STA347
Final Preparation

Yuchen Wang

March 26, 2020

Contents

CS GRAD SCHOOL
ADMISSION CONSULTING

Do you want offers from top CS grad schools?

My team offers

- Personal statements editing
- Resume editing
- Recommendation letter strategies
- Grad school/program list suggestion
- Application timeline planning
- 1:1 mentorship with top school Masters' and PhD graduates

Yuchen Wang

MSCS@Stanford, SWE@Microsoft

www.yuchenwyc.com

Stanford University | University of Toronto | Microsoft

GET STARTED

yuchenw@stanford.edu

1 Experiments, Events and Sample Spaces

Definition 1.1. Experiment, Sample space and event

- Experiment: Any process, real or hypothetical, in which the possible outcomes can be identified ahead of time;
- Sample space: The collection of all possible outcomes, denoted by S;
- Event: A well-defined subset of sample space

Definition 1.2 (countably infinity)**.** A set is **countably infinite** if its elements can be put in one-to-one correspondence with the set of natural numbers.

Definition 1.3 (At most countable sets)**.** A set that is either finite or countably infinite is called an **at most countable set**.

Theorem 1.1. Suppose $E, E_1, E_2, \ldots$ are events. The following are also events

1. E^c
2. $E_1 \cup E_2 \cup \ldots E_n$
3. $\sum_{i=1}^{\infty} E_i$

2 Definition and Properties of Probability

Definition 2.1 (σ-field)**.** Let χ be a space. A collection $\mathcal{F}$ of subsets of χ is called a **σ-field** if

1. $\chi \in \mathcal{F}$
2. (closure under complement) if $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$
3. (closure under countable union) if $E_1, E_2, \ldots \in \mathcal{F}$, then $\cup_{n=1}^{\infty} E_n \in \mathcal{F}$

Remark 2.1. A σ-field refers to the collection of subsets of a sample space that we should use in order to establish a mathematically formal definition of probability. The sets in the σ-field constitute the events from our sample space.

Axiom 2.1 (Axioms of Probability)**.** Let S be a sample space, and let $\mathcal{F}$ be a σ-field of S.

- Axiom 1 (non-negativity) $P(E) \geq 0$ for any event $E \in \mathcal{F}$.
- Axiom 2 $\quad P(S) = 1$
- Axiom 3 (countable additivity) For every sequence of disjoint events $E_1, E_2, \ldots \in \mathcal{F}$

$$P\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P\left(E_i\right)$$

Definition 2.2 (probability)**.** Any function P on a sample space S satisfying Axioms 1-3 is called a **probability**.

Definition 2.3 (disjoint sets)**.** Sets A and B are **disjoint** if $A \cap B = \emptyset$.

Theorem 2.1. Properties of Probability

1. $P(\emptyset) = 0$

2. (finite additivity) For any disjoint events $E_1, \ldots, E_n$,

$$P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i)$$

3. $P(A^c) = 1 - P(A)$
4. For $A \subset B, P(A) \leq P(B)$
5. $0 \leq P(A) \leq 1$
6. $P(A - B) = P(A) - P(A \cap B)$
7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
8. (subadditivity, Boole's inequality) For any events $E_1, \ldots, E_n$,

$$P(\cup_{i=1}^n E_i) \leq \sum_{i=1}^n P(E_i)$$

Theorem 2.2 (Continuity from below and above)**.** Let P be a probability.
(continuity from below) If $A_n \nearrow A$ (i.e. $A_1 \subset A_2 \subset \ldots$ and $\cup_n A_n = A$), then $P(A_n) \nearrow P(A)$
(continuity from above) If $A_n \searrow A$ (i.e. $A_1 \supset A_2 \supset \ldots$ and $\cap_n A_n = A$), then $P(A_n) \searrow P(A)$

2.1 Finite Sample Spaces

Suppose $|S| = n$, that is, $S = \{s_1, \ldots, s_n\}$. Then each member has probability, that is, $p_i = P(\{s_i\})$ such that

$$p_i \geq 0 \text{ and } \sum_{i=1}^n p_i = 1$$

3 Classical Equal Probability and Combinatorics

Definition 3.1 (permutation)**.** When there are n elements, the number of events pulling k elements out of n elements is called a **permutation** of n elements taken k at a time and denoted by $P_{n,k}$.

