PSTAT 120A Notes: Probability and Statistics I

Alex Mei

Winter, 2020

1 Counting

Counting Principle: Suppose the outcome of some experiment has k independent components. Then, the total number of outcomes is the product of the outcomes of each component, denoted by n_i.

$$\prod_{i=1}^{k} n_i$$

Addition Principle: Suppose a collection of outcomes is partitioned into k subcollections. Then, the total number of outcomes is the sum of the outcomes of each subcollection, denoted by n_i.

$$\sum_{i=1}^{k} n_i$$

Permutation Principle: Given n distinguishable objects, there are n! ways to arrange them.

$$0! = 1; (n > 0) : n! = \prod_{i=1}^{n} i$$

Distinguishable Principle: In a collection of n total outcomes, each outcome is part of a family of k distinguishable outcomes. The total number of distinguishable outcomes is equal to:

$$\frac{n}{k}$$

Choose Notation:

$$n\mathbf{C}k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Choosing Principle: The number of ways to choose k objects out of n objects where order doesn't matter is:

$$\binom{n}{k} = \binom{n}{n-k}$$

Note: When the order does matter, it is considered a permutation.
Permutation Notation:

$$n\mathbf{P}k = \binom{n}{k}k! = \frac{n!}{(n-k)!}$$

Word Counting Principle: Given n letters, k of which are distinct, such that the first letter repeated r_i times up to the kth letter repeated r_k times, the total number of words that can be creased from the n letters is:

$$\binom{n}{r_1, r_2, ..., r_k} = \frac{n!}{(r_1)!(r_2)!...(r_k)!} = \binom{n}{r_1}\binom{n-r_1}{r_2}...\binom{n-r_1-r_2-...r_{k-1}}{r_k}$$

Binomial Expansion: Given the polynomial $(x+y)^n$, the binomial expansion of the polynomial is:

$$\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$$

Note: when x and y both equal 1, then:

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

Non-Negative Integer Solutions: The number of ways to divide n indistinguishable objects into k bins (where bins may have 0 objects) is:

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

Positive Integer Solutions: The number of ways to divide n indistinguishable objects into k bins where each bin must have at least i objects is:

$$\binom{(n-i*k)+(k-1)}{k-1}$$

2 Set Theory

2.1 Sets

Set: an unordered collection of unique objects called **elements** or **members** denoted by {...}
Member Notation: a member a in set S is denoted by $a \in S$
Equal Sets: two sets A and B are equal if and only if $\forall x\ (x \in A \leftrightarrow x \in B)$; if $A \subseteq B$ and $B \subseteq A$, then $A = B$
Empty Sets: also known as a **null set**, an empty set has no elements denoted by $\emptyset$

Universal Set: denoted by the letter U, this set contains all the elements under consideration represented by a rectangle in a Venn Diagram
Cardinality: if there are exactly n distinct elements of set S, then the cardinality of the **finite** set S is n, denoted by $\mid S \mid$ = n
Subset: The set A is a subset of B, denoted by A $\subseteq$ B, if and only if every element in A is in B
Proper Subset: The set A is a proper subset of B, denoted by A $\subset$ B, if A $\subseteq$ B and there exists an element in B that isn't in A: $\forall$x (x $\in$ A $\rightarrow$ x $\in$ B) $\wedge$ $\exists$x (x $\in$ B $\rightarrow$ x $\notin$ A)

2.2 Set Operations

Union: The union of two sets A and B, denoted by A $\cup$ B is the set that contains the elements in A, B, or both.

$$A \cup B = \{x \mid x \in A \vee x \in B\}$$

Generalized Union: The union of a collections of sets is the set that contains all elements that are in at least one set of the collection.

$$A_1 \cup A_2 \cup ...A_n = \bigcup_{i=1}^{n} A_i$$

Intersection: The intersection of two sets A and B, denoted by A $\cap$ B is the set that contains the elements in both A and B.

$$A \cap B = \{x \mid x \in A \wedge x \in B\}$$

Generalized Intersection: The intersection of a collections of sets is the set that contains all elements that are in every set of the collection.

