} - \int_{0}^{\infty} vdu \right) \\
& \text{Use Integration By Parts } u = z \quad dv = ze^{\frac{-z^2}{2}} dz \\
& du = 1dz \quad v = -e^{\frac{-z^2}{2}} \\
&= \frac{1}{\sqrt{2\pi}} \left(z \cdot -e^{\frac{-z^2}{2}} \Big|_{-\infty}^{0} - \int_{-\infty}^{0} -e^{\frac{-z^2}{2}} dz + z \cdot -e^{\frac{-z^2}{2}} \Big|_{0}^{\infty} - \int_{0}^{\infty} -e^{\frac{-z^2}{2}} dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(z \cdot -e^{\frac{-z^2}{2}} \Big|_{-\infty}^{0} + \int_{-\infty}^{0} e^{\frac{-z^2}{2}} dz + z \cdot -e^{\frac{-z^2}{2}} \Big|_{0}^{\infty} + \int_{0}^{\infty} e^{\frac{-z^2}{2}} dz \right) \\
&= \frac{1}{\sqrt{2\pi}} \left((0 - 0) + (0 - 0) + \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}} dz \right) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \\
&= e^{\frac{t^2}{2}} \\
& \text{This is the probability density function of standard normal distribution} \\
&= 1
\end{aligned}
$$

Finally, we can calculate the **Moment Generating Function** $M_Z(t)$

$$\begin{aligned}
M_Z(t) &= E(e^{tz}) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tz} \cdot e^{-\frac{z^2}{2}} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\left(\frac{z^2}{2} - tz\right)} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\left(\frac{z^2}{2} - tz + \frac{t^2}{2} - \frac{t^2}{2}\right)} dz \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\left(\frac{z^2}{2} - tz + \frac{t^2}{2}\right)} \cdot e^{\frac{t^2}{2}} dz \\
&= e^{\frac{t^2}{2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\left(\frac{z^2}{2} - tz + \frac{t^2}{2}\right)} dz \\
&= e^{\frac{t^2}{2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(z-t)^2}{2}} dz \\
&= e^{\frac{t^2}{2}} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{u^2}{2}} dz && \text{Set } u = z - t, \quad du = dz \\
&\text{Above is a probability density function for standard normal ditsirbution} \\
&= e^{\frac{t^2}{2}}
\end{aligned}$$

Normal Distribution

Definition 2.0.0.0.2
*We say X has the **Normal Distribution** with parameters*

$$\begin{aligned}
&M \to mean \\
&\sigma^2 \to variance
\end{aligned}$$

if the probability density function of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-(x-M)^2}{2\sigma^2}}, \qquad x \in (-\infty, \infty)$$

Now, let's verify that $f_X(x)$ is a probability density function.
i) $f_X(x) \geq 0 \qquad \forall x$
This is the same as the Standard Normal Distribution. So it is verified.

ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

$$\begin{aligned}
\int_{-\infty}^{\infty} f_X(x)dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-(x-M)^2}{2\sigma^2}} \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\left(\frac{-(x-M)}{\sigma}\right)^2 \frac{1}{2}} \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}} dz && Z = \frac{x-M}{\sigma} \implies dz = \frac{dx}{\sigma} \\
&= 1 && \text{Note it is the pdf of standard normal distribution}
\end{aligned}$$

Definition 2.0.0.0.3
Suppose $X \sim N(M, \sigma^2)$, then

$$Z = \frac{X-M}{\sigma} \longrightarrow \textit{Z-Score of } X$$

Note: The standard normal distribution has parameters $M = 0$ and $\sigma^2 = 1$, i.e $Z \sim (0, 1)$

Theorem 2.1
Suppose $X \sim N(M, \sigma^2)$ and $Z \sim N(0, 1)$. Then

1. $\left(\frac{X-M}{\sigma}\right) \sim N(0, 1)$
2. $X = \sigma Z + M$

Proof.

$$
\begin{aligned}
F_{\frac{X-M}{\sigma}}(u) &= P\left(\frac{X-M}{\sigma} \leq u\right) \\
&= P\left(X \leq \sigma u + M\right) \\
&= \int_{-\infty}^{\sigma u + u} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{\frac{-(x-M)^2}{2\sigma^2}} dx \\
&= \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{dz}} && Z = \frac{X-M}{\sigma} \qquad dz = \frac{dx}{\sigma} \\
&= F_Z(u)
\end{aligned}
$$

Therefore, $\left(\frac{X-M}{\sigma}\right)$ are identically distributed, meaning $\left(\frac{X-M}{\sigma}\right)$ has the standard normal distribution. □

Theorem 2.2
*Suppose $X \sim N(M, \sigma^2)$. Note that $X = \sigma Z + M$. Then using the Theorem **0.2***

$$
\begin{aligned}
E(X) &= \sigma E(Z) + M = M \\
V(X) &= \sigma^2 \cdot V(Z) = \sigma^2 \\
M_X(t) &= e^{Mt} \cdot M_Z(\sigma t) \\
&= e^{Mt} \cdot \left(e^{\frac{t^2}{2}} \Big|_{t=\sigma t} \right) \\
&= e^{Mt} \cdot e^{\frac{\sigma^2 t^2}{2}} \\
&= e^{Mt + \frac{\sigma^2 t^2}{2}}
\end{aligned}
$$

Standard Gamma Distribution

Definition 3.0.0.0.1
*The **Gamma Function** is defined as for $\alpha > 0$,*

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \cdot e^{-t} dt$$

Theorem 3.1
*Some **Properties of the Gamma Function***

1. $\Gamma(\alpha) \geq 0$
2. $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
3. $\forall n \in \mathbb{N}$, $\Gamma(n+1) = n!$
4. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Proof. 1) $\Gamma(\alpha) \geq 0$

$$
\begin{aligned}
t^{\alpha-1} &\geq 0,\ t \in (0,\infty) \\
e^{-t} &\geq 0,\ t \in (0,\infty) \\
t^{\alpha-1} \cdot e^{-t} &\geq 0,\ \forall t \in (0,\infty) \\
\text{Then } \int_0^\infty t^{\alpha-1} \cdot e^{-t} &\geq 0,\ \forall \alpha > 0
\end{aligned}
$$

2) $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

$$
\begin{aligned}
\Gamma(\alpha+1) &= \int_0^\infty t^{\alpha+1-1} \cdot e^{-t} dt \\
&= \int_0^\infty t^{\alpha} \cdot e^{-t} dt \\
&= uv - \int_0^\infty u dv && \text{Set } v = t^\alpha,\ dv = \alpha t^{\alpha-1} dt \\
& && \text{Set } du = e^{-t},\ u = -e^{-t} \\
&= -t^\alpha e^{-t}\Big|_0^\infty - \int_0^\infty -\alpha t^{\alpha-1} \cdot e^{-t} \\
&= \alpha \int_0^\infty t^{\alpha-1} \cdot e^{-t} \\
&= \alpha\Gamma(\alpha)
\end{aligned}
$$

3) $\forall n \in \mathbb{N}$, $\Gamma(n+1) = n!$

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty t^{1-1} \cdot e^{-t} dt \\
&= \int_0^\infty e^{-t} dt \\
&= \lim_{c\to\infty} \int_0^c e^{-t} dt \\
&= \lim_{c\to\infty} \left(\frac{e^{-t}}{-1} \Big|_0^c \right) \\
&= \lim_{c\to\infty} \left(1 - e^{-c}\right) \\
&= 1
\end{aligned}$$

$$\begin{aligned}
\Gamma(n+1) &= n\Gamma(n) \\
&= n\left((n-1)\Gamma(n-1)\right) \\
&\;\;\vdots \\
&= n(n-1)(n-2) \cdot 3 \times 2 \times 1 \times \Gamma(1) \\
&= n!
\end{aligned}$$

4) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$$\begin{aligned}
\Gamma(\frac{1}{2}) &= \int_0^\infty t^{\frac{1}{2}-1} e^{-t} dt \\
&= \int_0^\infty t^{-\frac{1}{2}} e^{-t} dt \\
&= \int_0^\infty u^{-1} e^{-u^2} 2udu && \text{Set } u = t^{\frac{1}{2}},\ du = \frac{1}{2} t^{-\frac{1}{2}} dt \\
& && dt = 2uu^{-1}du,\ e^{-t} = e^{-u^2} \\
&= 2\int_0^\infty e^{-u^2} du \\
&= \int_{-\infty}^\infty e^{-u^2} du && e^{-u^2} \text{ is an even function} \\
&= \sqrt{\pi} && \text{Gaussian Integral}
\end{aligned}$$

□

Definition 3.1.0.0.1
We say T has the ***Standard Gamma Distribution*** *with parameter*

$$\alpha \to \ \textit{shape}$$

if the probability density function of T is given by

$$f_T(t) = \begin{cases} \frac{1}{\Gamma(\alpha)} \cdot t^{\alpha-1} \cdot e^{-t} & t > 0 \\ 0 & t \le 0 \end{cases}$$

We need to verify that f_T is in fact a probability density function.

Proof. i) $f_T(t) \ge 0$

$$\begin{aligned}
\frac{1}{\Gamma(\alpha)} &\ge 0 \\
t^{\alpha-1} &\ge 0 \text{ if } t \in (0,\infty) \\
e^{-t} &\ge 0 \\
\text{Then,} \qquad f_T(t) = \frac{1}{\Gamma(\alpha)} &\cdot t^{\alpha-1} \cdot e^{-t} > 0
\end{aligned}$$

ii) $\int_{-\infty}^{\infty} f_T(t)dt = 1$

$$\begin{aligned}
\int_{-\infty}^\infty f_T(t)dt &= \int_0^\infty \frac{1}{\Gamma(\alpha)} \cdot t^{\alpha-1} \cdot e^{-t} dt \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^\infty t^{\alpha-1} \cdot e^{-t} dt \\
&= \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) && \text{By the definition of Gamma function} \\
&= 1
\end{aligned}$$

□

First, let's calculate **Expect Value** $E(T)$

$$
\begin{aligned}
E(T) &= \int_{-\infty}^{\infty} t f_T(t) dt \\
&= \int_0^{\infty} t \cdot \frac{1}{\Gamma(\alpha)} \cdot t^{\alpha-1} \cdot e^{-t} dt \\
&= \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \cdot t^{\alpha} \cdot e^{-t} dt \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} t^{\alpha} \cdot e^{-t} dt \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} t^{\alpha+1-1} \cdot e^{-t} dt \\
&= \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha+1) && \text{Definition of } \Gamma(\alpha+1) \\
&= \frac{1}{\Gamma(\alpha)} \cdot \alpha\Gamma(\alpha) \\
&= \alpha
\end{aligned}
$$

Then, let's calculate the **Variance**. To do so, we need to calculate the $E(T^2)$.