Theorem 3.1.

$$P_{n,k} = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$$

Definition 3.2 (combination)**.** The number of combinations of n elements taken k at a time is denoted by $C_{n,k}$ or $\binom{n}{k}$.

Theorem 3.2.

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!} = P_{n,k}/k!$$

Theorem 3.3 (Binomial coefficients)**.**

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

Theorem 3.4 (Newton Expansion)**.** For $|z| < 1$, the term $(1+z)^r$ can be expanded as

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

Theorem 3.5.

$$\binom{n}{k} = \frac{r(r-1)\dots(r-k+1)}{k!} = \frac{\Gamma(r+1)}{\Gamma(r-k+1)\Gamma(k+1)}$$

with $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$

Theorem 3.6. For any numbers $x_1, \dots, x_k$ and non-negative integer n,

$$(x_1 + \dots + x_k)^n = \sum \binom{n}{n_1, \dots, n_k} x_1^{n_1} \dots x_k^{n_k}$$

It is easy to see that

$$\begin{aligned} \binom{n}{n_1, \dots, n_k} &= \binom{n}{n_1}\binom{n_2 + \cdots + n_k}{n_2}\binom{n_3 + \cdots + n_k}{n_3} \cdots \binom{n_k}{n_k} \\ &= \frac{n!}{n_1! \cdots n_k!} \end{aligned} \tag{1}$$

Theorem 3.7 (Stirling's formula)**.**

$$\lim_{n\to\infty} \left| \log(n!) - [\frac{1}{2}\log(2\pi) + (n + \frac{1}{2})\log(n) - n] \right| = 0$$

4 Inclusion-Exclusion Formula

For any n events $A_1, \dots, A_n$,

$$\begin{aligned} P\left(\cup_{i=1}^n A_i\right) = &\sum_{i=1}^n P\left(A_i\right) - \sum_{i<j} P\left(A_i \cap A_j\right) + \sum_{i<j<k} P\left(A_i \cap A_j \cap A_k\right) + \cdots \\ &+ (-1)^{n-1} P\left(A_1 \cap \cdots \cap A_n\right) \end{aligned} \tag{2}$$

5 Conditional Probability

Definition 5.1 (conditional probability)**.** When $P(B) > 0$, the **conditional probability** of an event A given B is defined by

$$P(A|B) = P(A \cap B)/P(B)$$

Theorem 5.1. If $P(B) > 0$, then $P(A \cap B) = P(A|B)P(B)$.

Theorem 5.2. Let $A_1, \dots, A_n$ be events with $P(A_1 \cap \dots \cap A_n) > 0$. Then

$$P\left(A_1 \cap \cdots \cap A_n\right) = P\left(A_1\right) P\left(A_2|A_1\right) P\left(A_3|A_1, A_2\right) \cdots P\left(A_n|A_1, \dots, A_{n-1}\right) \tag{3}$$

6 Independence

Definition 6.1 (independence)**.** Two events A and B are **independent** if and only if

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

. A collection of events $\{A_i\}_{i\in I}$ are said to be **(mutually) independent** if

$$P(\cap_{i\in J} A_i) = \Pi_{i\in J} P(A_i)$$

for any $\emptyset \neq J \subset I$.
A collection of events $\{A_i\}_{i\in I}$ are said to be **pair-wise independent** if

$$P(A_i \cap A_j) = P(A_i)P(A_j)$$

for $i \neq j \in I$.

Theorem 6.1. Two events A and B are independent if and only if A and B^c are independent.