$$A_1 \cap A_2 \cap ...A_n = \bigcap_{i=1}^{n} A_i$$

Complement: The complement of the set A, denoted by $\overline{A}$ or $A^{\complement}$, is the complement of A with respect to U (the universal set.)

$$A^{\complement} = \{x \in U \mid x \notin A\}$$

Disjoint: Two sets A and B are considered disjoint if their intersection is the empty set.

$$A \cap B = \emptyset$$

Pairwise Disjoint: An indexed family of sets $\{A_i\}_{i=1}^{n}$ is considered pairwise disjoint or **mutually exclusive** if $A_i \cap A_j = \emptyset$ whenever i $\neq$ j.

Partition: Let S be a non-empty set. A partition π of S is a family $\pi = \{A_i\}_{i=1}^{n}$ of non-empty subsets of S satisfying the condition that every element in S belongs to exactly 1 A_i.

$$\bigcup_{i=1}^{n} A_i = S$$

$$A_i \neq A_j; i \neq j$$

De Morgan's Laws:

$$(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$$

$$(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}$$

De Morgan's Laws, Generalized:

$$(\bigcup_{i=1}^{n} A_i)^{\complement} = \bigcap_{i=1}^{n} A_i^{\complement}$$

$$(\bigcap_{i=1}^{n} A_i)^{\complement} = \bigcup_{i=1}^{n} A_i^{\complement}$$

3 Probability

3.1 Basic Probability

Sample Space: A sample space, Ω, is a set which represents the collection of all possible outcomes of an experiment.
Event: An event, E, is a subset of the sample space Ω (to which we prescribe a probability).
Probability: Let Ω be a sample space. Then, a probability P on Ω is a function assigning a number to each event such that the axioms of probability are true.
Axioms of Probability:

1. $0 \leq$ P(E) ≤ 1 for every event E

2. P(Ω) = 1

3. If $\{E_1, E_2, ..., E_n\}$ are pairwise disjoint, then $P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$.

Properties of Probability:

- $P(E^c) = 1 - P(E)$

- If $E \subseteq F$, then $P(E) \leq P(F)$

- **Inclusion Exclusion Principle (for two events):**

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- **Inclusion Exclusion Principle (for three events):**

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- **Inclusion Exclusion Principle, Generalized:**

$$P(E_1 \cup E_2 \cup ...E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i<j} P(E_i \cap E_j) + ... + (-1)^{n+1} P(E_1 \cap E_2 \cap ...E_n)$$

Probability Space: A space defined to be (Ω, F, P) where Ω is the sample space, F is the set of all possible events, and P is probability associated with each event.
Set of All Events: Given the sample space Ω, the set of all events F is the set of all subsets of Ω.
Probability Theorem: Suppose (Ω, F, P) is a finite probability space such that all outcomes are equally likely. Then, for any event E:

$$P(E) = \frac{\mid E \mid}{\mid \Omega \mid}$$

3.2 Conditional Probability

Conditional Probability: Let (Ω, F, P) be a probability space. Suppose B is some event where P(B) > 0. Then, the conditional probability of A $\subseteq$ Ω given B is:

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

Multiplication Rule (Special Case): (derived from the formula of conditional probability)

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

Multiplication Rule: Let E_1, ..., E_n. Then,

$$P(E_1 \cap E_2 \cap ...E_n) = P(E_1) * P(E_2 \mid E_1) * ... * P(E_n \mid E_1 \cap E_2 \cap ...E_{n-1}).$$

Total Law of Probability: If E_1, ..., E_n creates a partition of Ω, then...

$$P(F) = \sum_{i=1}^{n} P(E_i) * P(F \mid E_i)$$

Baye's Formula (Simple Version):

$$P(E \mid F) = \frac{P(E) * P(F \mid E)}{P(E) * P(F \mid E) + P(E^{\complement}) * P(F \mid E^{\complement})}$$

Generalized Baye's Formula: If {E_1, ..., E_n} partitions Ω and F is any event, then...