$$
\begin{aligned}
E(T^2) &= \int_0^{\infty} t^2 \cdot \frac{1}{\Gamma(\alpha)} \cdot t^{\alpha-1} \cdot e^{-t} dt \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} t^{\alpha+2-1} \cdot e^{-t} dt \\
&= \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha+2) \\
&= \frac{1}{\Gamma(\alpha)} \cdot (\alpha+1)\Gamma(\alpha+1) \\
&= \frac{1}{\Gamma(\alpha)} \cdot (\alpha+1)(\alpha)\Gamma(\alpha) \\
&= (\alpha+1)(\alpha)
\end{aligned}
$$

Then,

$$
\begin{aligned}
V(T) &= E(T^2) - (E(T))^2 \\
&= (\alpha+1)(\alpha) - \alpha^2 \\
&= \alpha^2 + \alpha - \alpha^2 \\
&= \alpha
\end{aligned}
$$

Finally, we can calculate the **Moment Generating Function** $M_T(t)$

$$
\begin{aligned}
M_T(t) &= E(e^{tT}) \\
&= \int_0^{\infty} e^{ts} \cdot \frac{1}{\Gamma(\alpha)} \cdot s^{\alpha-1} \cdot e^{-s} ds \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} e^{ts} \cdot s^{\alpha-1} \cdot e^{-s} ds \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} e^{-s(1-t)} \cdot s^{\alpha-1} ds \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} \left(\frac{u}{1-t}\right)^{\alpha-1} \cdot e^{-u} \cdot \frac{du}{1-t} && \text{Set } u = s(1-t)\ du = (1-t)ds \\
& && ds = \frac{du}{(1-t)}\ \ s = \frac{u}{1-t}\ \ -s(1-t) = -u \\
&= \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} \frac{u^{\alpha-1}}{(1-t)^{\alpha-1}} \cdot e^{-u} \cdot \frac{du}{1-t} \\
&= \frac{1}{(1-t)^{\alpha}} \cdot \frac{1}{\Gamma(\alpha)} \cdot \int_0^{\infty} u^{\alpha-1} \cdot e^{-u} \cdot du \\
&= \frac{1}{(1-t)^{\alpha}} \cdot \frac{1}{\Gamma(\alpha)} \cdot \Gamma(\alpha) && \text{Above is } \Gamma(\alpha) \\
&= \frac{1}{(1-t)^{\alpha}}
\end{aligned}
$$

Suppose $T \sim Gamma(\alpha)$, $\alpha > 0$. This is a Gamma distribution with shape= α. Then for $\beta > 0$, we define

$$X = \beta T$$

X is T scaled by a factor of β. We want to calculate the probability density function of X using the probability density function of T.

First, we calculate the cumulative density function of X, i.e F_x

$$\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(\beta T \leq x) \\
&= P(T \leq \frac{x}{\beta}) \\
&= \int_0^{\frac{x}{\beta}} f_T(t)dt \\
&= F_T\left(\frac{x}{\beta}\right) \\
&= \int_0^{\frac{x}{\beta}} \frac{1}{\Gamma(\alpha)} \cdot t^{\alpha-1} \cdot e^{-t} dt
\end{aligned}$$

Then, we differentiate F_X to get f_X.

$$\begin{aligned}
f_X(x) &= \frac{d}{dx}(F_X(x)) \\
&= \frac{d}{dx}\left(\int_0^{\frac{x}{\beta}} \frac{1}{\Gamma(\alpha)} \cdot t^{\alpha-1} \cdot e^{-t} dt\right) \\
&= \frac{d}{dx}\left(\frac{x}{\beta}\right) \cdot g(b(x)) - \frac{d}{dx}(0) \cdot g(0) \\
&= \frac{1}{\beta} \cdot \frac{1}{\Gamma(\alpha)} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)} \\
&= \frac{1}{\Gamma(\alpha)\beta} \cdot \left(\frac{x}{\beta}\right)^{\alpha-1} \cdot e^{-\left(\frac{x}{\beta}\right)}
\end{aligned}$$

Gamma Distribution

Definition 3.1.0.0.2

*We say X has the **Gamma Distribution** with parameters*

$$\begin{aligned}
\alpha > 0 &\to Shape \\
\beta > 0 &\to Scale
\end{aligned}$$

if X has the probability density function defined as

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\cdot\beta^\alpha} \cdot x^{\alpha-1} \cdot e^{\frac{-x}{\beta}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Now, Let's Check that $f_X(x)$ is a probability density function.

Proof. i) $f(x) \geq 0$
This is same as the Standard Gamma Distribution.

ii) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

$$\begin{aligned}
\int_{-\infty}^{\infty} f_X(x)dx &= \int_0^{\infty} \frac{1}{\Gamma(\alpha)\cdot\beta^\alpha} \cdot x^{\alpha-1} \cdot e^{\frac{-x}{\beta}} dx \\
&= \frac{1}{\Gamma(\alpha)\cdot\beta^\alpha} \int_0^{\infty} (\beta u)^{\alpha-1} \cdot e^{-u} \beta du && \text{Set } u = \frac{x}{\beta} \quad du = \frac{dx}{\beta} \\
& && x = \beta u \quad dx = \beta du \\
&= \frac{1}{\Gamma(\alpha)\cdot\beta^\alpha} \int_0^{\infty} (\beta)^\alpha (u)^{\alpha-1} \cdot e^{-u} du \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{\infty} (u)^{\alpha-1} \cdot e^{-u} du && \text{This is } \Gamma(\alpha) \\
&= \frac{1}{\Gamma(\alpha)}\Gamma(\alpha) = 1
\end{aligned}$$

□

First, we calculate the **Expected Value** $E(X)$. Note that $X = \beta T$, then

$$\begin{aligned} E(X) &= E(\beta T) \\ &= \beta \cdot E(T) = \beta \cdot \alpha \end{aligned}$$

Then, we calculate the **Variance** $V(X)$.

$$\begin{aligned} V(X) &= V(\beta T) \\ &= \beta^2 \cdot V(T) \\ &= \beta^2 \cdot \alpha \end{aligned}$$

Lastly, we calculate the **Moment Generating Function** $M_X(t)$

$$\begin{aligned} M_X(t) &= M_{\beta T}(t) \\ &= M_T(\beta t) \\ &= \left(\frac{1}{1-\beta t}\right)^{\alpha} \end{aligned}$$

The Gamma distribution $X \sim Gamma(\alpha, \beta)$ has two important **special cases**.

When $\alpha = 1$, it is a **Exponential Distribution** with $\lambda = \frac{1}{\beta}$.
When $\beta = 2$, it is a **Chi-Squared Distribution** with $\nu = 2\alpha$

Exponential Distribution

Definition 4.0.0.0.1
*We say X has the **Exponential Distribution** with parameters*

$$\alpha \rightarrow \textit{Rate Parameter}$$

if X has the probability density function defined as

$$f_X(t) = \begin{cases} \lambda \cdot e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}$$

Note that this is same as Gamma Distribution with $\alpha = 1, \beta = \frac{1}{\lambda}$
Then, the **Expected Value** $E(X)$

$$\begin{aligned} E(X) &= \alpha \cdot \beta \\ &= 1 \cdot \frac{1}{\lambda} \\ &= \frac{1}{\lambda} \end{aligned}$$

The **Variance** $V(X)$ is

$$V(X) = \alpha \cdot \beta^2 = \frac{1}{\lambda^2}$$

The **Moment Generating Function** $M_X(t)$ is

$$\begin{aligned} M_X(t) &= \left(\frac{1}{1-t\beta}\right)^{\alpha} \\ &= \frac{1}{1-t\cdot\frac{1}{\lambda}} \\ &= \frac{\lambda}{\lambda - t} \end{aligned}$$

Chi-Squared Distribution

Definition 5.0.0.0.1
*We say X has the **Chi-Squared Distribution** with parameters*

$$\nu \rightarrow \textit{Rate Parameter}$$

if X has the probability density function defined as

$$f_X(x) = \begin{cases} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)\cdot 2^{\frac{\nu}{2}}} \cdot x^{\frac{\nu}{2}-1} \cdot e^{\frac{-x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Note that this is same as Gamma Distribution with $\alpha = \frac{\nu}{2}, \beta = 2$
Then, the **Expected Value** $E(X)$

$$\begin{aligned} E(X) &= \alpha \cdot \beta \\ &= 2 \cdot \frac{\nu}{2} \\ &= \nu \end{aligned}$$

The **Variance** $V(X)$ is

$$\begin{aligned} V(X) &= \alpha \cdot \beta^2 \\ &= 2^2 \cdot \frac{\nu}{2} \\ &= 2\nu \end{aligned}$$

The **Moment Generating Function** $M_X(t)$ is

$$\begin{aligned} M_X(t) &= \left(\frac{1}{1 - t\beta}\right)^{\alpha} \\ &= \left(\frac{1}{1 - 2t}\right)^{\frac{\nu}{2}} \end{aligned}$$

Beta Distribution

The support (values of the random variable) for the normal distribution and gamma distribution is infinite. We want a distribution/family of distribution whose support/values is a finite interval - Beta Family.

Definition 6.0.0.0.1
Given $\alpha, \beta > 0$, the beta function $B(\alpha, \beta)$ is defined as:

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} \cdot (1-x)^{\beta-1} dx$$

The Beta function is related to the gamma function as follows:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Definition 6.0.0.0.2
*We say that X has **Beta Distribution** with parameters*

$$\alpha, \beta > 0$$

if X has probability density function given by

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} & x \in [0, 1] \\ 0 & \textit{otherwise} \end{cases}$$

Using relationship with the Gamma function

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} & x \in [0, 1] \\ 0 & \textit{otherwise} \end{cases}$$

Note: The support/values of Beta(α, β) is $[0, 1]$. This has applications when dealing with modeling of probabilities.