Definition 6.2 (conditionally independence)**.** Two events A and B are **conditionally independent** given C if

$$P(A \cap B|C) = P(A|C)P(B|C)$$

Remark 6.1. Conditional independence does not imply independence.

7 Bayes Theorem

Definition 7.1. A collection of sets $B_1, \ldots, B_k$ is called a **partition** of A if and only if $B_1, \ldots, B_k$ are disjoint and $A = \cup_{i=1}^{k} B_i$.

Theorem 7.1 (Law of total probability)**.** Let events $B_1, \ldots, B_k$ be a partition of S with $P(B_j) > 0$ for all $j = 1, \ldots, k$. For any event A,

$$P(A) = \sum_{j=1}^{k} P(B_j)P(A|B_j)$$

Theorem 7.2 (Bayes' Theorem)**.** If $0 < P(A), P(B) < 1$, then

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

8 Random Variables

Definition 8.1. A real-valued function X on the sample space S is called a **random variable** if the probability of X is well-defined, that is, $\{s \in S : X(s) \leq r\}$ is an event for each $r \in \mathbb{R}$.

Definition 8.2 (Borel sets in $\mathbb{R}$)**.** The collection of all Borel sets $\mathcal{B}$ in $\mathbb{R}$ is the smallest collection satisfying the followings

1. $(a, b] \in \mathcal{B}$ for any $a < b \in \mathbb{R}$
2. (closure under complement) For any $B \in \mathcal{B}, B^c \in \mathcal{B}$
3. (closure under countable union) For any $B_1, B_2, \ldots \in \mathcal{B}, \cup_{j=1}^{\infty} B_j \in \mathcal{B}$

We call the collection $\mathcal{B}$ the **Borel σ-field**

Definition 8.3 (Probability of a random variable)**.** For any Borel set B in $\mathbb{R}$, an event $X \in B$ is defined as $\{s \in S : X(s) \in B\}$ and often denoted by $\{X \in B\}$ or $(X \in B)$. The corresponding probability is

$$P(X \in B) = P(\{s \in S : X(s) \in B\})$$

Lemma 8.1. If $|X(S)| < \infty$ and $(X = r)$ is an event for any $r \in X(S)$, then X is a random variable.

Definition 8.4 (distribution)**.** The **distribution** of X is the collection of all probabilities of all events induced by X, that is, $(B, P(X \in B))$. Two random variables X and Y are said to be **identically distributed** if they have the same distribution.

Remark 8.1. To show X and Y having the same distribution, we need to check for any event B on $\mathbb{R}$, $P(X \in B) = P(Y \in B)$. Since all Borel sets on $\mathbb{R}$ are induced by intervals, it is enough to prove

$$P(a < X \leq b) = P(a < Y \leq b)$$

for any $a < b \in \mathbb{R}$. Even $P(X \leq a) = P(Y \leq a)$ for any $a \in \mathbb{R}$ guarantees that X and Y are identically distributed.

Definition 8.5 (discrete random variable)**.** A random variable X is said to be **discrete** if $P(X = x) = 0$ or $P(X = x) > 0$ and $P(X \in \chi_0) = 1$ where $\chi_0 = \{x \in \mathbb{R} : P(X = x) > 0\}$

Definition 8.6 (probability mass function)**.** The **probability mass function** (pmf) of a discrete random variable X is

$$pmf_X(x) = P(X = x)$$

for any possible value of $x \in X(S)$.

Theorem 8.1. Let X be a discrete random variable. Then the set of x having $P(X = x)$ is at most countable.

Theorem 8.2. Let f be the pmf of a discrete random variable X. The set of possible values of X is $X(S) = \{x_1, x_2, \ldots\}$. For $x \notin X(S) \geq 0$ and $\sum_{i=1}^{\infty} f(x_i) = 1$.