$$P(E_k \mid F) = \frac{P(E_k) * P(F \mid E_k)}{\sum_{i=1}^{n} P(E_i) * P(F \mid E_i)}$$

3.3 Independence

Independent Events: Two events E and F are independent if...

$$P(E \cap F) = P(E) * P(F)$$

Note: If E and F are independent, then E and $F^{\complement}$, $E^{\complement}$ and F, and $E^{\complement}$ and $F^{\complement}$ are all independent.
Generalized Independence: Suppose several events E_1, ..., E_n are independent. Then, $P(E_i \cap E_j) = P(E_i) * P(E_j)$ whenever $i \neq j$ and $P(E_i \cap E_j \cap E_k) = P(E_i) * P(E_j) * P(E_k)$ whenever $i \neq j \neq k$ and ... until $P(E_1 \cap E_2 \cap ... E_n) = P(E_1) * P(E_2) * ... P(E_n)$

4 Random Variables

Random Variable: Let Ω be a sample space. A (real-valued) random variable is a function of X: Ω → R.
State Space: $S_X \subseteq R$ is the image of X; That is $X(\Omega) = S_X$ = all possible outcomes of X. (Transformation of the sample space by X.)
Cumulative Distribution Function: Abbreviated as a CDF, if X is a random variable, then the cdf of X, denoted, F_X, is the function $F_X : R \to [0, 1]$, where $F_X(t) = P(X \leq t)$.
Properties of CDF: $F_X(t)$ denotes the probability the the random variable X takes on a value that is less than or equal to t. The following properties are true:

- F is a non-decreasing function.
- $\lim_{t \to \infty} (F_X(t)) = 1$
- $\lim_{t \to -\infty} (F_X(t)) = 0$

4.1 Discrete Random Variables

Discrete Random Variable: A discrete random variable X is one where the state space of X, S_X, is countable.

Probability Mass Function: Abbreviated as pmf, the probability mass function of discrete random variable X is the function $p_X : S_X \rightarrow [0, 1]$, with $p_X(k) = P(X = k)$.
Note: $P(X{=}k) = F_X(k^+) - F_X(k^-)$
Note: The cdf of a discrete random variable looks like a step function with jumps at each point in S_X where the gap of the jump at $s \in S_X$ is $p_X(s) = P(X = s)$.
Bernoulli Random Variable: Suppose that one trial or experiment is performed where the result ends in success or failure. Let X = 1 denote a success and X = 0 denote a failure. Let $p \in [0, 1]$ be the probability of success. Then, the state space $S_X = \{0, 1\}$ and the pmf is $p_X(0) = 1 - p$ and $p_X(1) = p$. Any random variable with this property is called a Bernoulli Random Variable and is denoted by X = Ber(p)
Binomial Random Variable: Suppose we do n independent trials where each result ends in success or failure and the probability of success is constant, $p \in [0, 1]$. Let X be the number of successes in n trials. Then, X is a Binomial Random Variable denoted by X=Bin(n, p) with $S_X = \{0, 1, 2, ..., n\}$ and

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

Poisson Random Variable: Given the average rate of of the event, denoted by λ, and the events are independent of time since the last event, denoted X = Pois(λ)

$$S_X = \{0, 1, 2, ...\}$$

$$p_X(k) = e^{-\lambda} * \frac{\lambda^k}{k!}$$

Geometric Random Variable: Used to find the number of independent trials needed before the first success, denoted X = Geom(p) where the probability of success, $p \in [0, 1]$

$$S_X = 1, 2, ...$$

$$p_X(k) = (1-p)^{k-1} * p$$

Indication Function: An indicator function I for some random variable X will have probability of P of success, denoted by I_x, is...

$$I = \begin{cases} 0 & x\ occurs \\ 1 & x^{\complement}\ occurs \end{cases}$$

where I = Ber(p) and P(X) = p.