Now, we can calculate **Expected Value** $E(X)$

$$\begin{aligned}
E(X) &= \int_0^1 x f_X(x) dx \\
&= \int_0^1 x \cdot \frac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} dx \\
&= \frac{1}{B(\alpha,\beta)} \cdot \int_0^1 x^{\alpha+1-1} \cdot (1-x)^{\beta-1} dx \\
&= \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+1)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \\
&= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)\cdot\Gamma(\alpha)} \\
&= \frac{\alpha\cdot\Gamma(\alpha)\Gamma(\alpha+\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)\cdot\Gamma(\alpha)} \\
&= \frac{\alpha}{(\alpha+\beta)}
\end{aligned}$$

To calculate **Variance** $V(X)$, we first calculate $E(X^2)$. More Generally, we can calculate $E(X^n)$

$$\begin{aligned}
E(X^n) &= \int_0^1 x^n f_X(x) dx \\
&= \int_0^1 x^n \cdot \frac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} dx \\
&= \frac{1}{B(\alpha,\beta)} \cdot \int_0^1 x^n \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} dx \\
&= \frac{1}{B(\alpha,\beta)} \cdot B(\alpha+n,\beta) \\
&= \frac{\Gamma(\alpha+n)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta+n)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \\
&= \frac{\Gamma(\alpha+n)\cdot\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\alpha+\beta+n)} \\
&= \frac{B(\alpha+n,\beta)}{B(\alpha,\beta)}
\end{aligned}$$

Then,

$$\begin{aligned}
E(X^2) &= \frac{\Gamma(\alpha+2)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\alpha+\beta+2)} \\
&= \frac{(\alpha+1)\cdot\alpha\cdot\Gamma(\alpha)\cdot\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} \\
&= \frac{(\alpha+1)\cdot\alpha}{(\alpha+\beta+1)(\alpha+\beta)}
\end{aligned}$$

Finally, we got the **Variance**

$$\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
&= \frac{(\alpha+1)\cdot\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{(\alpha+1)\cdot\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{1}{(\alpha+\beta)}\left(\frac{(\alpha+1)\cdot\alpha}{(\alpha+\beta+1)} - \frac{\alpha^2}{\alpha+\beta}\right) \\
&= \frac{1}{(\alpha+\beta)}\left(\frac{(\alpha^2+\alpha)\cdot(\alpha+\beta) - (\alpha^2)(\alpha+\beta+1)}{(\alpha+\beta+1)(\alpha+\beta)}\right) \\
&= \frac{1}{(\alpha+\beta)^2(\alpha+\beta+1)}\left((\alpha^2+\alpha)\cdot(\alpha+\beta) - (\alpha^2)(\alpha+\beta+1)\right) \\
&= \frac{1}{(\alpha+\beta)^2(\alpha+\beta+1)}\left(\alpha^2\cdot(\alpha+\beta) + \alpha\cdot(\alpha+\beta) - (\alpha^2)(\alpha+\beta) - \alpha^2\right) \\
&= \frac{1}{(\alpha+\beta)^2(\alpha+\beta+1)}\left(\alpha^2+\alpha\beta) - \alpha^2\right) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

Depending on values of X, β the probability density function has different shapes

1. If $\alpha > 1, \beta = 1$, it is strictly increasing.
2. If $\alpha = 1, \beta > 1$, it is strictly decreasing
3. If $\alpha < 1, \beta < 1$, it is U-shaped.
4. If $\alpha = \beta$ it is symmetric about $\frac{1}{2}$, with $M_X = \frac{1}{2}$ and $\sigma_x^2 = \frac{1}{4(2\alpha+1)}$
5. If $\alpha = \beta = 1$, it will be a uniform distribution on $(0, 1)$

Cauchy Distribution

Definition 7.0.0.0.1
*We say X has the **Cauchy Distribution** with parameters*

$$\theta \to \textit{Mean}$$

X has the probability density function

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1 + (x-\theta)^2}, \; x \in (-\infty, \infty)$$

Then, we need to prove if $f_X(x)$ is actually a probability density function.

Proof. 1) $f_X(x) \geq 0$

$$\begin{aligned}(x-\theta)^2 &\geq 0 \\ 1 + (x-\theta)^2 &> 0 \\ \frac{1}{1+(x-\theta)^2} &> 0 \\ \frac{1}{\pi} &> 0 \\ \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2} &> 0\end{aligned}$$

2) $\int_{-\infty}^{\infty} f_X(x) = 0$
Let $g(x) = \frac{1}{1+x^2}$. Recall from calculus

$$\begin{aligned}\int_{-\infty}^{\infty} g(x)dx &= \int_{-\infty}^{0} \frac{1}{1+x^2} + \int_{0}^{\infty} \frac{1}{1+x^2}dx \\ &= \lim_{d \to -\infty} \int_{d}^{0} \frac{1}{1+x^2} + \lim_{c \to \infty} \int_{0}^{c} \frac{1}{1+x^2}dx \\ &= \lim_{d \to -\infty} \left(\tan^{-1}(x)\big|_d^0\right) + \lim_{c \to \infty} \left(\tan^{-1}(x)\big|_0^c\right) \\ &= \lim_{d \to -\infty} \left(\tan^{-1}(0) - \tan^{-1}(d)\right) + \lim_{c \to \infty} \left(\tan^{-1}(c) - \tan^{-1}(0)\right) \\ &= \lim_{d \to -\infty} \left(-\tan^{-1}(d)\right) + \lim_{c \to \infty} \left(\tan^{-1}(c)\right) \\ &= \frac{\pi}{2} + \frac{\pi}{2} \\ &= \pi\end{aligned}$$

Then,

$$\begin{aligned}\int_{-\infty}^{\infty} f_X(x) &= \frac{1}{\pi} \int_{-\infty}^{\infty} g(x)dx \\ &= \frac{1}{\pi}\pi \\ &= 1\end{aligned}$$

□

The **mean, Variance, Moment Generating Function** does not exist for the Cauchy distribution.

6 Joint Distribution and Sum of Random Variable

Arithmetic with Random Variable

Given random variables **X, Y** defined on a common sample space. For example, suppose we toss a coin four times, $P(H) = P$.

$$\begin{aligned} X &:= \# \text{ of H in four tosses} \\ Y &:= \# \text{ of Y in four tosses} \\ W &:= \text{ are there more heads then tails? 0 if no 1 if yes.} \\ S &= \{HHHH, HHHT, \cdots, TTTT\} \\ |S| &= 2^4 = 16 \end{aligned}$$

Then we can consider the following:
1. Scalar Multiplication: Given fixed constant $c \in \mathbb{R}$,

$$\begin{aligned} Z &:= cX \\ Z(\omega) &= c \cdot X(\omega) \end{aligned}$$

For example, if $c = 5$.

$$\begin{aligned} Z &= \{0, 5, 10, 15, 20\} \\ Z(HHTT) &= 5 \cdot X(HHTT) = 5 \times 2 = 10 \\ Z(HTTT) &= 5 \cdot X(HTTT) = 5 \times 1 = 5 \\ P(Z \geq 7) &= P(5X \geq 7) = P(X \geq \frac{7}{5}) = 1 - P(x < \frac{7}{5}) \end{aligned}$$

If $Z = y^3$, then

$$\begin{aligned} Z &= \{0, 1, 8.27, 64\} \\ P(Z \leq 10) = P(Y^3) &= P(Y \leq \sqrt[3]{10}) \end{aligned}$$

2. Addition of Two Random Variables:

$$\begin{aligned} Z =& X + Y \\ \omega \mapsto & X(\omega) + Y(\omega) \end{aligned}$$

In the context of the example, let

$$\begin{aligned} Z &= X + Y \\ R &= X + W \end{aligned}$$

Then,

$$\begin{aligned} Z(HHTT) &= X(HHTT) + Y(HHTT) \\ &= 2 + 2 = 4 \\ Z &: \omega \mapsto 4 \text{ The total of tails and heads is always 4} \end{aligned}$$

$$\begin{aligned} R(HHTT) &= X(HHTT) + W(HHTT) \\ &= 1 + 0 = 1 \end{aligned}$$

3. Linear Combination of Random Variables: Given $a, b \in \mathbb{R}$

$$\begin{aligned} Z &= aX + by \\ Z(\omega) &= aX(\omega) + bY(\omega) \ \forall \omega \in S \end{aligned}$$

Joint Distributions Introduction

Often times we want probabilities associated to value coming from two or more random variables, which is the **joint** behavior of more than one random variables.

Definition 2.0.0.0.1
Given two random variable $\boldsymbol{X}$ *and* $\boldsymbol{Y}$*. The* ***joint sample space*** *of X and Y is*

$$X \times Y = \{(x, y); x \in X, y \in Y\}$$

Now, we want to calculate the probabilities of joint events, i.e want $P(x, y) := P(X = x \text{ and } Y = y)$. This would depends on

1. Distribution of X and Y
   i.e We need information about the individual behavior of X and Y
2. The intervention of X with Y
   i.e We need information about how the outcomes of X affect the outcomes of Y.

Definition 2.0.0.0.2
If X and Y are both discrete, the ***joint probability mass function*** *is defined as:*

$$p(x, y) := P(X = x \text{ and } Y = y)$$

$p(x, y) = 0$ *if* $x \notin X$ *or* $y \notin Y$

Definition 2.0.0.0.3
Given the joint probability mass function, we can define ***Marginal for X****:*

$$\text{For fixed } x, \quad p_X(x) = \sum_{y \in Y} p(x, y)$$

Marginal for Y*:*

$$\text{For fixed } y, \quad p_Y(y) = \sum_{x \in X} p(x, y)$$

Conditional for X

$$\text{For fixed } y,$$
$$\begin{aligned} p_{X|Y=y}(x|y) &:= P(Y = y|X = x) \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

Conditional for Y

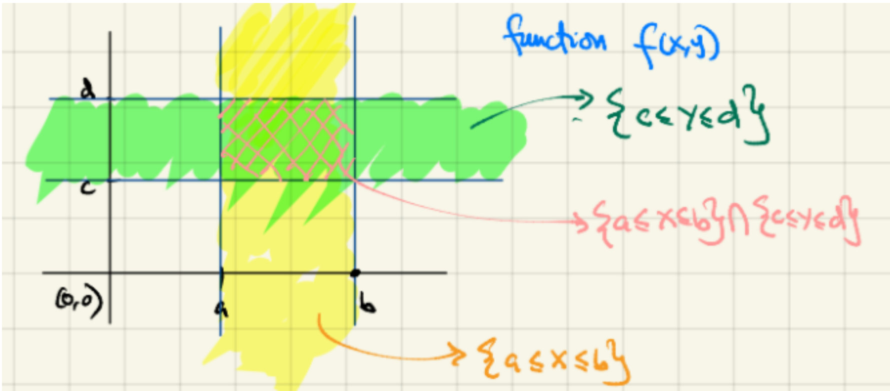
$$\text{For fixed } x,$$
$$\begin{aligned} p_{Y|x=x}(y|x) &:= P(X = x|Y = y) \\ &= \frac{p(x, y)}{p_X(x)} \end{aligned}$$

Using the ***Multiplication Principle,*** *we have*

$$\begin{aligned} p(x, y) &= p_{x|y}(x|y) \cdot p_Y(y) \quad \forall x, y \in X \times Y \\ &= p_{y|x}(y|x) \cdot p_X(x) \quad \forall x, y \in X \times Y \end{aligned}$$

Definition 2.0.0.0.4
Then if X and Y are both continuous, the ***joint probability density function*** *is a function* $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *satisfying the following:*

$$P(a \leq X \leq b, c \leq y \leq d) = \int_a^b \int_c^d f(x, y)dydx$$



Definition 2.0.0.0.5
*Given a joint probability density function $f(x, y)$ for X and Y. Then **Marginal for X**:*

$$\text{For fixed } x,$$
$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

***Marginal for Y**:*

$$\text{For fixed } y,$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Conditional for X

$$\text{For fixed } y,$$
$$f_{x|y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Conditional for Y

$$\text{For fixed } x,$$
$$f_{y|x}(y|x) = \frac{f(x, y)}{f_X(x)}$$

*Using the **Multiplication Principle,** we have*

$$\begin{aligned} f(x, y) &= f_{x|y}(x|y) \cdot f_Y(y) \quad \forall x, y \in X \times Y \\ &= f_{y|x}(y|x) \cdot f_X(x) \quad \forall x, y \in X \times Y \end{aligned}$$