Theorem 8.3. Let $X(S) = \{x_1, x_2, \ldots\}$ be the set of possible values of a discrete random variable X. Then for any subset A of $\mathbb{R}$,

$$P(X \in A) = \sum_{x \in A} P(\{x\}) = \sum_{x \in A} pmf_X(x)$$

Definition 8.7 (absolutely continuity and probability density function)**.** A random variable X is said to be **absolutely continuous** if the probability of each interval $[a, b]$ is of the form

$$P(a < X \leq b) = \int_a^b f(x)\, dx$$

where $a < b \in \mathbb{R}$ and f is a non-negative function on $\mathbb{R}$. Such function f is called a **probability density function** (pdf) of X.

Theorem 8.4. Let X be a continuous random variable. Then

$$pdf_X(x) = \frac{d}{dx} P(X \leq x)$$

8.1 Examples of Random Variables

Definition 8.8 (Bernoulli)**.** A random variable X taking value 0 or 1 with $P(X = 1) = p$ and $P(X = 0) = 1-p$ for some $p \in [0, 1]$ is called a **Bernoulli** random variable with success probability p and often denoted by $X \sim$ Bernoulli(p).

Definition 8.9 (discrete uniform)**.** Let χ be a non-empty finite set. A random variable X taking values in χ with equal probability is called a uniform random variable on χ and denoted by $X \sim uniform(\chi)$.
The probability mass function of $X \sim uniform(\chi)$ is

$$pmf_X(x) = \begin{cases} \frac{1}{|\chi|} & \text{if } x \in \chi \\ 0 & \text{otherwise} \end{cases}$$

Definition 8.10 (binomial)**.** A random variable X is called a **binomial** random variable if it has the same distribution as Z which is the number of success in n independent trails with success probability p, and denoted by $X \sim$ binomial(n, p).
The probability mass function of $X \sim$ binomial(n, p) is

$$pmf_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } n = 0, 1, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Definition 8.11 (continuous uniform)**.** A random variable X defined on (a, b) for finite real numbers $a < b$ satisfying $P(c < X \leq d) = \frac{d-c}{b-a}$ for any c, d such that $a \leq c \leq d \leq b$ is called a **uniform** random variable on (a, b) which is denoted by $X \sim$ uniform(a, b). The probability mass function of $X \sim uniform(a, b)$ is

$$pmf_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

Definition 8.12 (geometric)**.** Consider an independent Bernoulli trial with success probability p. The number of trials until the first success is called a **geometric** distribution with parameter p, denoted by geometric(p). The geometric random variable $X \sim geometric(p)$ has probability mass function as

$$pmf_X(n) = (1-p)^{n-1}p$$

for $n \in \mathbb{N}$.

Definition 8.13 (negative binomial)**.** Consider an independent Bernoulli trial with success probability p. The number of trials until k-th success is called a **negative binomial** distribution with parameter k and p, denoted by neg-bin(k, p).
The negative binomial random variable $X \sim neg - bin(k, p)$ has probability mass function as

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

for $n \in \mathbb{N}$ s.t. $n \geq k$.

Definition 8.14 (hypergeometric)**.** Consider a jar containing n balls of which r are black and the remainder $n-r$ are white. The random variable X is the number of black balls when m balls are drawn without replacement. The probability of k black balls are drawn is

$$\text{pmf}_X(k) = \begin{cases} \left(\begin{array}{c} n-r \\ m-k \end{array}\right) / \left(\begin{array}{c} n \\ m \end{array}\right) & \text{if } k = 0, \dots, \min(r, m) \\ 0 & \text{otherwise.} \end{cases}$$

Such distribution is called a **hypergeometric** distribution.

Definition 8.15 (zeta/zipf)**.** A positive integer valued random variable X follows a **Zeta** or **Zipf** distribution if

$$\text{pmf}_X(n) = \frac{n^{-s}}{\zeta(s)}$$

for $n = 1, 2, \dots$ and $s > 1$ where $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$

Definition 8.16 (Poisson)**.** A **Poisson** distribution with parameter $\mu > 0$ has the probability mass function

$$pmf_X(n) = e^{-\mu}\frac{\mu^n}{n!}$$

for non-negative integer n.