4.2 Expected Value

Expected Value: The expected value of a random variable X with CDF $F_x(t) = P(X \leq t)$, denoted by E(X), is

$$E(X) = \int_0^\infty (1 - F_X(t))dt - \int_{-\infty}^0 F_X(t)dt = \int_\Omega X(\omega)dP(\omega)$$

Linearity: Expected value is a linear function which is preserves addition and scalar multiplication.

$$E(a_n * X^n + a_{n-1} * X^{n-1} + ...a_0) = a_n * E(X^n) + a_{n-1} * E(X^{n-1}) + ... + a_0$$

Expected Values of Common Discrete RV:

- $E(Bin(n, p)) = n * p$
- $E(Pois(\lambda)) = \lambda$
- $E(Geom(p)) = \frac{1}{p}$

Discrete EV Theorems:

- If X is a discrete random variable, then

$$E(X) = \sum_{k \in S_X} k * p_X(k)$$

- For a function g : R → R and a discrete random variable X, then

$$E(g(X)) = \sum_{k \in S_X} g(k) * p_X(k)$$

- For two random variables X and Y in the same sample space,

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

Moment: The nth moment of the discrete random variable X is the expectation $E(X^n)$.

$$E(X^n) = \sum_{k \in S_X} k^n * p_X(k)$$

Variance: The variance of a random variable X is defined as $E((X - \mu)^2)$ where $\mu = E(X)$, which conveys how spread out the values of X are from its mean.

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

Non-linearity: Variance is a non-linear function. In fact, $Var(a * X + b) = a^2 * Var(x)$.
Variance of a Constant: The variance of any constant is 0.
Variances of Common Discrete RV:

- $Var(Bin(n,p)) = n * p * (1 - p)$
- $Var(Pois(\lambda)) = \lambda$
- $Var(Geom(p)) = \frac{1-p}{p^2}$

4.3 Approximating Binomial with Poisson

Poisson Limit Theorem: Also known as the **Law of Rare Events:**, Suppose that $\lim_{n \to \infty} n * p = \lambda < \infty$ (p approaches 0 and n approaches infinity in such a way that n * p approaches lambda is finite). Then, $P(Bin(n,p) = k) = P(Pois(\lambda) = k)$ as $n \to \infty$.
Conditions for Poisson Approximation

- n is large
- n * p << n

4.4 Continuous Random Variables

Continuous Random Variable: X is a continuous random variable when there exists a function f_X: R → R such that $F_X(t)$ is continuous.

$$F_X(t) = \int_{-\infty}^{t} f_X(s)ds$$

Probability Density Function: The pdf of a random variable x is denoted by $f_X(t)$.
PDF Properties:

- for all t, $f_X(t) \geq 0$
- $\int_{-\infty}^{\infty} f_X(t)dt = 1$
- $\frac{d}{dt}F_X(t) = f_X(t)$
- P(X = k) = 0

Uniform Random Variable: A continuous random variable X, denoted X = Unif(a, b) = U(a, b) is said to be uniformly distributed on [a, b] if X has pdf:

$$f_X(t) = \begin{cases} \frac{1}{b-a} & a \leq t \leq b \\ 0 & otherwise \end{cases}$$

Note: the strictness of end points don't matter for continuous random variables.
Note: A ***discrete uniform random variable*** *is* $Z \in [a, b]$.
Exponential Random Variables: An exponential random variable describes the amount of time until a specific event occurs. X is an exponential random variable with parameter $\lambda > 0$ if it has pdf:

$$f_X(t) = \begin{cases} \lambda * e^{-\lambda * t} & t \geq 0 \\ 0 & otherwise \end{cases}$$

Normal Random Variables: X is a normal random variable (X is normally distributed with parameters $\mu = E(X)$ and $\sigma^2 = Var(X)$), denoted by X = N(μ, σ^2), if the density of X is given by:

$$f_X(t) = \frac{1}{\sqrt{2\pi * \sigma^2}} * e^{\dfrac{-(t-\mu)^2}{2\sigma^2}}$$

Standard Deviation: The standard deviation of a random variable is denoted by σ.
Standard Normal Random Variable: Z is defined as a standard normal random variable when Z = N($\mu = 0$, $\sigma = 1$)
Note: A change of variables from X to Z can be made such that $Z = \dfrac{X - \mu}{\sigma}$.
CDF of Standard Normal RV: $F_z(a) = \Phi(a)$.
Property of $\Phi(a)$: $\Phi(-a) = 1 - \Phi(a)$