Definition 2.0.0.0.6
We say two random variables X and Y with joint probability mass function $p(x, y)$ or joint probability density function $f(x, y)$ are independent if

$$p(x, y) = p_X(x) \cdot p_Y(y) \ \forall(x, y) \in X \times Y$$
$$f(x, y) = f_X(x) \cdot f_Y(y) \ \forall(x, y) \in X \times Y$$

Example: Let experiment is as follows:
Step 1: Toss a coin with $P(H) = 0.8$ (more generally) $p \in (0, 1)$
Step 2: If coin lands H, roll a fair six-sided die. If coin lands T, roll an unfair die with distribution

1	2	3	4	5	6
0.1	0.1	0.1	0.1	0.1	0.5

Let random variables X, Y defines as below

$$X = \begin{cases} 0 & \text{If coin lands T} \\ 1 & \text{If coin lands H} \end{cases}$$
$$Y = \text{ outcome of the die roll}$$

Then, we know that

$$X = \{0, 1\}$$
$$Y = \{1, 2, 3, 4, 5, 6\}$$
$$X \times Y = \begin{Bmatrix} (0, 1), (0, 2), (0, 3), (0, 4), (0, 5), (0, 6) \\ (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6) \end{Bmatrix}$$

The **Joint Distribution tables:**

	1	2	3	4	5	6
0	$p(0,1)$ 0.2×0.1	$p(0,2)$ 0.2×0.1	$p(0,3)$ 0.2×0.1	$p(0,4)$ 0.2×0.1	$p(0,5)$ 0.2×0.1	$p(0,6)$ 0.2×0.5
1	$p(1,1)$ $0.8 \times \frac{1}{6}$	$p(1,2)$ $0.8 \times \frac{1}{6}$	$p(1,3)$ $0.8 \times \frac{1}{6}$	$p(1,4)$ $0.8 \times \frac{1}{6}$	$p(1,5)$ $0.8 \times \frac{1}{6}$	$p(1,6)$ $0.8 \times \frac{1}{6}$

Now, we can calculate the Marginals and the conditionals.
When $x = 0$,

$$\begin{aligned} p_X(0) = \sum_{y=1}^{6} p(0,y) &= 0.2 \times (0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.5) \\ &= 0.2 \end{aligned}$$

When $x = 1$,

$$\begin{aligned} p_X(1) = \sum_{y=1}^{6} p(1,y) &= 0.8 \times (\frac{1}{6} \times 6) \\ &= 0.8(1) = 0.8 \end{aligned}$$

Then, the marginal distribution of X is

x	0	1
$p_X(x)$	0.2	0.8

Similarly, we can get Marginal Distribution for Y

y	1	2	3	4	5	6
$p_Y(y)$	$0.2 \times 0.1 + 0.9 \times \frac{1}{6} =$ 0.1533	0.1533	0.1533	0.1533	0.1533	$0.2 \times 0.5 + 0.8 \times \frac{1}{6} =$ 0.233

To calculate the conditionals for Y.
When $x = 0$

$$\begin{aligned} p_{y|x=0}(y|x=0) &= \frac{p(0,y)}{p_X(0)} \\ &= \frac{0.2 \times 0.1}{0.2} \\ &= 0.1 \end{aligned}$$

Then conditional distribution table for $Y|X = 0$ is

y	1	2	3	4	5	6
$p_{Y\|X=0}$	0.1	0.1	0.1	0.1	0.1	0.5

Notice that this is the distribution table of the unfair die.

Similarly, we can calculate when $x = 1$.

$$\begin{aligned} p_{y|x}(y|x=1) &= \frac{p(1,y)}{p_X(1)} \\ &= \frac{0.8 \times \frac{1}{6}}{0.8} \\ &= \frac{1}{6} \end{aligned}$$

Then, the conditional distribution table for $Y|X = 1$ is

y	1	2	3	4	5	6
$p_{Y\|X=1}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Notice this is the distribution table for the fair die.
Intuition: IF we know that $\{X = 1\}$ has happened, then we rolled the fair die.

Then we can calculate the condition of X.
When $y = 1$

$$\begin{aligned} p_{x|y}(x|y=1) &= \frac{p(x,1)}{p_Y(1)} \\ \frac{p(0,1)}{p(0,1)+p(1,1)} &= \frac{0.2\times 0.1}{0.2\times 0.1 + 0.8\times\frac{1}{6}} \\ \frac{p(1,1)}{p(0,1)+p(1,1)} &= \frac{0.8\times\frac{1}{6}}{0.2\times 0.1 + 0.8\times\frac{1}{6}} \end{aligned}$$

So $Y = 1, 2\cdots 5$, the conditional of X is

x	0	1
$p_{x\|y=1,2,3,4,5}$	$\frac{0.2\times0.1}{0.2\times0.1+0.8\times\frac{1}{6}}$	$\frac{0.8\times\frac{1}{6}}{0.2\times0.1+0.8\times\frac{1}{6}}$

When $Y = 6$, the conditional of Y is

x	0	1
$p_{x\|y=1}$	$\frac{0.2\times0.5}{0.2\times0.5+0.8\times\frac{1}{6}}$	$\frac{0.8\times\frac{1}{6}}{0.2\times0.5+0.8\times\frac{1}{6}}$

Observe that all of these calculation are really Bayes Theorem at play.

Then, we need to see of X and Y are independent.
Intuitively, the outcome of the die roll depends on the outcome of the coin toss. To show that, let's take $(x_0, y_0) = (1, 6) \in X \times Y$

$$\begin{aligned} p(1,6) &= 0.8\times\frac{1}{6} \\ p_X(1) &= 0.8 \\ p_Y(6) &= 0.8\times o.5 + 0.8\times\frac{1}{6} \\ p_X(1)\cdot p_Y(6) &= 0.8(0.2\times 0.5 + 0.8\times\frac{1}{6}) = 0.18667 \neq 0.1333 \end{aligned}$$

Therefore, X and Y are not independent.

Distribution of the Sum of Two Independent Random Variables

Theorem 3.1
Uniqueness of Moment Generating Functions
If X, Y are two random variables with cumulative density functions $F_X(u), F_Y(u)$ respectively and moment generating functions $M_X(t)$ and $M_Y(t)$ respectively.
The,

$$\begin{aligned} M_X(t) = M_Y(t)\ \forall t\in(-\delta,\delta) &\implies F_X(u) = F_Y(u)\ \forall u\in\mathbb{R} \\ &\implies X, Y \textit{ are identically distributed} \end{aligned}$$

Theorem 3.2
Suppose X, Y are two independent random variables. (i.e $p(x,y) = p_X(x)\cdot p_Y(y)\ \forall x,y\in\mathbb{R}\times\mathbb{R}$).
Then if

$$Z = aX + bY$$

then,

$$\begin{aligned} M_Z(t) &= M_{aX+bY}(t) \\ &= M_X(at)\cdot M_Y(bt) \end{aligned}$$

where $M_X(t), M_Y(t)$ are moment generating functions of X, Y respectively.
In particular: if $a = b = 1$, if X, Y are independent,

$$M_{X+Y}(t) = M_X(t)\cdot M_Y(t)$$

Proof. Suppose X, Y are independent

$$\begin{aligned} M_{aX+bY}(t) = E(e^{(aX+bY)t}) :&= \sum_{(x,y)} e^{(aX+bY)t} \cdot p_X(x) \cdot p_Y(y) \\ &= \sum_{(x,y)} \left(e^{axt} \cdot p_X(x)\right) \left(e^{bxt} \cdot p_Y(y)\right) \\ &= \left(\sum_{x \in X} e^{axt} \cdot p_X(x)\right) \left(\sum_{y \in Y} e^{bxt} \cdot p_Y(y)\right) \\ &= M_X(t) \cdot M_Y(t) \end{aligned}$$

□

Theorem 3.3
Suppose X, Y are two binomial distribution.

$$\begin{aligned} X &\sim \ Binom(n,p) \\ Y &\sim \ Binom(m,p) \end{aligned}$$

Then if X, Y are independent

$$X + Y \sim \ Binom(n+m,p)$$

Proof. Suppose X, Y are independent. To calculate the moment generating function of $X+Y$:

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= \left(p \cdot e^t + (1-p)\right)^n \cdot \left(p \cdot e^t + (1-p)\right)^m \\ &= \left(p \cdot e^t + (1-p)\right)^{n+m} \end{aligned}$$

This is moment generating function of Binom$(n+m,p)$ □

Theorem 3.4
Suppose X, Y are two binomial distribution.

$$\begin{aligned} X &\sim \ Gamma(\alpha_1,\beta) \\ Y &\sim \ Gamma(\alpha_2,\beta) \end{aligned}$$

If X, Y are independent

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) + M_Y(t) \\ &= \left(\frac{1}{1-\beta t}\right)^{\alpha_1} \cdot \left(\frac{1}{1-\beta t}\right)^{\alpha_2} \\ &= \left(\frac{1}{1-\beta t}\right)^{\alpha_1+\alpha_2} \end{aligned}$$

This is moment generating function of $Gamma(\alpha_1 + \alpha_2, \beta)$

Similarly, can check that the sum of two independent normal distributions is also normal.
Note: Sum of two independent Poisson distributions is also a Poisson distribution.

Theorem 3.5
Suppose

$$\begin{aligned} X &\sim Binom(n,p) \\ Y &\sim Binom(m,p) \end{aligned}$$

X, Y are independent. We say

$$Z = X + Y \sim \ Binom(n+m,p)$$

$X|Z \sim$*Hypergeo*$(N = n+m, M = n,$ *sample size* $= Z)$

Proof.

$$\begin{aligned}
p_{X|Z}(x|z) &= \frac{P(X = x \text{ and } Z = z)}{P(Z = z)} \\
&= \frac{P(X = x \text{ and } X + Y = z)}{P(Z = z)} \\
&= \frac{P(X = x, Y = z - x)}{P(Z = z)} \\
&= \frac{\binom{n}{x} \cdot p^x (1-p)^{n-x} \cdot \binom{m}{z-x} \cdot p^{z-x} \cdot (1-p)^{m-(z-x)}}{\binom{n+m}{z} \cdot p^z \cdot (1-p)^{n+m-z}} \\
&= \frac{\binom{n}{x} \cdot \binom{m}{z-x} \cdot p^{z-x+x} \cdot (1-p)^{m-z+x+n-x}}{\binom{n+m}{z} \cdot p^z \cdot (1-p)^{n+m-z}} \\
&= \frac{\binom{n}{x} \cdot \binom{m}{z-x} \cdot p^{z} \cdot (1-p)^{m+n-z}}{\binom{n+m}{z} \cdot p^z \cdot (1-p)^{n+m-z}} \\
&= \frac{\binom{n}{x} \cdot \binom{m}{z-x}}{\binom{n+m}{z}}
\end{aligned}$$

This is probability mass function of the Hypergeometric distribution with parameters

$$\begin{aligned}
\text{Population Size} &: n + m \\
\text{\# Successes} &: n \\
\text{Sample Size} &: z
\end{aligned}$$

□

7 | Properties of Expectation

Covariance

Given two random variables X, Y with joint probability mass function $p(x, y)$ or probability density function $f(x, y)$, and $h : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ is a function.
Let $Z = h(x, y)$ is random variable.