Theorem 8.5. If $X \sim Poisson(\lambda)$ and the distribution of Y, conditional on $X = k$, is a binomial distribution, $Y|(X = k) \sim Binom(k, p)$, then the distribution of Y follows a Poisson distribution $Y \sim Poisson(\lambda \cdot p)$

Theorem 8.6 (Sums of Poisson-distributed random variables)**.** If $X_i \sim Poisson(\lambda_i)$ for $i = 1, \dots, n$ are independent, and $\lambda = \sum_{i=1}^{n} \lambda_i$, then $Y = (\sum_{i=1}^{n} X_i) \sim Poisson(\lambda)$.

Definition 8.17 (Exponential)**.** A continuous random variable W having the probability density

$$\text{pdf}_W(w) = \lambda e^{-\lambda w} 1(w > 0)$$

is distributed from an exponential distribution with parameter $\lambda > 0$, which is denoted by $W \sim$ exponential (λ).

8.2 Cumulative Distribution Function

The **(cumulative) distribution function** of a random variable X is the function

$$\mathrm{cdf}_X(x) = F_X(x) = P(X \leq x)$$

for $-\infty < x < \infty$.

Theorem 8.7 (properties of distribution functions)**.** Let F be a distribution function. Then
(a) F is nondecreasing,
(b) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$,
(c) F is right continuous, that is, $\lim_{y\searrow x} F(y) = F(x)$,
(d) $F(x-) := \lim_{y\nearrow x} F(y) = P(X < x)$
(e) $P(X = x) = F(x) - F(x-)$

Theorem 8.8. If a real function F satisfies (a)-(c) in the above properties, then it is a distribution function of a random variable.

Definition 8.18 (p-quantile)**.** The p-quantile of a random variable X is x such that $P(X \leq x) \geq p$ and $P(X \geq x) \geq 1 - p$.

Definition 8.19. The median, lower quartile, upper quartile are 0.5-, 0.25-, 0.75-quantile. The inter quartile range (IQR) is the difference between upper and lower quartile.

8.3 Multivariate Distributions

8.3.1 Bivariage Distributions

Definition 8.20. The **joint/bivariate distribution** of two random variables X and Y is the collection of all possible probabilities, that is, $P((X,Y) \in B)$ where B is a Borel set in $\mathbb{R}^2$.

Definition 8.21. Two random variables X and Y are jointly continuously distributed if and only if there exists a non-negative function f such that for any Borel set B in $\mathbb{R}^2$

$$P((X,Y) \in B) = \iint_B f(x,y)\,dx\,dy$$

Such function f is called a **joint density function** of (X,Y).

Theorem 8.9 (Properties of joint density functions)**.** Joint density functions satisfies

1. $$pdf_{X,Y}(x,y) \geq 0$$

2. $$\iint pdf_{X,Y}(x,y)\,dx\,dy = 1$$

Definition 8.22. The **joint (cumulative) distribution function** of X and Y is

$$cdf_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

Definition 8.23. When X and Y are discrete, then the **joint probability mass function** of X and Y is defined by

$$pmf_{X,Y}(x,y) = P(X = x, Y = y)$$

Theorem 8.10 (Properties of joint probability mass functions)**.** Satisfies

1. $$pmf_{X,Y}(x,y) \geq 0$$

2. $$\sum_{x,y} pmf_{X,Y}(x,y) = 1$$

Theorem 8.11. Consider two random variables X and Y.

$$\begin{aligned}
\lim_{y \to -\infty} cdf_{X,Y}(x,y) &= 0 \\
\lim_{x \to -\infty} cdf_{X,Y}(x,y) &= 0 \\
\lim_{y \to \infty} cdf_{X,Y}(x,y) &= cdf_X(x) \\
\lim_{x \to \infty} cdf_{X,Y}(x,y) &= cdf_Y(y)
\end{aligned}$$

8.3.2 Marginal Distributions

Suppose X and Y are random variables. The cdf or pmf or pdf of X (or Y) derived from the joint cdf or pmf or pdf is called the **marginal** cdf or pmf or pdf of X (or Y).

