STAT 333

S18

Dina Dawoud

Learning Outcomes

- <u>Understand and apply</u>:
- Discrete-time Markov Chains
- Poisson Process
- Continuous-time Markov Chains

Some Applications

- Discrete Time Markov Chains:
- Marketers use Markov Chains to predict brand switching behavior within their customers. For example having consumers switch from “Android” to “iPhone”.

- A car rental agency might be interested in modelling the probability that a car picked up at a particular location is delivered back to that location, or a different location.

- Use a Markov chain to create a statistical model of a piece of music. Simulate the Markov chain to generate pieces of music.
- Google Rank Page algorithm
- Text prediction and speech recognition
- Genetic application

- <u>The Poisson Process:</u>
- The number of bankruptcies that are filed in a month.
- The number of arrivals at a car wash in one hour.
- The number of network failures per day.
- The number of hungry persons entering McDonald's restaurant.

- The number of birth, deaths, marriages, divorces, suicides, and homicides over a given period of time.
- The number of customers who call to complain about a service problem per month.
- The number of visitors to a Web site per minute.
- The number of calls to consumer hot line in a 5-minute period.
- The number of telephone calls per minute in a small business.

- How will your understanding and knowledge be tested:
- We will make use of

- A review quiz and tutorial tests
- The in-class tutorial assignments
- Take home assignments
- The final exam

- These assessments are *spaced out* throughout the semester to help you manage your time and gradually build up your knowledge by staying on track with the course.
- Further information can be found in the Course Outline on Learn.

Strategies to stay on track and help you succeed

1. Download and READ the course outline
2. Make yourself aware of the important dates in the course.

Planned schedule (subject to change)

Week	**Monday**	**Wednesday**	**Friday (Tutorial)**
1 (May 1-4)	N/A – classes start Tues	Chapter 1	Optional tutorial (SSO presentation)
2 (May 7-11)	Chapter 1	Chapter 2	**Review Quiz**
3 (May 14-18)	Chapter 2	Chapter 3	No tutorial
4 (May 21-25)	No class Monday, class on Tuesday instead: Chapter 3	Chapter 3	**Tutorial activity 1**
5 (May 28-June 1)	Chapter 3	Chapter 4 **Assignment 1 due 11:59 pm**	Review for Test 1
6 (June 4-8)	Chapter 4	Chapter 4	**Test 1**
7 (June 11-15)	Chapter 4	Chapter 4	No tutorial
8 (June 18-22)	Chapter 4	Chapter 4	**Tutorial activity 2**
9 (June 25-29)	Chapter 4	Chapter 5 **Assignment 2 due 11:59 pm**	Review for Test 2
10 (July 3-6)	No class Monday	Chapter 5	**Test 2**

MONTH May

YEAR 2018

SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY

Strategies to stay on track and help you succeed

3. Download and read through skeleton slides before coming to class.

4. Come prepared to class with skeleton slides so that you can annotate and add additional notes as we cover the material in class.