Then,

$$E(Z) := \begin{cases} \sum_{(x,y)\in X\times Y} h(x,y)\cdot p(x,y) & \text{If both } X, Y \text{ are discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y) & \text{If both } X, Y \text{ are countinous} \end{cases}$$

If $h(x, y) = z$

$$\begin{aligned} E(h(x,y)) &= \sum_{x,y} h(x,y)p(x,y) \\ &= \sum_{x,y} x \cdot p(x,y) \\ &= \sum_{x}\left(\sum_{y} x \cdot p(x,y)\right) \\ &= \sum_{x} x \cdot \sum_{y} p(x,y) \\ &= \sum_{x} x \cdot p_X(x) && \text{The Expected Value of X with respect to the marginal} \\ &= E(X) \end{aligned}$$

If $h(x, y) = x^2$, then $E(h(x, y)) = E(x^2)$

Definition 1.0.0.0.1
The covariance of X, Y is the expected value of product of the deviations of X, Y from their expected value. To calculate so, let

$$h(x, y) = (x - M_X)(y - M_Y)$$

Then, the covariance is

$$\begin{aligned} Cov(X, Y) &:= E(h(x, y)) \\ &= E((X - M_X)(Y - M_Y)) \end{aligned}$$

Note that covariance is the multiplication product of deviation of x from M_X and deviation of y from M_Y.

In term of applications, the expect value is the measure of location for the values of a random variable. i.e Under reasonable assumption, we would expect most of the values under repeated sampling of a random variable to be close to the expected value of the random variable.
We say a value of X is large if it is much larger than $E(X)$. If it is much smaller than $E(X)$ we say it is small.
To know how the values are distributed about $E(X)$, we need the variance of X.

Definition 1.0.0.0.2
If large values of X are related to large values of Y. We could have pairs (x, y) such that

$$\begin{gathered} x >> M_X \text{ and } y >> M_y \\ (x - M_X) >> 0 \text{ and } (y - M_Y) >> 0 \\ (x - M_X)(y - M_Y) \geq 0 \end{gathered}$$

This makes a positive contribution to $Cov(X, Y)$

Definition 1.0.0.0.3
If small values of X are related to small values of Y. We could have pairs (x, y) such that

$$x << M_X \text{ and } y << M_y$$
$$(x - M_X) << 0 \text{ and } (y - M_Y) << 0$$
$$(x - M_X)(y - M_Y) \geq 0$$

This (x, y) will make a positive contribution to $Cov(X, Y)$

Definition 1.0.0.0.4
We will say X and Y are ***positively related*** *if large values of X are related to large values of Y and small values of X are related to small values of Y.*
i.e For most pairs (x, y) we will have

$$(x - M_X)(y - M_Y) >> 0$$

Then, $Cov(X, Y)$ will likely to be positive.

Definition 1.0.0.0.5
We will say X and Y are ***negatively related*** *if large values of X are related to small values of Y and small values of X are related to large values of Y.*
i.e For most pairs (x, y) we will have

$$(x - M_X)(y - M_Y) << 0$$

Then, $Cov(X, Y)$ will likely to be negative.

Definition 1.0.0.0.6
We say the relationship between X and Y is neither positive nor negative if ***i)*** *large values of X are related to both large values of Y and small values of Y*
ii) *and small values of X are related to both small values of Y and large values of Y*

i.e For most pairs (x, y) we have

$$x >> M_X \text{ is related to } \begin{cases} y >> M_Y \\ or \\ y << M_Y \end{cases}$$

and

$$x << M_X \text{ is related to } \begin{cases} y >> M_Y \\ or \\ y << M_Y \end{cases}$$

Then $Cov(X, Y) \approx 0$

Theorem 1.1
When calculating $Cov(X, Y)$, use the formula

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

Proof.

$$\begin{aligned} Cov(X, Y) &= E((X - M_X)(Y - M_Y)) \\ &= E(XY - XM_Y - YM_X + M_X \cdot M_Y) \\ &= E(XY) - E(XM_Y) - E(Y \cdot M_X) + E(M_X M_y) \\ &= E(XY) - M_Y \cdot E(X) - M_X E(y) + M_X M_Y E(1) \\ &= E(XY) - M_X M_Y - M_X M_Y + M_X M_Y \\ &= E(XY) - E(X) \cdot E(Y) \end{aligned}$$

□

Definition 1.1.0.0.1
Correlation coefficient *measure the extent of the relationship between X, Y. Defined as*

$$Corr(X, Y) = \rho_{x,y} := \frac{Cov(X, Y)}{\sqrt{V(X)} \cdot \sqrt{V(Y)}}$$

Theorem 1.2
The correlation coefficient satisifies the following:

1. $-1 \leq Corr(X, Y) \leq 1$
2. $\rho_{x,y} = \pm 1 \iff \exists a, b$ *such that* $Y = aX + b$
3. $\rho_{x,y} = \pm 1 \implies$ *perfect linear relationship between X and Y*
   $\rho_{x,y} = 0 \implies$ *no linear relationship between X and Y (but can be other relationship*

Theorem 1.3
Suppose $Z = aX + bY$

1. $E(Z) = E(aX + bY) = aE(X) + bE(Y)$
2. $V(Z) = a^2V(X) + b^2V(Y) + 2abCov(X, Y)$

Proof.

$$\begin{aligned} V(Z) &:= E((Z - M_Z)^2) \\ &= E((ax + by - aM_X - bM_Y)^2) \\ &= E((ax - aM_X) + (bY - bM_Y)^2)) \\ &= E(a^2(X - M_X)^2 + 2ab(X - M_X)(Y - M_Y) + b^2(Y - M_Y)^2) \\ &= a^2E((X - M_X)^2) + 2abE(X - M_X)(Y - M_Y) + b^2E((Y - M_Y)^2) \\ &= a^2V(X) + b^2V(Y) + 2abCov(X, Y) \end{aligned}$$

□

Theorem 1.4
If $X_2, X_2, \cdots, X_n$ *is a collection of random variable and* $a_1, a_2, \cdots a_n \in \mathbb{R}$ *i.e*

$$Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

then

$$V(Z) = \sum i = 1^n a_i^2 V(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j Cov(X_i, X_j)$$

Hierarchical Models

Definition 2.0.0.0.1
We X has a ***mixture distribution*** *if the distribution of X depends on a quality that is also a distribution.*

Theorem 2.1
Let $N \sim Pois(\lambda)$, $Y|N \sim Binom(N, P)$. *Then,* $Y \sim Pois(\lambda p)$

Example: An insect lays a bunch of eggs. We want to know how many eggs survive.
Assume
1. P(Survival of a single egg) $= p$
2. Survival of different eggs is independent of each other

In this setting we are interested in the number of survivals given that there were certain numbers of egg laid.
Let n be the number of eggs that the insect has laid is n then # survival among n eggs $\sim$ Binom(n, p). Then, we can setup the Hierarchy as follow:
  Step 1: Sample n from a distribution.
  Step 2: Use the n from step 1, in Binom(n, p)
i.e
  Step 1: $N \sim Pois(\lambda)$
  Step 2: $Y|N \sim Binom(N, p)$, where Y = # surviving eggs

We want to know the distribution of Y. i.e

$$\begin{aligned}
P(Y=y) &= \sum_{n=0}^{\infty} P(Y=y, N=n) \\
&= \sum_{n=0}^{\infty} P(Y=y|N=n) \cdot P(N=n) \\
&= \sum_{n=0}^{\infty} \binom{n}{y} \cdot p^y \cdot (1-p)^{n-y} \cdot \frac{e^{-\lambda}\lambda^n}{n!} \\
&= p^y \cdot e^{-\lambda} \sum_{n=0}^{\infty} \binom{n}{y} \cdot (1-p)^{n-y} \cdot \frac{\lambda^n}{n!} \\
&= p^y \cdot e^{-\lambda} \sum_{n=0}^{\infty} \frac{n!}{(n-y)!y!} \cdot (1-p)^{n-y} \cdot \frac{\lambda^n}{n!} \\
&= p^y \cdot e^{-\lambda} \sum_{n=0}^{\infty} \frac{1}{(n-y)!y!} \cdot (1-p)^{n-y} \cdot \lambda^n \\
&= \frac{p^y \cdot e^{-\lambda}}{y!} \sum_{n=y}^{\infty} \frac{1}{(n-y)!} \cdot (1-p)^{n-y} \cdot \lambda^n \text{ Since } n < y \ (n-y)! = 0
\end{aligned}$$

Then, set $n - y = m$. Then $n = m + y$, $n = y \implies m = 0$, and $n = \infty \implies m = \infty$

$$\begin{aligned}
&= \frac{p^y \cdot e^{-\lambda}}{y!} \sum_{m=0}^{\infty} \frac{1}{m!} \cdot (1-p)^m \cdot \lambda^{m+y} \\
&= \frac{(p\lambda)^y \cdot e^{-\lambda}}{y!} \sum_{m=0}^{\infty} \frac{1}{m!} \cdot (1-p)^m \cdot \lambda^m \\
&= \frac{(p\lambda)^y \cdot e^{-\lambda}}{y!} \sum_{m=0}^{\infty} \frac{((1-p)\lambda)^m}{m!} \text{Power series expansion of } e^{(1-p)\lambda} \\
&= \frac{(p\lambda)^y \cdot e^{-\lambda}}{y!} \cdot e^{\lambda} \cdot e^{-\lambda p} \\
&= \frac{e^{-\lambda p} \cdot (\lambda p)^y}{y!}
\end{aligned}$$

This is the probability mass function of $Pois(\lambda p)$

Conditional Expectation and Variance

Theorem 3.1
If X, Y are two random variables, then we have the following equality

$$E(X|Y) = E(E(X|Y))$$

Where $E(X|Y)$ is a function of Y only.