Theorem 8.12. 1. $$pmf_X(x) = \sum_y pmf_{X,Y}(x,y)$$

2. $$pdf_X(x) = \int pdf_{X,Y}(x,y)\, dy$$

Definition 8.24. Two random variables X and Y are **independent** if and only if

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

Theorem 8.13. If two random variables X and Y are independent, then the following hold if the functions exist.

1. $cdf_{X,Y}(x,y) = cdf_X(x) \times cdf_Y(y)$ for all x, y
2. $pmf_{X,Y}(x,y) = pmf_X(x) \times pmf_Y(y)$ for all x, y
3. $pdf_{X,Y}(x,y) = pdf_X(x) \times pdf_Y(y)$ for all x, y

Theorem 8.14. If one of the following hold, then two random variables X and Y are independent.

1. $cdf_{X,Y}(x,y) = cdf_X(x) \times cdf_Y(y)$ for all x, y
2. $pmf_{X,Y}(x,y) = pmf_X(x) \times pmf_Y(y)$ for all x, y
3. $pdf_{X,Y}(x,y) = pdf_X(x) \times pdf_Y(y)$ for all x, y

8.3.3 Conditional Distributions

Definition 8.25. The conditional density of X given $Y = y$ is

$$pdf_{X|Y}(x|y) = \frac{pdf_{X,Y}(x,y)}{pdf_Y(y)}$$

Theorem 8.15.

$$pdf_{X,Y}(x,y) = pdf_X(x)pdf_{X|Y}(x|y)$$

8.3.4 Multivariate Distributions

Definition 8.26. The joint cumulative distribution function of n variables $X_1, \dots, X_n$ is defined by

$$cdf_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 \le x_1, \dots, X_n \le x_n)$$

The joint probability mass/density function of n discrete/continuous random variables $X_1, \dots, X_n$ is

$$pmf_{X_1,\dots,X_n}(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

$$P((X_1, \dots, X_n) \in B) = \int \underset{B}{\cdots} \int pdf_{X_1,\dots,X_n}(x_1, \dots, x_n)\, dx_n \dots dx_1$$

Definition 8.27. Let $X_1, \dots, X_n$ be random variables. Marginal cumulative distribution, probability mass, probability density functions of $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$ are

$$cdf_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)} = \lim_{x_i \to \infty} cdf_{X_1,\dots,X_{i-1},X_i,X_{i+1},\dots,X_n}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \quad (4)$$

$$pmf_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)} = \sum_{x_i} pmf_{X_1,\dots,X_{i-1},X_i,X_{i+1},\dots,X_n}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \quad (5)$$

$$pdf_{X_1,\dots,X_{i-1},X_{i+1},\dots,X_n(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)} = \int pdf_{X_1,\dots,X_{i-1},X_i,X_{i+1},\dots,X_n}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)\, dx_i \quad (6)$$

Theorem 8.16. Let $X_1, \dots, X_n$ be continuous random variables having cdf. Then

$$pdf_{X_1,\dots,X_n}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, \dots, x_n)$$

Definition 8.28. Random variables $X_1, \dots, X_n$ are **independent** if and only if for any Borel sets $B_1, \dots, B_n$