4.5 Expected Value for Continuous Random Variables

Expected Value Formulas: When X is a continuous random variable,

$$E(X) = \int_{-\infty}^{\infty} t * f_X(t)dt$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(t) * f_X(t)dt$$

4.6 Poisson Process

Poisson Process: A process which counts the number of events at which these events occur in a given time interval. When X = Exp(1/λ), E(X) = 1 / λ and X=Pois(λ) where E(X) is the average number of events in a given time interval.
Memoryless Property: We say a non-negative random variable X is memoryless if

$$\forall s, t \geq 0[P(X > s + t \mid X > t) = P(X > s)]$$

Note: Exp(λ) and Geom(p) is memoryless

4.7 Functions of Random Variables

Process: To find the pdf of Y given a continuous random variable X such that Y = g(X), follow the following 3 steps:

1. Find the cdf of Y in terms of the cdf of X or in terms of t.
2. Take derivative to get the pdf of Y.
3. Find bounds where defined.

Note: this method words for non-continuous random variables, too.

5 Jointly Distributed Random Variables

Jointly Distributed: X and Y are jointly distributed if they are both random variables on the same sample space Ω. The joint CDF of X and Y is...

$$F_{X,Y}(s,t) = P(X \leq s, Y \leq t)$$

Marginal Distribution: Given jointly distributed random variables X and Y, the marginal distribution of X is...

$$F_X(s) = P(X \leq s, Y < \infty)$$

5.1 Jointly Discrete

Jointly Discrete: If random variables X and Y are jointly discrete, then the joint pmf of X and Y is...

$$p_{X,Y}(k,m) = P(X = k, Y = m)$$

Marginal Mass Function: Given jointly discrete random variables X and Y, the marginal mass function of Y is...

$$p_y(m) = \sum_{k \in S_X} p_{X,Y}(k,m)$$

Joint CDF: The joint cdf of jointly discrete random variables X and Y is...

$$F_{x,y}(k,m) = \sum_{a \leq k} \sum_{b \leq m} p_{X,Y}(a,b)$$

Joint Expectation: The joint expectation of jointly discrete random variables X and Y is...

$$E(g(x,y)) = \sum_{k \in S_X} \sum_{m \in S_Y} g(k,m) * p_{X,Y}(k,m)$$

5.2 Jointly Continuous

Jointly Continuous: If X and Y are random variables on a common sample space Ω, X and Y are jointly continuous when there is a function $f : R^2 \to R_0^+$ such that...

$$P((x,y) \in D) = \int \int_D f_{X,Y}(x,y) dA$$

Joint Density: If X and Y are jointly continuous random variables, then the joint density is...

$$f_{X,Y}(x,y) = \frac{\delta}{\delta x \delta y} F_{X,Y}(x,y)$$

Joint CDF: The joint cdf of jointly continuous random variables X and Y is...

$$F_{X,Y}(a,b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f_{X,Y}(x,y) dx dy$$

Marginal Density: The marginal density of X for jointly continuous random variables X and Y is...

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

5.3 Independent Random Variables

Independent: Random variables X and Y are independent when

$$P(X \in A, Y \in B) = P(X \in A) * P(Y \in B)$$

Independence Theorem: If X and Y are jointly distributed random variables, then the following statements are equivalent:

- X and Y are independent
- $F_{X,Y}(a,b) = f_X(a) * f_Y(b)$
- If X and Y are discrete, then $p_{X,Y}(a,b) = p_X(a) * p_Y(b)$
- If X and Y are continuous, then $f_{X,Y}(a,b) = f_X(a) * f_Y(b)$