Example: Recall that given two independent random variables, X and Y where

$$\begin{aligned}
X &\sim Binom(n,p) \\
Y &\sim Binom(m,p) \\
\text{Let } Z &= X + Y
\end{aligned}$$

Then, $X|Z \sim HyperGeo(N = m+n, M = n, \text{ samples zie } = Z)$. Therefore

$$\begin{aligned}
E(X|Z) &= (\text{ Sample Size } \times \frac{M}{N}) = Z \cdot \frac{n}{m+n} \\
E(E(X|Z)) &= E\left(Z \cdot \frac{n}{m+n}\right) = \frac{n}{n+m} \cdot E(Z) \\
&= \frac{n}{n+m} \cdot (n+m) \cdot p \\
&= np = E(X)
\end{aligned}$$

Theorem 3.2
Conditional Variance
If X, Y are two random variable, then

$$V(X) = E(V(X|Y)) + V(E(X|Y))$$

Example: Let $X \sim Binom(n,p), Y \sim Binom(m,p)$, where X, Y are independent. Let $Z = X + Y$. Then $X|Z \sim HyperGeo(N = m+n, M = n, \text{ samples zie } = Z)$. Then

$$\begin{aligned} E(X|Z) &= \frac{n}{n+m} \cdot Z \\ V(X|Z) &= \frac{n+m-Z}{n+m-1} \cdot Z \cdot \frac{n}{n+m} \cdot \frac{m}{n+m} \\ V(X) &= n \cdot p(1-p) \end{aligned}$$

Then ,we want to verify

$$\begin{aligned} E(V(X|Z)) &= E\left(\frac{n+m-Z}{n+m-1} \cdot Z \cdot \frac{n}{n+m} \cdot \frac{m}{n+m}\right) \\ &= \frac{n}{n+m} \cdot \frac{m}{n+m} \cdot E\left(\frac{n+m-Z}{n+m-1} \cdot Z\right) \\ &= \frac{n}{n+m} \cdot \frac{m}{n+m} \cdot E\left(\frac{nZ+mZ-Z^2}{n+m-1}\right) \\ &= \frac{n}{n+m} \cdot \frac{m}{n+m} \cdot \frac{1}{n+m-1} E\left(nZ+mZ-Z^2\right) \\ &= \frac{n}{n+m} \cdot \frac{m}{n+m} \cdot \frac{1}{n+m-1} (n(n+m) \cdot p + m(n+m) \cdot p - n(n-1)p^2 - np) \end{aligned}$$

$$\begin{aligned} V(E(X|Z)) &= V\left(\frac{n}{n+m} \cdot Z\right) \\ &= \left(\frac{n}{n+m}\right)^2 \cdot V(Z) \\ &= \left(\frac{n}{n+m}\right)^2 \cdot (n+m)p(1-p) \\ &= \frac{n^2}{n+m} \cdot p \cdot (1-p) \end{aligned}$$

8 Limit Theorem

Inequalities for Parameters

We want to arrive at reasonable estimates on certain parameters/probabilities without knowing all the needed information.

Example: Given two random variables X, Y. Say we want to know $E(XY)$, then we need the joint probability mass density/probability density function, so that we can calculate

$$E(XY) = \sum_{x,y} xy \cdot p(x,y)$$

To be able to calculate $p(x,y)$, we need the marginals $p_X(x)$, $p_Y(y)$ for all x, y and conditionals $p_{X|Y}(X|Y)$, $p_{Y|X}(y|x)$. So that we can use the multiplication principle.

1. Holders Inequality

Theorem 1.1
Given two random variables X, Y and $p, q > 1$ such that

$$\frac{1}{p} + \frac{1}{q} = 1$$

then

$$|E(XY)| \leq E(|X|^p)^{\frac{1}{p}} \cdot E(|Y|^q)^{\frac{1}{q}}$$

Note:

1. Calculation of $E(XY)$ needs both the conditionals and marginals for X and Y
2. Calculation of $E(|X|^p)$ needs marginal of X. Calculation of $E(|Y|^q)$ needs marginal of X.

Application of Holder: Cauchy-Schwartz Inequality take $p = q = 2$, so that $\frac{1}{p} + \frac{1}{q} = \frac{1}{2} + \frac{1}{2} = 1$
In this setting

$$|E(XY)| \leq E(|X|^2)^{\frac{1}{2}} \cdot E(|Y|^2)^{\frac{1}{2}}$$

Application of Cauchy-Schwartz: Given X, Y with expected value M_X, M_Y respectively. Then,

$$R = (X - M_X)$$
$$S = (Y - M_Y)$$

Now apply **Cauchy Schwarz** to R, S to get

$$|E(R \cdot S)| \leq E(|R|^2)^{\frac{1}{2}} \cdot E(|S|^2)^{\frac{1}{2}}$$
$$|E((X - M_X)(Y - M_X))| \leq E((X - M_X)) \cdot E((Y - M_Y)^2)^{\frac{1}{2}}$$
$$|Cpv(X,Y)| \leq \sqrt{V(X)} \cdot \sqrt{V(X)}$$

This is the **Covariance Inequality**
Note: If $V(X), V(Y) > 0$ (ie $\neq 0$). We get

$$\frac{|Cov(X,Y)|}{\sqrt{V(X)} \cdot \sqrt{V(Y)}} \leq 1$$

Using covariance identity we have shown that

$$|Corr(X,Y)| \leq 1$$

Example: Suppose X, Y have the following table

	1	2	3	4	5	6
0	$p(0,1)$ 0.2×0.1	$p(0,2)$ 0.2×0.1	$p(0,3)$ 0.2×0.1	$p(0,4)$ 0.2×0.1	$p(0,5)$ 0.2×0.1	$p(0,6)$ 0.2×0.5
1	$p(1,1)$ $0.8 \times \frac{1}{6}$	$p(1,2)$ $0.8 \times \frac{1}{6}$	$p(1,3)$ $0.8 \times \frac{1}{6}$	$p(1,4)$ $0.8 \times \frac{1}{6}$	$p(1,5)$ $0.8 \times \frac{1}{6}$	$p(1,6)$ $0.8 \times \frac{1}{6}$

Marginal distribution of X is

x	0	1
$p_X(x)$	0.2	0.8

Marginal Distribution for Y

y	1	2	3	4	5	6
$p_Y(y)$	$0.2 \times 0.1 + 0.9 \times \frac{1}{6} =$ 0.1533	0.1533	0.1533	0.1533	0.1533	$0.2 \times 0.5 + 0.8 \times \frac{1}{6} =$ 0.233

Then, we calculate

$$\begin{aligned}
E(XY) &= \sum_{x,y} xy \cdot p(x,y) \\
&= \sum_{y=1}^{6} 1 \cdot y \cdot p(1,y) \ x = 0 \text{ does not contribute} \\
&= 1 \cdot 0.8 \cdot \frac{1}{6} + 2 \cdot 0.8 \cdot \frac{1}{6} + \cdots + 6 \cdot 8 \cdot \frac{1}{6} \\
&= \frac{0.8}{6} \cdot (1 + 2 + 3 + \cdot + 6) \\
&= \frac{0.8}{6} \cdot \frac{6 \cdot 7}{2} \\
&= 0.4 \cdot 7 = 2.8
\end{aligned}$$

Now: Suppose $p = q = 2$, we than calculate $E(X^2)$ and $E(Y^2)$

$$\begin{aligned}
E(X^2) &= \sum_x x^2 \cdot p_X(x) \\
&= 1 \times 0.8 = 0.8
\end{aligned}$$

$$\begin{aligned}
E(Y^2) &= \sum_y y^2 \cdot p_Y(y) \\
&= 1^2 \left(0.2 \times 0.1 + 0.8 \times \frac{1}{6}\right) + \cdots \\
&+ 6^2 \left(0.2 \times 0.5 + 0.8 \times \frac{1}{6}\right) \\
&= 16.833
\end{aligned}$$

$$2.8 \leq \sqrt{0.8} \cdot \sqrt{16.833} = 3.66965$$

So Holder's identity is true!

Example: Suppose

$$\begin{aligned}
X &\sim \mathit{Binom}(n = 10, p = 0.2) \\
Y &\sim \mathit{Gamma}(\alpha = 2, \beta = 3)
\end{aligned}$$

Now, we want to estimate $|E(XY)|$.

Let's take $p = 1 = 2$. Then by Holder's Inequality.

$$E(XY) \leq (E(X^2))^{\frac{1}{2}} \cdot (E(Y^2))^{\frac{1}{2}}$$

Then

$$\begin{aligned}
E(X^2) = V(X) + E(X)^2 & \\
&= n \cdot p(1-p) + (np)2 \\
&= 10 \times 0.2 \times 0.8 + (10 \times 0.2)^2 \\
&= 1.6 + 4 = 5.6
\end{aligned}$$

$$\begin{aligned} E(Y^2) &= V(Y) + E(Y)^2 \\ &= \alpha \cdot \beta^2 + (\alpha\beta)^2 \\ &= 2 \times 9 + (2 \times 3)^2 \\ &= 18 + 36 = 54 \end{aligned}$$

Then,

$$\begin{aligned} |E(XY)| &\leq \sqrt{5.6} \times \sqrt{54} \\ |Cov(X,Y)| &\leq \sqrt{V(X)} \cdot \sqrt{V(Y)} \\ &\leq \sqrt{1.6} \cdot \sqrt{18} \\ &\leq 5.3665 \end{aligned}$$

2. Minkowski's Inequality

Theorem 1.2
Suppose X, Y are random variable and $p \in [1, \infty]$, then

$$E(|X+Y|^p)^{\frac{1}{p}} \leq E(|X|^p)^{\frac{1}{p}} + E(|Y|^p)^{\frac{1}{p}}$$

Exercise: Suppose

$$\begin{aligned} X &\sim Binom(n = 10, p = 0.2) \\ Y &\sim Gamma(\alpha = 2, \beta = 3) \end{aligned}$$

We want a bound for

$$\left(E(|X+Y|^3)\right)^{\frac{1}{3}} \leq E(X^3)^{\frac{1}{3}} + E(Y^3)^{\frac{1}{2}}$$

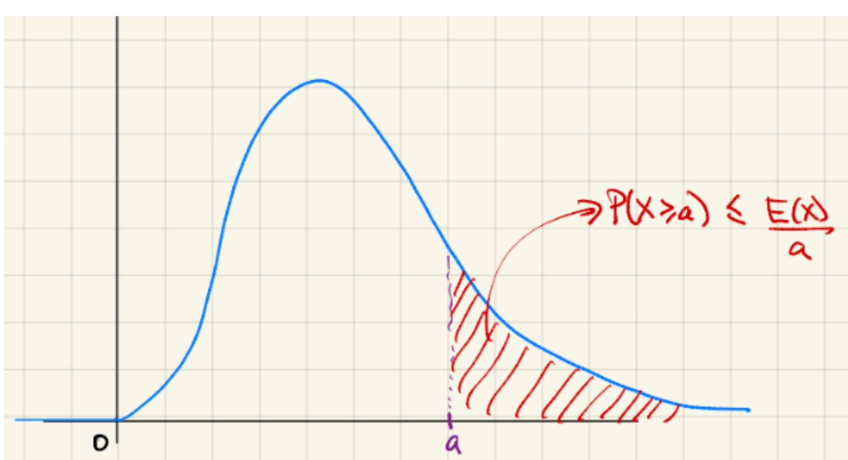
We want to using moment generating functions of X and Y to get there.