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n)$$

Theorem 8.17. Random variables $X_1, \dots, X_n$ are **independent** if and only if

$$cdf_{X_1,\dots,X_n}(x_1, \dots, x_n) = cdf_{X_1}(x_1) \dots cdf_{X_n}(x_n)$$

8.4 Functions of Random Variables

Theorem 8.18. Let X be a discrete random variable and $Y = g(X)$ be a transformed random variable where $g : \mathbb{R} \to \mathbb{R}$ is a function. The pmf of Y is

$$pmf_Y(y) = \sum_{x:g(x)=y} pmf_X(x)$$

Theorem 8.19. Let X be a continuous random variable and $Y = g(X)$ be a transformed random variable where g is an appropriate transformation like continuous increasing. The cdf of Y is

$$cdf_Y(y) = \int_{\{x:g(x)\le y\}} pdf_X(x)\, dx$$

The probability density function of Y is

$$pdf_Y(y) = \frac{d}{dy} cdf_Y(y)$$

Theorem 8.20. Let X be a continuous random variable and $F(x) = cdf_X(x)$. Then new random variable $Y = F(X)$ is uniformly distributed on $(0, 1)$, that is, $Y \sim uniform(0, 1)$.

Theorem 8.21 (change of variable)**.** Let X be a continuous random variable and g be a one-to-one and differentiable function. Then the density of random variable $Y = g(X)$ is

$$pdf_Y(y) = pdf_X(g^{-1}(y))\Big|\frac{d}{dy}g^{-1}(y)\Big|$$

whenever y is in the range of $Y(S)$.

Theorem 8.22. Consider discrete random variables $X_1, \ldots, X_n$. There exist m functions $g_1, \ldots, g_m$ so that $Y_i = g_i(X_1, \ldots, X_n)$. The joint probability mass function of $Y = (Y_1, \ldots, Y_m)$ is

$$pmf_Y(y) = \sum_{x:g_i(x)=y_i, i=1,\ldots,m} pmf_X(x)$$

Definition 8.29. Random variables $X_1, \ldots, X_n$ are said to be **independent** and **identically distributed (i.i.d)** if all random variables have the same distribution and are independent.

Theorem 8.23. Let X and Y be jointly continuous random variables. The density of $Z = X + Y$ is

$$pdf_Z(z) = \int pdf_{X,Y}(x, z - x)\, dx$$

If X and Y are independent, then

$$pdf_X(z) = \int pdf_X(x) pdf_Y(z - x)\, dx$$

Theorem 8.24 (change of variable)**.** Suppose $X_1, \ldots, X_n$ have a joint density function $f(x_1, \ldots, x_n)$ and $Y_i = g_i(X_1, \ldots, X_n)$ for one-to-one correspondent and differentiable functions g_i's, say $y = g(x)$. The joint density of $Y_1, \ldots, Y_n$ is

$$pdf_Y(y) = pdf_X(x)\Big|\det\left(\frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)}\right)\Big|$$

where $x = (x_1, \ldots, x_n) = g^{-1}(y)$

8.5 Expectation

Definition 8.30. expectation The **expectation** (or expected value or mean value) of a discrete random variable is

$$\mathbb{E}[X] = \sum_x x \times P(X = x) = \sum_x x \times pmf_X(x)$$

when the sum is absolutely convergent.

Definition 8.31. The expectation of a continuous random variable X is defined by

$$\mathbb{E}[X] = \int x \times pdf_X(x)\, dx$$

Theorem 8.25. Assume a discrete random variable X is non-negative. Then

$$\mathbb{E}[X] = \int_0^\infty P(X > z)\, dz = \int_0^\infty x\, dF(x)$$

Corollary 8.1. Let X be a non-negative integer valued random variables. Then

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} P(X \geq n)$$

Lemma 8.2. Let F be the cumulative distribution function of a random variable X. For an interval,

$$P(a < X \leq b) = \mathbb{E}[1(a < X \leq b)]$$

In general, for each event A of X,

$$P(X \in A) = \mathbb{E}[1(X \in A)]$$

Theorem 8.26. For any random variable X with finite expectation,

$$\mathbb{E}[X] = \int_0^\infty P(X > z)\, dz - \int_{-\infty}^0 P(X < z)\, dz = \int_{-\infty}^\infty x\, dF(x)$$

