MATH 394 Notes

Contents

1 Counting

1.1 Sum Rule

If you can choose from one of n choices or one of m choices then the total number of outcomes is $n + m$

1.2 Product Rule

If each outcome is constructed by a sequential process where there are

- n_1 choices for the first step
- n_2 choices for the second step (given the choice for the first step)
- n_k choices for the k^{th} step (given the choice for the previous step)

then the total number of outcomes is $n_1 \times n_2 \times ... \times n_k$

1.3 Power Set

The power set of a set A is the set of all subsets of A, including the empty set and A itself

- $P(A) = \{S \mid S \subseteq A\}$
- $P(\varnothing) = \{\varnothing\}$
- $P(\{x, y\}) = \{\varnothing, \{x\}, \{y\}, \{x, y\}\}$

A set with n elements has 2^n power sets

1.4 Permutations

There are $n!$ ways to order n distinct objects

1.5 Complementary Counting

Let U be a set and S a subset of interest. Let $U \backslash S$ denote the set difference. Then $|U \backslash S| = |U| - |S|$

1.6 ${}^nP_k$ Permutations

There are ${}^nP_k = \frac{n!}{k!}$ ways to *arrange* k out of n distinct objects without repetition

- n permute k

1.7 ${}^nC_k$ Combinations

There are ${}^nC_k = \binom{n}{k} = \frac{n!}{(n-k)! \times k!}$ ways to *choose* k out of n distinct objects without repetition

- n choose k

1.8 Combinatorial Argument/Proof

- Let S be a set of objects
- Show how to count $|S|$ one way, let $|S| = M$
- Show how to count $|S|$ another way, let $|S| = N$
- Then $M = N$

1.9 Binomial Theorem

Let $x, y \in \mathbb{R}$ and $n \in \mathbb{N}$ a positive integer, then

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

- Symmetry in Binomial Coefficients $\binom{n}{k} = \binom{n}{n-k}$
- Pascal's Identity $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
- Application of Binomial Theorem $\sum_{k=0}^{n} \binom{n}{k} = 2^n$

1.10 Inclusion-Exclusion

If $A_1, A_2, ..., A_N$ are sets, then

$$\begin{aligned} |A_1 \cup A_2 \cup ... \cup A_n| &= \text{singles} - \text{doubles} + \text{triples} - \text{quads} + ... \\ &= (|A_1| + ... + |A_n|) - (|A_1 \cap A_2| + ... + |A_{n-1} \cap A_n|) + ... \end{aligned}$$

1.11 Pigeonhole Principle

If there are n pigeons in $k < n$ holes, then one hole must contain at least $\left\lceil \frac{n}{k} \right\rceil$ pigeons

To use the Pigeonhole Principle

1. Identify pigeons
2. Identify pigeonholes
3. Specify how pigeons are assigned to pigeonholes
4. Apply Pigeonhole Principle

1.12 Sleuth's Criterion

For each object constructed, it should be possible to reconstruct the unique sequence of choices that led to it

- If an example has no sequence, then we are undercounting
- If an example has multiple sequences, then we are overcounting

2 Probability

2.1 Sample Space

A sample space Ω is the set of all possible outcomes of an experiment

2.2 Events

An event $E \subseteq \Omega$ is a subset of possible outcomes

- Events E and F are mutually exclusive if $E \cap F = \varnothing$

2.3 Probability Measure

A probability measure is a function $P : \omega \to [0, 1]$ such that

- $\mathbb{P}(\omega) \geq 0$ for all $\omega \in \Omega$
- $\displaystyle\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$

2.4 Probability Space

A probability space is a pair $(\omega, \mathbb{P})$ where

- ω is a set called the sample space
- $\mathbb{P}$ is the probability measure

If $(\omega, \mathbb{P})$ is a probability space, then for any event $A \in \Omega$ it has probability $\mathbb{P}(A) = \displaystyle\sum_{\omega \in \mathcal{A}} \mathbb{P}(\omega)$

2.5 Uniform Probability Space

A uniform probability space is a pair $(\Omega, \mathbb{P})$ such that $\mathbb{P}(x) = \dfrac{1}{|\Omega|}$ for all $x \in \Omega$

If $(\omega, \mathbb{P})$ is a uniform probability space, then for any event $E \in \Omega$ it has probability $\mathbb{P}(E) = \dfrac{|E|}{|\Omega|}$