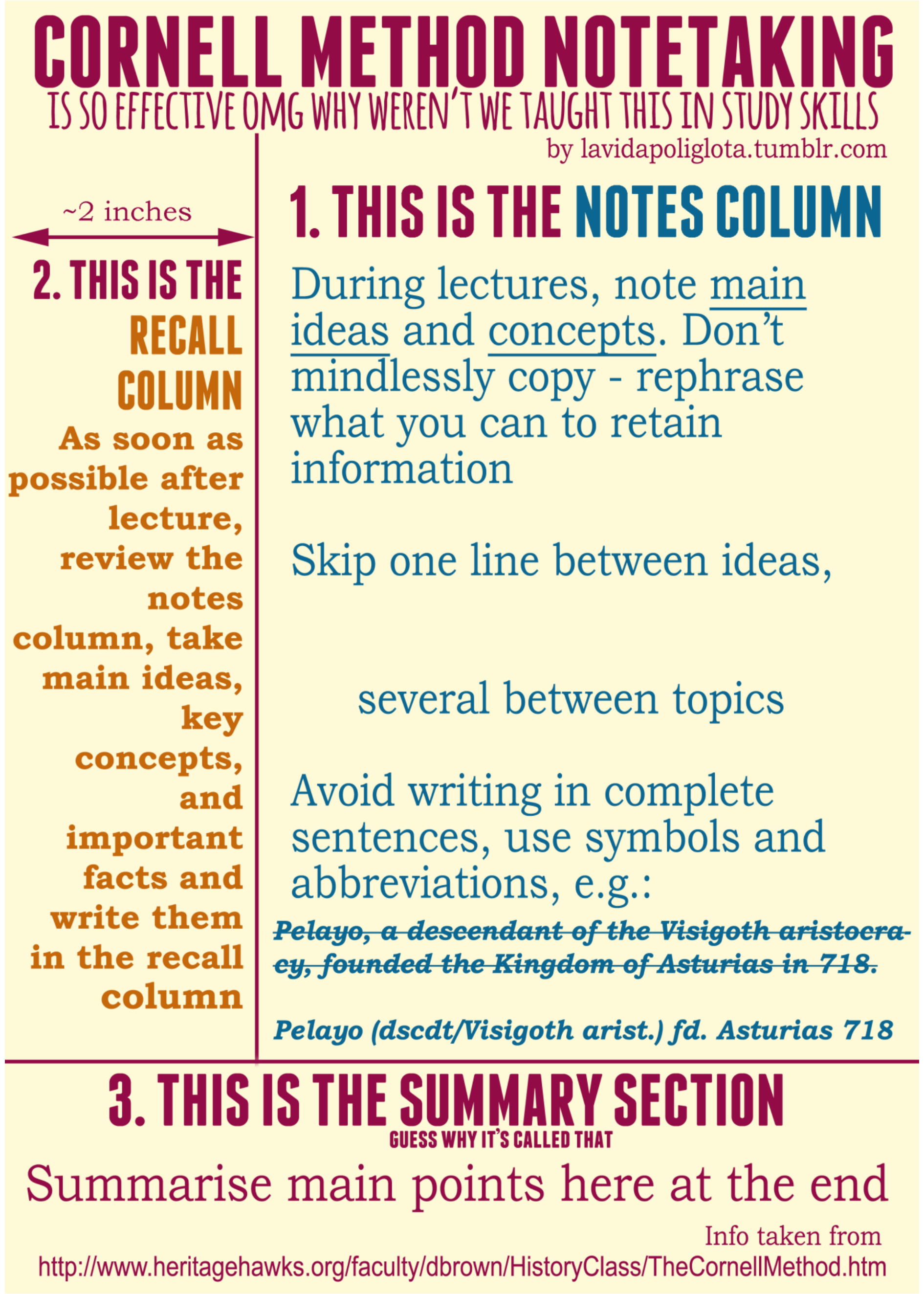
5. Revise the material after each lecture, or perhaps revise the lectures at the end of each week.

Strategies to stay on track and help you succeed

6. Once we reach the end of a chapter create summary sheets to capture the main ideas from each chapter.

7. Once we reach the end of a chapter practice the end of chapter questions. Do not just look through solutions, attempt the question yourself first. Learn from possible mistakes made.

8. Go to instructor and/or TA office hours immediately when struggling with a concept.



When taking notes it has been recommended to:

- Rephrase what you hear, don't just write down what I say word for word.
- Use bullet points which can later be expanded on.
- Visualise concepts, i.e capturing them using pictures.

Chapter 1: Review of Elementary Probability

- Chapter Outline

- Sample Space and Events

- Probability of Events

- Conditional Probabilities

- **<u>Sample Space and Events (1.2):</u>**
- A Sample Space, S, is a list of all possible outcomes in an experiment
- Recall, with each repetition of the experiment one and only one outcome must occur, i.e they are mutually exclusive.