5.4 Conditional Distributions

Notation: The following denotes the conditional mass and conditional distribution functions of X given Y = m respectively for the discrete case:

$$p_{X|Y}(k \mid m) = P(X = k \mid Y = m)$$

$$F_{X|Y}(k \mid m) = P(X \leq k \mid Y = m)$$

For the continuous case,

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

$$F_{X|Y}(a \mid y) = \int_{-\infty}^{a} f_{X|Y}(x \mid y)dx$$

Independence: If joint random variables X and Y are independent, then the conditional mass and density functions are:

$$p_{X|Y}(k \mid m) = p_X(m)$$

$$f_{X|Y}(x \mid y) = f_X(x)$$

5.5 Conditional Expectation

Conditional Expectation: Suppose Y is a discrete random variable. Then the expectation of Y given event A is:

$$E(Y \mid A) = \sum_{m \in S_Y} m * P(Y = m \mid A)$$

Given Random Variable X = x instead of event A, the conditional expectation of Y is (for the jointly discrete case)...

$$E(Y \mid X = x) = \sum_{m \in S_Y} m * P(Y = m \mid X = x)$$

For the jointly continuous case...

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} f_{Y|X}(y \mid x)dy$$

Properties of Conditional Expectation:

- Linearity persists: $E(X + Y \mid A) = E(X \mid A) + E(Y \mid A)$
- Total Law of Expectation: $E(Y) = \sum_{i=1}^{n} P(A_i) * E(Y \mid A_i)$ for partition of $\{A_i\}$.

5.6 Sums of Independent Random Variables

Convolution: For independent random variables X and Y, the joint pmf or density is found by breaking {X + Y = n} into a disjoint union of sets. For the discrete case,

$$p_{X+Y}(n) = \sum_{k \in S_X} p_X(k) * p_Y(n-k)$$

Special Cases:

- $Pois(\lambda_1) + Pois(\lambda_2) = Pois(\lambda_1 + \lambda_2)$
- $Bin(n, p) + Bin(m, p) = Bin(n + m, p)$
- $N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Independent and Identically Distributed: Random variables X and Y are independent and identically distributed (iid) if they are independent and share the same probability distribution.
Negative Binomial Distribution: A random variable X has a negative binomial distribution, denoted X = NegBin(k, p), with parameters k and p where k is nonnegative, if $S_x = \{k, k+1, ...\}$ with pmf for n ≥ k (where n denotes number of trials until the kth success):

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

5.7 Covariance

Covariance: The covariance of jointly distributed random variables X and Y, denoted Cov(X, Y), is...

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

Properties of Covariance:

- Cov(X, Y) = Cov(Y, X)
- Cov(X, X) = Var(X)
- Cov(a*X + b, Y) = a*Cov(X, Y)
- Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)
- If X and Y are independent, then Cov(X, Y) = 0

Note: Cov(X, Y) = 0 doesn't imply X and Y are independent.
Variance Theorem: Let X and Y be random variables. Then,

$$Var(X) + Var(Y) + 2 * Cov(X, Y) = Var(X + Y)$$

Note: When X and Y are independent, Cov(X, Y) = 0 so Var(X) + Var(Y) = Var(X + Y)
General Case: For n random variables,

$$Var(X_1 + X_2 + ...X_n) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} Cov(x_i, x_j)$$

Correlation: Suppose that X and Y are jointly distributed random variables with nonzero variances. Then, the correlation of X and Y, denoted Corr(X, Y), is:

$$Corr(X, Y) = r = \frac{Cov(X, Y)}{\sqrt{Var(X)} * \sqrt{Var(Y)}}$$

Note: If Corr(X, Y) = 0, we say X and Y are uncorrelated.
Note: If X and Y are independent, then Corr(X, Y) = 0.
Note: Corr(X, Y) = 0 doesn't imply X and Y are independent.

5.8 Moment Generating Functions

Moment Generating Function: The moment generating function (MGF) of a random variable X with real parameter t (where t is the t-th moment) is defined by...

$$M_X(t) = E(e^{t*X})$$

MGF Theorem: Let X and Y be 2 random variables with MGFs $M_X(t)$ and $M_Y(t)$. If $M_X(t) = M_Y(t)$, then X and Y have the same distribution. (Random variables have unique MGFs)