2.6 Axioms of Probability

1. Non-negativity: $\mathbb{P}(E) \geq 0$
2. Normalization: $\mathbb{P}(\Omega) = 1$
3. Countable Additivity: If E and F are mutually exclusive, then $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$

Corollaries of the axioms

1. Complementation: $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
2. Monotonicity: If $E \subseteq F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$
3. Inclusion-Exclusion: $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$

2.7 Conditional Probability

The conditional probability of event A given an event B occurred is $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

- We can rearrange the equation such that $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$
- If A and B are independent events, then $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A) \times \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$

2.8 Bayes' Theorem

The probability of an event A, based on prior knowledge of conditions related to the event is $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$

2.9 Partitions

Non-empty events $E_1, E_2, ..., E_n$ partition the sample space Ω if

$$E_1 \cup E_2 \cup ... \cup E_n = \bigcup_{i=1}^{n} E_i = \Omega$$

- The union of partitions cover the sample space
- The intersection of partitions is the null set

2.10 Law of Total Probability

If events $E_1, E_2, ..., E_n$ partition the sample space Ω, then for any event F

$$\mathbb{P}(F) = \mathbb{P}(F \cap E_1) + \mathbb{P}(F \cap E_2) + ... + \mathbb{P}(F \cap E_n) = \sum_{i=1}^{n} \mathbb{P}(F \cap E_i)$$

2.11 Chain Rule

$$\mathbb{P}(A_1 \cap ... \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2) \cdot ... \cdot \mathbb{P}(A_n \mid A_1 \cap ... \cap A_{n-1})$$

2.12 Independence

Two events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$

- If $\mathbb{P}(A) \neq 0$, then $\mathbb{P}(B \mid A) = \mathbb{P}(B)$
- If $\mathbb{P}(B) \neq 0$, then $\mathbb{P}(A \mid B) = \mathbb{P}(A)$
- Independent events with non-zero probabilities are never mutually exclusive

2.13 Conditional Independence

Two events A and B are independent conditioned on C if $\mathbb{P}(C) \neq 0$ and $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C) \cdot \mathbb{P}(B \mid C)$

- If $\mathbb{P}(A \cap C) \neq 0$, then $\mathbb{P}(B \mid A \cap C) = \mathbb{P}(B \mid C)$
- If $\mathbb{P}(B \cap C) \neq 0$, then $\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid C)$

3 Discrete Random Variables

3.1 Discrete Random Variables

A discrete random variable for a probability space $(\Omega, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$

- Discrete random variables partition the sample space
  - Every event must have a probability
  - Every event has exactly one probability

3.2 Probability Mass Function (PMF)

The probability mass function of a discrete random variable $X : \Omega \to \mathbb{R}$ specifies, for any real number x, the probability that $X = x$

$$\mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$$

where $\{\omega \in \Omega \mid X(\omega) = x\}$ is the event space

3.3 Cumulative Distribution Function (CDF)

The cumulative distribution function of a random variable $X : \Omega \to \mathbb{R}$ specifies, for any real number x, the probability that $X \leq x$

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

3.4 Converting a CDF to a PMF

Let F_X be piecewise constant. Then X is a discrete random variable and the possible values of X are the locations where F_X has jumps. If x is such a point, then $\mathbb{P}(X = x)$ equals the magnitude of the jump of F_X at x

3.5 Expectation

Given a discrete random variable $X : \Omega \to \mathbb{R}$, the expectation or expected value of X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega) \text{ or equivalently } \mathbb{E}[X] = \sum_{x \in \Omega_X} x \cdot \mathbb{P}(X = x)$$

3.6 Linearity of Expectation

For any two random variables X and Y

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

- Linearity of expectations applies for both independent and dependent variables

For any random variables $X_1, X_2, ..., X_n$ and real numbers $a_1, a_2, ..., a_n \in \mathbb{R}$

$$\mathbb{E}[a_1 X_1 + a_2 X_2 + ... + a_n X_n] = a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2] + ... + a_n \mathbb{E}[X_n]$$