- *Discrete Sample Spaces*:
- Finite # of outcomes, S={1,2,…,6} – in our dice example
- Countably Infinite # of outcomes, S={1,2,…}

- *Continuous Sample Spaces*:
- Uncountably infinite # of outcomes, S={t; t>0}

- An event, A, is a subset of the Sample Space ($A \subseteq S$):
- A can be a <u>Simple event </u>(consisting of just one outcome) or a <u>compound event </u>(consisting of several outcomes).

- For example:
- Going back to the dice example, S={1,2,…,6}
- A={1}

- A={roll an even #}={2,4,6}

- We also introduced more complex types of probabilities that we may be interested in:
- Having events A and B we may be interested in their:
- Union, $A \cup B$ (A only or B only or Both)

- Intersection, $A \cap B \equiv AB$ (Both)

- Complement, $\bar{A} \equiv A^c$

- **<u>Probability of Events (1.3):</u>**
- For each event A of a sample space S, P(A) is defined as the "probability of event A", satisfying

i) $0 \leq P(A) \leq 1$

ii) $P(S) = 1, \mathrm{P}(\emptyset) = 0$ where ∅ is the null event.

iii) For $n \in \mathbb{Z}^+, \mathrm{P}(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$ if the sequence $\{A_i\}_{i=1}^{n}$ is mutually exclusive (i.e $A_i \cap A_j = A_i A_j = \emptyset \quad \forall \ i \neq j$)

- From (ii) and (iii) we note that A^c is the complement of A, it follows that

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c) \Longrightarrow P(A^c) = 1 - P(A)$$

iv) If $A \subseteq B$ then $P(A) \leq P(B)$

v) Two events A and B are independent if and only if

$$P(AB) = P(A)P(B)$$

This idea can be extended to n events.

- Recall some of these proofs:

Example

- Suppose we roll a fair six-sided die once. Let A denote the event of rolling a number less than 4 and let B denote the event of rolling an odd number.

a) Define the Sample space of this experiment
b) Define the outcomes in event A and B
c) What is P(A), P(B), $P(A \cap B), P(A \cup B)$
d) Are events A and B independent? Mutually exclusive?

- **<u>Conditional Probability (1.4):</u>**

- The "conditional probability of event A given event B occurs" is defined as

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

Provided that $P(B) > 0$

- Notes:
- When $B = S, P(A|S) = P(A)$
- **Multiplication Rule**: $P(A \cap B) = P(A|B)P(B)$
- **Law of total probability**: For $n \in \mathbb{Z}^+$, suppose that $S = \cup_{i=1}^{n} B_i$, where the sequence $\{B_i\}_{i=1}^{n}$ is mutually exclusive. Then,

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

- **Bayes' Theorem (1.6):**

$$P\left(B_j\middle|A\right) = \frac{P\left(A\middle|B_j\right)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

Example

- Going back to our Dice example: Suppose we roll a fair six-sided die once. Let A denote the event of rolling a number less than 4 and let B denote the event of rolling an odd number.

- Given that we rolled an odd number, what is the probability that the number obtained is less than 4?

Example: The Monty Hall Problem

- Monty Hall was the host of an old gameshow and one of the games worked like this.
- There are 3 doors: behind one of them is a car and behind two of them are goats. The contestant chooses a door. Monty opens one of the other (non-chosen) doors, revealing a goat, and offers the contestant a chance to switch to the other non-open door. What should the contestant do?

Chapter 2: Random Variables

- <u>Chapter outline</u>
- Discrete and Continuous Random Variables

- Expectation of a Random Variable

- Jointly distributed Random Variables

- Moment Generating functions

- **<u>Definition of a Random Variable (2.1 and 2.2):</u>**
- A random variable is a variable whose value is a numerical outcome of a random phenomenon.
- The random variable X is a function mapping outcomes to real numbers: $X: S \to \mathbb{R}$