Inequalities For Probabilities

1. Markou's Inequality

Theorem 2.1
If X is a positive random variable (i.e support is ≥ 0). For any $a > 0$, we have

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



Note: The only number need to estimate $P(X \geq a)$ using Markov's inequality is $E(X)$. As a result the bounds will be pretty coarse (sometime useless)

1. Chebychev's Inequality

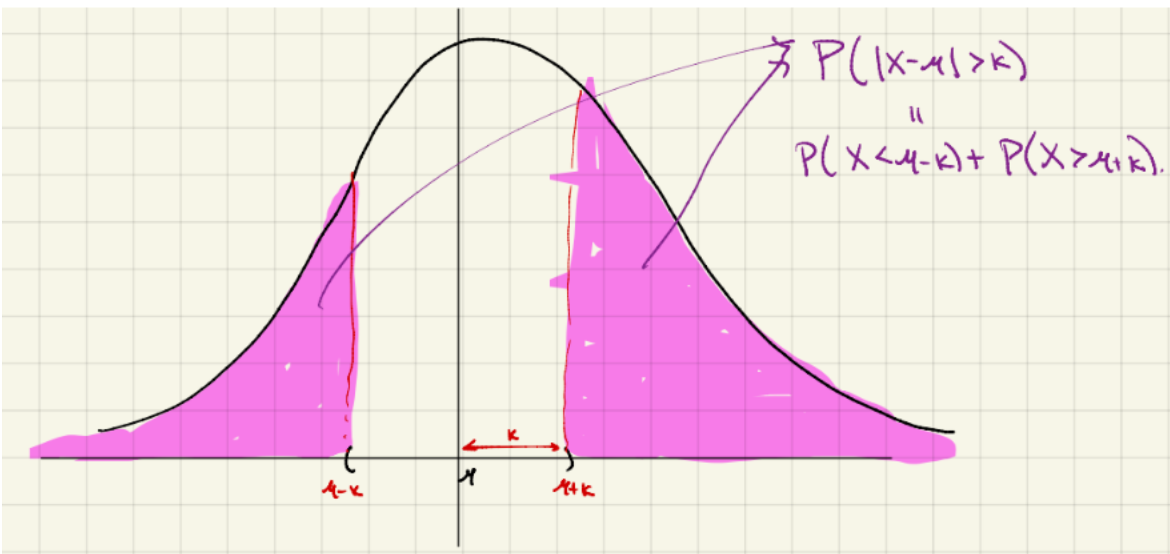
Theorem 2.2
If X is any random variable with $E(X) = M$ and $V(X) = \sigma^2 < \infty$, then for any $k \leq 0$,

$$P(|X - M| \geq k) \leq \frac{\sigma^2}{k^2}$$

Proof. Notice $(X - M_X)^2$ is a positive random variables.
Using Markous, we have

$$\begin{aligned} P(|X - M| \geq k) &= P((X - M)^2 \geq k^2) \\ &\leq \frac{E((X - M)^2)}{k^2} \\ &= \frac{V(X)}{k^2} \\ &= \frac{\sigma^2}{k^2} \end{aligned}$$

□

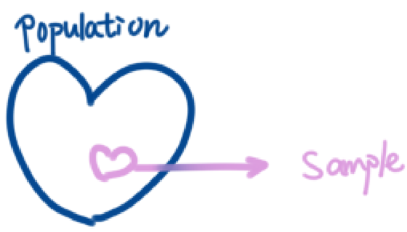


9 | Random Sample and Statistic

Given a population, we can uses a random variable to analyze the information about this characteristic remains fixed once the population of interest is identified.
However, calculating the number is usually not possible due to the following

1. The population might be to large or inaccessible.
2. Not all individuals in the population might be accessible
3. Computational/physical resources might not exist

Then, we can have a sample that is typically much smaller than the size of the population



And the information obtained using a sample is called a statistic, which has the following properties

1. The value of a statistic depends on much smaller set of individuals obtained from the population, which is easier to calculate
2. The value of the statistic will depend on the sample. Therefore we have to be careful about how we interpret this value

Definition 0.0.0.0.1
*A **random sample** of size n is a collection of n independent and identical distributed random variables.*
More precisely: We say $\{X_1, X_2, \cdots, X_n\}$ is a random sample of size n if

1. *$X_1, X_2, \cdots, X_n$ are identical to a fixed common distributed say X.*
   i.e $X_i \sim X$ $\forall i = 1, 2, \cdots, n$

2. *$X_1, X_2, X_3, \cdots, X_n$ are independent.*
   i.e If $P_X(x)$ is the probability mass of X (the common distribution), then

$$p\left(X_1, X_2, \cdots, X_n\right) = p_X(X_1) \cdot p_X(X_2) \cdots p_X(X_n)$$

Note: The common distribution for the $X_i's$, i.e X is called the population distribution.
Note: If x is the values of X, then the joint sample space for $X_1, X_2, \cdots, X_n$ is

$$X_1 \times X_2 \times \cdots \times X_n = x^n$$

Example: Suppose the population is the set

$$\{1, 2, 3, 4, 5\} \longrightarrow \text{population}$$

Choose a random sample of size 3.

$$\{N_1, N_2, N_3\}$$

N_i is sample a number from the population. If N_1 and N_2 have to be independent, we need to sample replacement. Then, the joint sample space for N_1, N_2, N_3 is

$$\begin{Bmatrix} (1,1,1), (1,2,1), (1,1,2) \cdots \\ \cdots (5,5,3), (5,5,4), (5,5,5) \end{Bmatrix}$$

which is all possible samples of size 3 from the population.
Note:

1. The random sample is a collection of random variables.
2. Sample data is one possible entry in the joint sample space of a random sample.

Definition 0.0.0.0.2
A statistic is a quantity that is calculated only using a random sample.
i.e: Given a random sample $\{X_1, X_2, \cdots, X_n\}$*, a statistic is a function of* $X_1, X_2, \cdots, X_n$
i.e: $\hat{\theta}$ *is a statistic, then* $\hat{\theta}$ *=function*$(X_1, \cdots, X_n)$

Note: The value of a statistic change everytime we sample the population.
i.e Statistic is s a random variable. This is what we interested in the distribution of this random variable. This is sampling distribution of the statistic.

Definition 0.0.0.0.3
Given a population $\{1, 2, 3, \cdots, n\}$*, and random sample* $\{N_1, N_2, \cdots, N_i\}$*. Then, they will have the statistics as follow:*
The ***sample mean*** *is*

$$\overline{X} = \frac{N_1 + N_2 + \cdots + N_i}{i}$$

The ***sample total*** *is*

$$T_o = N_1 + N_2 + \cdots + N_i$$

The ***sample max*** *is*

$$Max(N_1, N_2, \cdots, N_i)$$

The ***sample median*** *is*

$$\widetilde{X} = Median(N_1, N_2, \cdots, N_i)$$

Note: A statistic $\hat{\theta}$ is a function defined on the joint sample space X^n of $X_1, \cdots, X_n$.i.e

$$\hat{\theta} : X^n \longrightarrow \mathbb{R}$$

Note: Value of $\hat{\theta}$ depends on the sample distribution. It is a random variable with sample space X^n.

Given a statistic, calculate its sampling distribution.

Brute Force

Given a a statistic, we can explictily list out X^n and the values of the statistic, and then use the distribution of X and the independence of $X_1, X_2, \cdots, X_n$ to get the distribution of the statistic explicitly.

Advantages: The process is easy and explicit and we get the exact distribution
Disadvantages: Even for small values of n, and small population size, the time and space needed to list out X^n might be intractable.

Example: Given a population $\{1, 2, 3, 4, 5\}$ with population distribution

x	1	2	3	4	5
$p_X(x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

$\sim X$

(i) Let sample size = 1. The statistic sample mean is

$$\overline{X} = \frac{X_1}{1} = X_1$$

And the values of $\overline{X} = \overline{x} = \{1, 2, 3, 4, 5\}$. And then the sampling distribution of $\overline{X}$ is

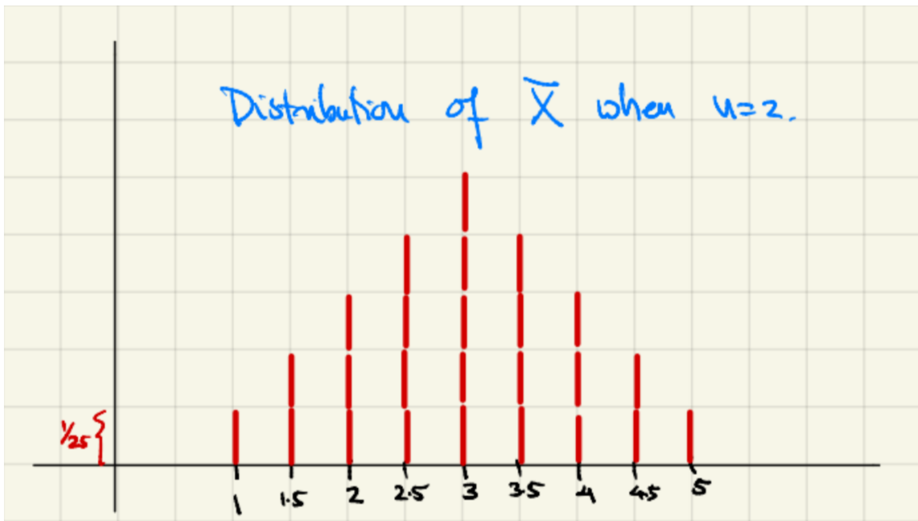
x	1	2	3	4	5
$p_X(x)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

(ii) Suppose $n = 2$, $\overline{X} = \frac{X_1+X_2}{2}$. The joint sample space is

$$\begin{Bmatrix} (1,1),(1,2),(1,3),(1,4),(1,5) \\ (2,1),(2,2),(2,3),(2,4),(2,5) \\ \vdots \\ (5,1),(5,2),(5,3),(5,4),(5,5) \end{Bmatrix}$$

The value of $\overline{X} = \overline{x} = \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6\}$. The distribution is

$\overline{x}$	1	1.5	2	2.5	3	3.5	4	4.5	5
$p_X(x)$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{25}$



Using Moment Generating Function

Theorem 2.1
Uniqueness of Moment Generating Functions
X and Y have identical moment generating functions if and only if X and Y have identical distribution.