3.7 Law of the Unconscious Statistician

Given a discrete real variable $X : \Omega \to \mathbb{R}$, the expectation or expected value of $Y = g(X)$ is

$$\mathbb{E}[Y] = \sum_{\omega \in \Omega} g(X(\omega)) \cdot \mathbb{P}(\omega)$$

or equivalently

$$\mathbb{E}[Y] = \sum_{x \in \Omega_X} g(x) \cdot \mathbb{P}(X = x)$$

or equivalently

$$\mathbb{E}[Y] = \sum_{y \in \Omega_y} y \cdot \mathbb{P}(Y = y)$$

3.8 Variance

The variance of a discrete real variable X is $\mathrm{Var}(X) = \sum_{x \in X} \mathbb{P}_X(x) \cdot (x - \mathbb{E}[X])^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

- $\mathrm{Var}(a \cdot X + b) = a^2 \cdot \mathrm{Var}(X)$
- $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- $\mathrm{Var}(X) = \mathrm{Var}(-X)$
- If X and Y are independent, then $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

3.9 Standard Deviation

The standard deviation of a random variable X is $\sigma(X) = \sqrt{\mathrm{Var}(X)}$

3.10 Median

The median of a random variable X is any $a \in \mathbb{R}$ that satisfies $\mathbb{P}(X \geq a) \geq \frac{1}{2}$ and $\mathbb{P}(X \leq a) \geq \frac{1}{2}$

3.11 p^{th} Quantile

The p^{th} quantile of a random variable X is any $a \in \mathbb{R}$ that satisfies $\mathbb{P}(X \geq a) \geq 1 - p$ and $\mathbb{P}(X \leq a) \geq p$

3.12 Independent Random Variables

Two random variables X, Y are mutually independent if for all x, y

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)$$

3.13 Discrete Uniform Random Variables

A discrete random variable X is equally likely to take any integer value between integers a and b inclusive, denoted $X \sim \mathrm{Unif}(a, b)$

- $\mathbb{P}(X = x) = \frac{1}{b-a+1}$
- $\mathbb{E}[X] = \frac{a+b}{2}$
- $\mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$

3.14 Bernoulli Random Variables

A Bernoulli random variable X takes value 1 with probability p, and value 0 with probability $1 - p$, denoted $X \sim \mathrm{Ber}(p)$

- $\mathbb{P}(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{otherwise} \end{cases}$
- $\mathbb{E}[X] = p$
- $\mathrm{Var}(X) = p(1 - p)$

3.15 Binomial Random Variables

A binomial random variable X is the number of successes in n independent random variables $Y_i \sim \mathrm{Ber}(p)$ where $X = \sum_{i=1}^{n} Y_i$, denoted $X \sim \mathrm{Bin}(n, p)$

- $\mathbb{P}(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$
- $\mathbb{E}[X] = np$
- $\mathrm{Var}(X) = np(1 - p)$

3.16 Geometric Random Variables

A geometric random variable X models the number of independent trials $Y_i \sim \mathrm{Ber}(p)$ before seeing the first success, denoted $X \sim \mathrm{Geo}(p)$

- $\mathbb{P}(X = x) = (1 - p)^{x-1} p$
- $\mathbb{E}[X] = \frac{1}{p}$
- $\mathrm{Var}(X) = \frac{1-p}{p^2}$

The geometric random variable $\frac{1}{n}\mathrm{Geo}\left(\frac{\lambda}{n}\right)$ well approximates the exponential distribution $\mathrm{Exp}(\lambda)$ when $\frac{\lambda}{n} < 1$ and n is very large

3.17 Negative Binomial Random Variables

A negative binomial random variable X models the number of independent trials $Y_i \sim \mathrm{Ber}(p)$ before seeing the r^{th} success. $X = \sum_{i=1}^{r} Z_i$ where $Z_i \sim \mathrm{Geo}(p)$, denoted $X \sim \mathrm{NegBin}(r, p)$

- $\mathbb{P}(X = x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$
- $\mathbb{E}[X] = \frac{r}{p}$
- $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$

3.18 Hypergeometric Random Variables

A hypergeometric random variable X measures the number of white balls you draw when you draw n balls uniformly at random from a total of N of which K are white and the rest are black, denoted $X \sim \mathrm{HypGeo}(N, K, n)$

- $\mathbb{P}(X = x) = \dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$
- $\mathbb{E}[X] = n\frac{K}{N}$
- $\mathrm{Var}(X) = n\frac{K(N-K)(N-n)}{N^2(N-1)}$