- **Discrete type**: A random variable (rv) X takes on a finite or countable number of possible values.
- **Probability Mass Function (pmf):** $p(a) = P(X = a)$

$$\sum_{all\ x} p(x) = 1$$

- **Cumulative distribution function (cdf):**

$$F(a) = P(X \leq a) = \sum_{x \leq a} p(a)$$

- Recall:
- If X takes on values in the set $\{a_1, a_2, a_3, \ldots\}$ where $a_1 < a_2 < a_3 < \cdots$ such that $p(a_i) > 0\ \ \forall i$, then we have

$$p(a_1) = F(a_1),$$

$$p(a_i) = F(a_i) - F(a_{i-1}), i = 2,3,4, \ldots$$

Special Discrete Distributions

Name		Probability Function	Range
$\text{Bernoulli}(p)$	$x = \begin{cases} 0 & \text{if failure} \\ 1 & \text{if success} \end{cases}$	$p(x) = p^x(1-p)^{n-x}$	$x = 0,1$
Binomial $Bin(n, p)$	# of successes in $n \in \mathbb{Z}^+$ independent Bernoulli trials	$p(x) = \binom{n}{x} p^x(1-p)^{n-x}$	$x = 0, 1, 2, \ldots, n$
Negative Binomial $NB(k, p)$	# of Bernoulli trials required to observe $\mathrm{k} \in \mathbb{Z}^+$ "successes"	$p(x) = \binom{x-1}{k-1} p^k(1-p)^{x-\mathrm{k}}$	$x = \mathrm{k}, \mathrm{k}+1, \mathrm{k}+2, \ldots$

Name		**Probability Function**	**Range**
Geometric $X \sim Geo(p)$ $X \sim NB(1, p)$	# of Bernoulli trials required to observe the first "success"	$p(x) = p(1-p)^{x-1}$	$x = 1, 2, 3, \dots$
Hypergeometric $X \sim HG(N, r, n)$	# of "successes" in n draws without replacement from a finite population of size N having exactly r "successes"	$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$	$x = \max\{0, \mathrm{n} - \mathrm{N} + \mathrm{r}\}, \dots, \min\{\mathrm{n}, \mathrm{r}\}$
Poisson $X \sim Poi(\lambda)$ $\lambda > 0$	# of occurrences observed within a specified interval of length *t*.	$p(x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$	$x = 0, 1, 2, \dots$

- **<u>Continuous type(2.3)</u>**: A rv X takes on a continuum of possible values (which is uncountable).

- **Probability density Function (pdf):** $f(x) = \frac{d}{dx}F(x) = F'(x)$

Which is a non-negative real-valued function that satisfies

$$P(X \in B) = \int_{x \in B} f(x)dx$$

Where B is a set of real numbers

With $\int_{all\ x} f(x)dx = 1$ i.e total area under the density curve=1

- **Cumulative distribution function (cdf):** $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$

Special Continuous Distributions

Name		Probability Function	Range
Uniform $U(a,b)$	X is equally likely to take on any value within the interval (a,b)	$f(x) = \frac{1}{b-a}$	$a < x < b$ Where $a, b \in \mathbb{R}$ with $a < b$
Beta* $Beta(m,n)$		$f(x) = \frac{(m+n-1)!}{(m-1)!\,(n-1)!} x^{m-1}(1-x)^{n-1}$	$0 < x < 1$

* Note: if we let m=n=1, then Beta(m,n) distribution becomes a U(0,1)

Name	**Probability Function**	**Range**
Gamma $Gam(\alpha, \lambda)$ $\alpha \in \mathbb{Z}^+$ and $\lambda > 0$	$f(x) = \dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$	$x > 0$
Exponential# $Exp(\lambda)$	$f(x) = \lambda e^{-\lambda x}$	$x > 0$

Note: if we let n=1, then Gam(1,λ) distribution becomes an $Exp(\lambda)$

- Recall:

The <u>memoryless property</u>

$$P(X > t + s | X > s) = P(X > t)$$

- **<u>Expectation (2.4)</u>:**

- If $g(.)$ is an arbitrary real-valued function, then

$$E[g(X)] = \begin{cases} \sum_{x} xg(x) & \text{,if X is a discrete rv} \\ \int_{-\infty}^{\infty} xg(x)dx & \text{,if X is a continuous rv} \end{cases}$$

- **Special Choices of $\boldsymbol{g(.)}$:**

1. $g(X) = X^n$ for $n \in \mathbb{N} \Longrightarrow E[g(X)] = E[X^n]$ is the n^{th} moment of X. In general, moments serve to describe the shape of a distribution.