Theorem 2.2
Moment Generating Functions of Linear Combination.
If $X_1, X_2, \cdots, X_n$ are mutually independent random variables with moment generating functions $M_{X_1}(t), \cdots, M_{X_n}(t)$ and

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

then

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t)$$

Proof.

$$\begin{aligned} M_Y(t) &= E(e^{Yt}) = E(e^{\sum a_i x_i t}) \\ &= \sum_{x \in X^n} e^{a_1x_1t} \cdot e^{a_2x_2t} \cdot \cdots \cdot e^{a_nx_nt} \cdot p_{x_1,\cdots,x_n} \\ &= \sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} e^{a_1x_1t} \cdot e^{a_2x_2t} \cdot \cdots \cdot e^{a_nx_nt} \cdot p_X(x_1)\cdots p_X(x_n) \\ &= \left(\sum_{x_1} e^{a_1x_1t} p_X(x_2)\right)\left(\sum_{x_2} e^{a_2x_2t} p_X(x_2)\right)\cdots\left(\sum_{x_n} e^{a_nx_nt} p_X(x_n)\right) \\ &= M_{X_1}(a_1t)\cdot \cdots \cdot M_{X_n}(a_2t) \end{aligned}$$

□

Corollary 2.2.1
If $Y = aX + b$ and $M_X(t)$ is the moment generating functions of X, then

$$M_Y(t) = e^{bt} \cdot M_X(at)$$

Corollary 2.2.2
If $X_1, X_2, \cdots, X_n$ are mutually independent random variables with moment generating functions $M_{X_1}(t), \cdots, M_{X_n}(t)$ and

$$Y_i = a_i X_i + b_i$$
$$\text{and } Y = \sum_{i=1}^{n} Y_i$$

then

$$M_Y(t) = e^{\sum_{i=1}^{n} b_i t} \cdot \prod_{i=1}^{n} M_{X_i}(a_i t)$$

Applications of theorem to calculate sampling distribution of statistics:
Suppose $\{X_1, X_2, \cdots, X_n\}$ is a random sample from population with distribution X. X has moment generating function $M_X(t)$, and $Y = a_1 X_1 + \cdots + a_n X_n$.
Then,

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(a_i t) \\ &= \prod_{i=1}^{n} M_X(a_i t) && \text{Since they are identically distributed to X} \end{aligned}$$

i) the sample total. If

$$T_o := \text{Sample\_Total} \qquad := X_1 + X_2 + \cdots + X_n$$

then

$$\begin{aligned} M_{T_o}(t) &= \prod_{i=1}^{n} M_X(t) \\ &= (M_X(t))^n \end{aligned}$$

ii) If

$$\begin{aligned} \overline{X} &:= \text{Sample\_Mean} \\ &:= \frac{X_1 + \cdots + X_n}{n} = \frac{T_o}{n} \end{aligned}$$

then

$$\begin{aligned} M_{\overline{X}}(t) &= \prod_{i=1}^{n} M_X\left(\frac{t}{n}\right) \\ &= M_{T_o}\left(\frac{t}{n}\right) \\ &= \left(M_X\left(\frac{t}{n}\right)\right)^n \end{aligned}$$

Example: X ~Bernoulli distribution with parameter p. And $\{X_1, \cdots, X_n\}$ is a random sample from X. Recall that

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

Then

$$\begin{aligned} M_{T_o}(t) &= (M_X(t))^n \\ &= \left(pe^t + (1-p)\right)^n \end{aligned}$$

This is the moment generating function of Binom(n,p). Then sample total for a random sample from Bernoulli(p) has the binomial distribution.

$$\begin{aligned} M_{\overline{X}}(t) &= \left(M_X\left(\frac{t}{n}\right)\right)^n \\ &= \left(pe^{\frac{t}{n}} + (1-p)\right)^n \end{aligned}$$

We cannot use the moment generating function of $\overline{X}$ to get its sampling distribution.

Example: X ~Binom(n,p), and sample size $= m$. Recall that

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

Then,

$$\begin{aligned} M_{T_o}(t) &= (M_X(t))^n = ((pe^t + (1-p)^n)^m \\ &= (pe^t + (1-p)^{nm} \end{aligned}$$

This is the moment generating function of Binom(nm, p)

Example: X ~N(μ,σ^2), $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$. Then

$$\begin{aligned} M_{t_o} &= (M_X(t))^n = \left(e^{\mu t + \frac{\sigma^2 t^2}{2}}\right)^n \\ &= e^{n\mu t + \frac{n\sigma^2 t^2}{2}} \\ &= e^{n\mu t + \frac{(\sqrt{n}\sigma)^2 t^2}{2}} \end{aligned}$$

This is moment generating functions of N(mean $= n\mu$, Variance $= n\sigma^2$).
Also:

$$\begin{aligned} M_{\overline{X}}(t) &= \left(M_X\left(\frac{t}{n}\right)\right)^n \\ &= \left(e^{\frac{\mu t}{n} + \frac{\sigma^2 (\frac{r}{n})^2}{2}}\right)^n \end{aligned}$$

This is the moment generating function of $N\left(\mu, \frac{\sigma^2}{n}\right)$

Example: $X \sim Gamma(\alpha, \beta)$, $M_X(t) = \left(\frac{1}{1-\beta t}\right)^\alpha$.
Then

$$\begin{aligned} M_{t_o} &= (M_X(t))^n = \left(\left(\frac{1}{1-\beta t}\right)^\alpha\right)^n \\ &= \left(\frac{1}{1-\beta t}\right)^{\alpha n} \end{aligned}$$

This is $Gamma(n\alpha, \beta)$.
Also:

$$M_{\overline{X}}(t) = \left(\frac{1}{1-\frac{\beta t}{n}}\right)^{\alpha n}$$

This is not clear about distribution.

This is limits the applicability to simple statistic that are linear combinations.

Order Statistic

Definition 3.0.0.0.1
Given a random sample $\{X_1, X_2, \cdot, X_n\}$. The ***ith order statistic is defined as:***

$$X_{(j)} = j^{th} \textit{ smallest number in the random sample}$$

Note: $X_{(j)}$ *is defined for* $j = 1, 2, \cdots, n$
$X_{(j)} = 1$*st smallest number*

We can use order statistic to calculate the sample median, sample range.

$$\text{Sample Median} = \begin{cases} X_{\frac{n+1}{2}} & n \in 2\mathbb{Z}+1 \\ \frac{X_{\frac{n}{2}} + X_{\frac{n+1}{2}}}{2} & n \in 2\mathbb{Z} \end{cases}$$

Theorem 3.1
Order Statistics for a Discrete Distribution:
Suppose X has values $x = \{x_1, x_2.x_3, \cdots, x_n\}$ arranged in ascending order and the distribution of X is

x	x_1	x_2	$\cdots$	x_n
$p_X(x)$	p_1	p_2	$\cdots$	p_n

where $\sum p_i = 1$. Nowe we define P_i as follows

$$\begin{aligned} p_1 &= p_1 \\ p_2 &= p_1 + p_2 \\ \cdots&\cdots \\ p_n &= p_1 + p_2 + \cdots + p_n \end{aligned}$$

Set $\{X_{(1)}, X_{(2)}, \cdots, X_{(n)}\}$ be the order statistic for a random sample of size n. Then, if $X_{(j)}$ is j^{th} order statistic.

1.
$$P(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} p_i^k \cdot (1 - p_i)^{n-k}$$

2.
$$\begin{aligned} P(X_{(j)} = x_i) &= P(X_{(j)} \leq x_i) - P(X_{(j)} \leq x_{i-1}) \\ &= \sum_{k=j}^{n} \left(\binom{n}{k} \left(p_i^k (1 - p_i)^{n-l} - p_{i-1}^k (1 - p_{i-1})^{n-k} \right) \right) \end{aligned}$$

Proof.
If we want to calculate $P(X_{(j)} \leq x_i)$, we can convert the problem into a problem with a familiar underlying distribution.

$$\text{The event of interest} \longrightarrow \{X_{(j)} \leq x_i\}$$

We can then define a new random variables

$$\begin{aligned} Y &= \#X_k \text{ in the sample that are less or equal to } x_i \\ \{Y \geq j\} &= \{j \text{ or more entries that are less than or equal to } x_i\} \end{aligned}$$

Note that $\{X_j \leq x_i\}$ is the event that jth smallest entry is less zthan or equal to x_i, which is same as the number of entries that less than or equal to x_i is greater than or equal to j, which is $\{Y \geq j\}$. Therefore,

$$P(X_{(j)} \leq x_i) = P(Y \geq j)$$

Now we want to calculate the distribution of Y. Let arbitary entries in X be x_k

$$P(Y = r) = P(\# \ x_k \text{ less or equal to } x_i \text{ is exactly equal to } r)$$

Suppose $\{X_k \leq x_i\}$ to be a success, then $P(X_k \leq x_i)$ is the same for all $k = 1, 2, ..., n$

$$\text{Y} \sim \text{Counting the number of successes in } n \text{ trials where } P(S) = P(x_k \leq x_i) = p.$$

Therefore

$$P(Y = r) = \binom{n}{r} p^r \cdot (1 - p)^{n-r}$$

So now we want to find the p.

$$\begin{aligned} P(S) &= P(X_k \leq x_i) \\ &= P(X_k = x_1) + P(X_k = x_2) + \cdots + P(X_k = x_i) \\ &= p_1 + p_2 + \cdots + p_i \\ &= P_i \end{aligned}$$

Then

$$P(Y = r) = \binom{n}{r} P_i^r \cdot (1 - P_i)^{n-r}$$

Finally,

$$\begin{aligned} P(X_j \le x_i) &= P(Y \ge k) \\ &= P(Y = j) + P(Y = j+1) + \cdots + P(Y = n) \\ &= \sum_{k=j}^{n} P(Y = k) \\ &= \sum_{k=j}^{n} \binom{n}{k} \cdot P_i^k \cdot (1 - P_i)^{n-k} \end{aligned}$$

Now

$$\begin{aligned} P(X_j = x_i) &= P(X_{(j)} \le x_i) - P(X_{(j)} \le x_{i-1}) \\ &= \sum_{k=j}^{n} \binom{n}{k} \left(P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \right) \end{aligned}$$

□

Example: X has distribution

x	2	5	8	10
$p_X(x)$	0.1	0.4	0.2	0.3

Say $n = 4$, the

$$\begin{aligned} &X_{(1)}((2, 2, 5, 8)) = 2 \\ &X_{(1)}((5, 2, 8, 5)) = X_{(1)}((2, 2, 5, 8)) = 2 \\ &X_{(3)}((2, 5, 2, 8)) = X_{(3)}((2, 2, 5, 8)) = 5 \end{aligned}$$

For small sample size n, it is possible to calculate the distribution of the order statistic explicitly. However, calculating the distribution of $X_{(j)}$ when sample size is large is intractable problem. But we c.

Example: Toss a coin 7 times. i.e $X \sim Bernoulli(p)$ with a random sample of size 7. We want to calculate the distribution of $X_{(7)}$, the sample max.
Note value of $X_{(7)} = \{0, 1\}$. We want to calculate

$$\begin{aligned} P(X_{(7)}) &= 0 \\ &= P(\text{7th smallest entry is 0}) \\ &= P(\text{All entries in the sample data are 0}) \\ &= (1 - p)^7 \end{aligned}$$

So the distribution table for $X_{(7)}$ is

x	0	1
$p_{X_{(7)}}(x)$	$(1-p)^7$	$1 - (1-p)^7$

Now, let's calculate the distribution of $X_{(4)}$. This also sample median.
First, let's calculate

$$\begin{aligned} P(X_{(4)} = 0) &= P(\text{At least 4 zeroes in 7 tosses}) \\ &= P(\text{At most 3 ones in 7 tosses}) \\ &= \binom{7}{4}(1-p)^4 \cdot p^3 + \binom{7}{5}(1-p)^5 \cdot p^2 + \binom{7}{6}(1-p)^6 \cdot p + \binom{7}{7}(1-p)^7 \end{aligned}$$

Setting $q = \binom{7}{4}(1-p)^4 \cdot p^3 + \binom{7}{5}(1-p)^5 \cdot p^2 + \binom{7}{6}(1-p)^6 \cdot p + \binom{7}{7}(1-p)^7$. The distribution of $X_{(4)}$ is

x	0	1
$p_{X_{(4)}}(x)$	q	$1 - q$