3.19 Poisson Random Variables

A Poisson random variable X is the actual number of events happening per unit time given events happen independently at an average rate of λ per unit time, denoted $X \sim \mathrm{Poi}(\lambda)$

- $\mathbb{P}(X = x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}$
- $\mathbb{E}[X] = \lambda$
- $\mathrm{Var}(X) = \lambda$

The Poisson random variable $\mathrm{Poi}(np)$ well approximates the binomial random variable $\mathrm{Bin}(n, p)$ when np^2 is small

3.20 Sum of Independent Poisson Random Variables

Let $X \sim \mathrm{Poi}(\lambda_1)$ and $Y \sim \mathrm{Poi}(\lambda_2)$ such that $\lambda = \lambda_1 + \lambda_2$. Let $Z = X + Y$

- $\mathbb{P}(Z = z) = \dfrac{e^{-(\lambda_1+\lambda_2)}}{z!}(\lambda_1 + \lambda_2)^z$
- $\mathbb{E}[Z] = \lambda_1 + \lambda_2$

3.21 Poisson Process

A Poisson process describes the actual number of events happening in the unit interval $[a, b]$ given events happen independently at an average rate of λ per unit time, denoted $N([a, b]) \sim \mathrm{Poi}(\lambda(b - a))$

- The average rate λ per unit time is called the intensity of the process
- If $I_1, I_2, ..., I_n$ are non-overlapping intervals in $[0, \infty)$, then the Poisson processes $N(I_1), N(I_2), ..., N(I_n)$ are mutually independent

4 Continuous Random Variables

4.1 Probability Density Function (PDF)

A probability density function $f_X : \mathbb{R} \to \mathbb{R}$ represents a continuous random variable X

- $f_X(x) \geq 0$ for all $x \in \mathbb{R}$
- $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$
- $f_X(x)$ may be greater than 1

4.2 Cumulative Distribution Function (CDF)

The cumulative distribution of a continuous random variable X specifies, for any real number x, the probability that $X \leq x$

$$F_X(a) = \mathbb{P}(X \leq a) = \int_{-\infty}^{a} f_X(x)\, dx$$

- The probability $\mathbb{P}(X < a)$ is obtained by taking the left-hand limit $\lim\limits_{x \to a^-} F_X(a)$

4.3 Converting a CDF to a PDF

Let F_X be continuous and its derivative F'_X exist everywhere on the real line, except possibly at finitely many points. Then X is a continuous random variable and $f(x) = F'_X(x)$ is the probability density function of x. If F_X is not differentiable at x, then the value $f(x)$ can be set arbitrarily

4.4 Expectation

Given a continuous random variable X, the expectation or expected value of X is

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} f_X(x) \cdot x\, dx$$

4.5 Variance

The variance of a continuous random variable X is

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot (x - \mathbb{E}[X])^2\, dx = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

4.6 Continuous Uniform Random Variables

A continuous uniform random variable X is denoted $X \sim \mathrm{Unif}(a, b)$

- $f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$
- $\mathbb{E}[X] = \frac{b+a}{2}$
- $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$

4.7 Exponential Distribution

An exponential random variable X models the waiting time before the next event occurs given that λ events occur per unit time, denoted $X \sim \mathrm{Exp}(\lambda)$

- $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$
- $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$
- $\mathbb{E}[X] = \frac{1}{\lambda}$
- $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

The exponential distribution $n\mathrm{Exp}(np)$ well approximates the geometric random variable $\mathrm{Geo}(p)$ when p is very small

4.8 Memoryless Random Variables

A random variable is memoryless if for all $s, t > 0$, $\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$

- $X \sim \mathrm{Exp}(\lambda)$ is a memoryless random variable

4.9 Normal Distribution

A normal random variable X with parameters $\mu \in \mathbb{R}$ and $\sigma \geq 0$ is denoted $X \sim \mathcal{N}(\mu, \sigma^2)$

- $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $\mathbb{E}[X] = \mu$
- $\mathrm{Var}(X) = \sigma^2$

Properties of the normal distribution

- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
- If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ where X and Y are independent, then $aX + bY + c \sim \mathcal{N}(a\mu_x + b\mu_y + c,\ a^2\sigma_X^2 + b^2\sigma_Y^2)$