2. $g(X) = (X - E[X])^2 \Longrightarrow E[g(X)] = E[(X - E[X])^2]$ is the variance of X.

Recall that $Var(X) = \sigma_X^2 = E[X^2] - E[X]^2$ and

Standard deviation of X is $\sqrt{Var(X)} = \sigma_X$

3. $g(X) = aX + b, \quad a, b \in \mathbb{R}$ we have

$$\mu_{aX+b} = E[aX + b] = aE[X] + b$$

$$\sigma^2_{aX+b} = Var(aX + b) = a^2 Var(X)$$

4. Moment generating Functions (2.6):

- $g(X) = e^{tX}, t \in \mathbb{R} \Longrightarrow E[g(X)] = E[e^{tX}] = \phi_X(t)$ is the moment generating function (mgf) of X.

Note: $\phi_X(0) = E[e^{0X}] = E[1] = 1$ and,

$$\boxed{E[X^n] = \phi_X^{(n)}(0) = \frac{d^n}{dt^n}\phi_X(t)|_{t=0} = \lim_{t\to 0}\frac{d^n}{dt^n}\phi_X(t)\,, n \in \mathbb{N}}$$

Represents the n^{th} moment of X.

- Recall: A mgf **uniquely** determines the distribution of a rv, i.e there exists a one-to-one correspondence between the mgf and the pmf/pdf of a rv.

Example

- For $X \sim Poi(\lambda)$, show that the mgf X is $\phi_X(t) = e^{\lambda(e^t - 1)}, t \in \mathbb{R}$. And use it to find E[X] and Var(X)

Discrete Distribution	Probability Mass Function of X	Mean $\mathrm{E}[X]$	Variance $\mathrm{Var}(X)$
$\mathrm{BIN}(n,p)$	$p(x)=\binom{n}{x}p^x(1-p)^{n-x}$, $x=0,1,...,n$	np	$np(1-p)$
$\mathrm{Bernoulli}(p)$	$p(x)=p^x(1-p)^{1-x}$, $x=0,1$	p	$p(1-p)$
$\mathrm{HG}(N,r,n)$	$p(x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, $x=\max\{0,n-N+r\},...,\min\{n,r\}$	$\frac{nr}{N}$	$\frac{nr(N-r)(N-n)}{N^2(N-1)}$
$\mathrm{POI}(\lambda)$	$p(x)=\frac{e^{-\lambda}\lambda^x}{x!}$, $x=0,1,2,...$	λ	λ
$\mathrm{NB}(k,p)$	$p(x)=\binom{x-1}{k-1}p^k(1-p)^{x-k}$, $x=k,k+1,k+2,...$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
$\mathrm{GEO}(p)$	$p(x)=(1-p)^{x-1}p$, $x=1,2,3,...$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Continuous Distribution	Probability Density Function of X	Mean $E[X]$	Variance $Var(X)$
$U(a,b)$	$f(x) = \frac{1}{b-a}$, $a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Beta(m,n)$	$f(x) = \frac{(m+n-1)!}{(m-1)!(n-1)!} x^{m-1}(1-x)^{n-1}$, $0 < x < 1$	$\frac{m}{m+n}$	$\frac{mn}{(m+n)^2(m+n+1)}$
$Erlang(n,\lambda)$	$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$, $x > 0$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$
$EXP(\lambda)$	$f(x) = \lambda e^{-\lambda x}$, $x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

- **<u>Jointly Distributed Random Variables (2.5)</u>:**
- We will consider the bivariate case only but all these results extend naturally to an arbitrary number of random variables.
- The **joint cdf** of X and Y is

$$
\begin{aligned}
F(a,b) &= P(X \le a, Y \le b) \\
&= P(\{X \le a\} \cap \{Y \le b\}), \qquad a, b \in \mathbb{R}
\end{aligned}
$$

- Where the joint cdf can be used to solve for the **marginal cdfs**

$$F_X(a) = P(X \le a) = F(a, \infty) = \lim_{b \to \infty} F(a,b)$$

Similarly

$$F_Y(b) = P(X \le b) = F(\infty, b) = \lim_{a \to \infty} F(a,b)$$

- ***Jointly Discrete Case***
- **Joint pmf:**

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

- **Marginal pmfs:**

$$p_X(x) = P(X = x) = \sum_y p(x, y)$$

$$p_Y(y) = P(Y = y) = \sum_x p(x, y)$$

- ***Jointly Continuous Case:***
- **Joint pdf:** The joint pdf $f(x, y)$ is a non-negative real-valued functions which enables one to calculate probabilities of the form

$$P(X \in A, Y \in B) = \int_B \int_A f(x,y)dxdy = \int_A \int_B f(x,y)dydx$$

Where A and B are sets of real numbers (e.g intervals).

- Hence we have

$$F(a,b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)dxdy = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)dydx$$

- **Marginal pdfs:**

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

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

Important property:

$$\boxed{f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)}$$

- Expectation:
- If $g(.,.)$ denotes an arbitrary real-valued function, then

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y) & \text{, If X and Y are jointly discrete} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,dx\,dy & \text{, If X and Y are jointly continuous} \end{cases}$$

- **Special choices of g(.,.):**

1. $g(X,Y) = (X - E[X])(Y - E[Y]) \Longrightarrow E[g(X,Y)]$
$= E[(X - E[X])(Y - E[Y])]$ is the covariance of X and Y

Note that:

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

And $Cov(X,X) = Var(X)$

2. $g(X,Y) = aX + bY, a, b \in \mathbb{R}$ we have

$$E[aX + bY] = aE[X] + bE[Y],$$

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$$

- **Independence of Random Variables:**
- X and Y are independent if and only if (iff)

$$F(a, b) = P(X \le a, Y \le b) = P(X \le a)P(Y \le b) = F_X(a)F_Y(b),$$

$\forall a, b, \in \mathbb{R}$

- Equivalently, independence exists iff $p(x, y) = p_X(x)p_Y(y)$ (in the jointly discrete case) or $f(x, y) = f_X(x)f_Y(y)$ (in the jointly continuous case) $\forall x, y, \in \mathbb{R}$

- <u>Recall</u>:
- if **X and Y are independent**

1. Then for arbitrary real-valued functions $g(.)$ and $h(.)$,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

2. Then $Cov(X, Y) = 0$

Note: if $Cov(X, Y) = 0$ we cannot conclude that X and Y are independent, only that they are *uncorrelated*.

3. If $X_1, X_2, \dots, X_n$ are independent random variables where $\phi_{X_i}(t)$ is the mgf of $X_i, i = 1,2,\dots,n$ then $T = \sum_{i=1}^{n} X_i$ has mgf

$$\phi_T(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$$

- In other words, the mgf of a sum of independent random variables is the product of their individual mgfs.
- In the case that $X_1, X_2, \dots, X_n$ are iid random variables:

$$\phi_T(t) = \prod_{i=1}^{n} \phi_{X_i}(t) = \phi_{X_1}(t)^n$$

4. The Strong Law of Large Numbers (SLLN):

If $X_1, X_2, \dots, X_n$ is an iid sequence of random variables, each having a finite mean $\mu = E[X_i]$, then with probability 1,

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \rightarrow \mu \; as \; n \rightarrow \infty$$

- In other words, the sample mean will, with probability 1, converge to the true mean of the underlying distribution as the sample size approaches infinity.