The normal distribution $\mathcal{N}(np, \sqrt{np(1-p)}^{\,2})$ well approximates the binomial random variable $\mathrm{Bin}(n, p)$ when $np(1-p) > 10$

4.10 Standard Unit Normal Distribution

The standard unit normal distribution Z is a normal random variable with parameters $\mu = 0$ and $\sigma^2 = 1$, denoted $Z \sim \mathcal{N}(0, 1)$

- $\mathbb{P}(Z \leq z) = \mathbb{P}(-z \leq Z) = \Phi(z)$
- $\mathbb{P}(z \leq Z) = \mathbb{P}(Z \leq -z) = 1 - \Phi(z)$

4.11 Standardizing Normal Distributions

Given a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, the CDF of X is given by

$$\mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

where $Z = \frac{X-\mu}{\sigma}$

4.12 Central Limit Theorem

Let $S_n = X_1 + ... + X_n$, where $X_1, ..., X_n$ are independent and identically distributed (iid) random variables each with expectation μ and variance σ^2

- $\mathbb{E}[S_n] = n\mu$
- $\mathrm{Var}(S_n) = n\sigma^2$

The CDF of $Y_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ converges to the CDF of the standard unit normal $\mathcal{N}(0, 1)$

- $\mathbb{E}[Y_n] = 0$
- $\mathrm{Var}(Y_n) = 1$

Alternately, the CDF of $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ converges to the CDF of normal variable $\mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$

- $\mathbb{E}\left[\bar{X}\right] = \mu$
- $\mathrm{Var}\left(\bar{X}\right) = \dfrac{\sigma^2}{n}$

4.13 Continuity Correction

To estimate the probability that a discrete random variable lands in the integer interval $[a, b]$, compute the probability that the continuous approximation lands in the interval $\left[a - \frac{1}{2}, b + \frac{1}{2}\right]$

4.14 Stirling's Formula

If $n \to \infty$, then $n! \sim n^n e^{-n}\sqrt{2\pi n}$

4.15 Minimum of IID Random Variables

If $Y_1, ..., Y_m$ are iid continuous uniform random variables $\mathrm{Unif}(0, 1)$, then $\mathbb{E}[\min(Y_1, ..., Y_m)] = \frac{1}{m+1}$

- Let $\mathrm{val} = \min(Y_1, ..., Y_m)$. Then $m = \frac{1}{\mathbb{E}[\mathrm{val}]} - 1$

4.16 Discrete Counting

Suppose we have an unknown number of iid random variables $Y_1, ..., Y_m$ and k independent hash functions $h_i : U \to [0, 1]$. Let $\mathrm{val}_i = \min(h_i(Y_1), ..., h_i(Y_m))$. Then

$$\mathbb{E}[\mathrm{val}] \approx \frac{1}{k}\sum_{i=1}^{k} \mathrm{val}_i \text{ such that } m \approx \frac{1}{\frac{1}{k}\sum_{i=1}^{k} \mathrm{val}_i} - 1$$

5 Transforms and Transformations

5.1 Moment Generating Functions

The moment generating function of a random variable X is given by $M(t) = \mathbb{E}\left[e^{tX}\right]$ where $t \in \mathbb{R}$

- If X is discrete, then $M(t) = \sum_x e^{tx} \cdot \mathbb{P}(X = x)$
- If X is continuous, then $M(t) = \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)\, dx$

5.2 Moments

If the moment generating function of X is finite around the origin, then the moment $\mathbb{E}[X^n]$ is given by the n^{th} derivative of the moment generating function $M(t)$ at $t = 0$

$$\mathbb{E}\left[X^n\right] = M^{(n)}(0)$$

5.3 Equality in Distribution

If two random variables X and Y have the same probability distribution, then $X \stackrel{d}{=} Y$

- Discrete X and Y are equal in distribution if and only if $\mathbb{P}(X = \alpha) = \mathbb{P}(Y = \alpha)$ for all $\alpha \in \Omega$
- Continuous X and Y are equal in distribution if and only if $f_X(\alpha) = f_Y(\alpha)$ for all $\alpha \in \mathbb{R}$

5.4 Moment Generating Functions and Equality in Distribution

Let X and Y be random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then X and Y are equal in distribution if and only if there exists $\delta > 0$ such that $M_X(t), M_Y(t)$ are finite and $M_X(t) = M_Y(t)$ for all $t \in (-\delta, \delta)$