Theorem 8.27. Let X be a random variable and g be a function on $\mathbb{R}$. If expectation of $Y = g(X)$ is defined, then

$$\mathbb{E}[Y] = \int g(x)\, d\, cdf_X(x) = \int_{-\infty}^\infty g(x) \cdot pdf_X(x)\, dx$$

or

$$\mathbb{E}[Y] = \int g(x)\, d\, cdf_X(x) = \sum_x g(x) \cdot pdf_X(x)$$

Lemma 8.3. Assume $X, Y \geq 0$ with probability 1, that is, $P(X \geq 0, Y \geq 0) = 1$, then

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

and

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

Theorem 8.28 (Properties of Expectation)**.** Satisfies

1. (linearity) Let $Y = aX + b$, then
$$\mathbb{E}[Y] = a\mathbb{E}[X] + b$$
2. (monotonicity) If $X \geq 0$, that is, $P(X \geq 0) = 1$, then $E(X) \geq 0$
3. (additivity) $\mathbb{E}[(]X + Y) = \mathbb{E}[X] + \mathbb{E}[Y]$
4. For constant random variable 1, $\mathbb{E}[1] = 1$

Theorem 8.29. Let X and Y be two independent random variables and g and h be real functions satisfying $g(X)$ and $h(Y)$ are random variables with finite expectations. Then

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

8.6 Moments

Definition 8.32. For positive integer k, the **k-th moment** of X is $\mathbb{E}[X^k]$ and the **k-th central moment** is $\mathbb{E}[(X - \mathbb{E}[X])^k]$.

Theorem 8.30. If $\mathbb{E}[|X|^t] < \infty$ for some $t > 0$, then $\mathbb{E}[|X|^s] < \infty$ for any $0 \leq s \leq t$.

Definition 8.33 (variance)**.** The **variance** of a random variable X is

$$\mathrm{VAR}\, X = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

The **covariance** and **correlation** between two random variables X and Y are

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

and

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{\mathrm{VAR}\, X\, \mathrm{VAR}\, Y}}$$

Theorem 8.31 (Properties of variance)**.** satisfies

1. $\mathrm{VAR}\, X \geq 0$
2. $\mathrm{VAR}\, X = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
3. $\mathrm{VAR}\, aX + b = a^2\, \mathrm{VAR}\, X$
4. $\mathrm{VAR}\, X + Y = \mathrm{VAR}\, X + \mathrm{VAR}\, Y + 2Cov(X, Y)$
5. $\mathrm{VAR}\, X + Y = \mathrm{VAR}\, X + \mathrm{VAR}\, Y$ if and only if X and Y are uncorrelated.
6. If a random variable X is bounded, then it must has finite variance.
7. $\mathrm{VAR}\, X = 0$ if and only if $P(X = c) = 1$ for some $c \in \mathbb{R}$.

Theorem 8.32 (Properties of covariance)**.**

$$Cov[X, Y] = \mathbb{E}[X, Y] - \mathbb{E}[X]\mathbb{E}[Y]$$

Definition 8.34 (skewness and kurtosis)**.** The standardized third and fourth moments are said to be **skewness** and **kurtosis**, that is,
skewness $= \mathbb{E}[(X-\mu)^3]/\sigma^3$, kurtosis $= \mathbb{E}[(X-\mu)^4]/\sigma^4$
where $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{VAR}\, X$.

9 Inequalities

Theorem 9.1 (Chebychev's inequality)**.** Let X be a random variable with mean μ and variance σ^2. Then, for any $\alpha > 0$,

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

Equivalently, for $\alpha > 0$,

$$P(|X - \mu| > \alpha) \leq \frac{\mathrm{VAR}\, X}{\alpha^2}$$

Theorem 9.2 (Markov's inequality)**.** If $X \geq 0$ with $\mu = \mathbb{E}[X] < \infty$, then for any $\alpha > 0$,

$$P(X \geq \alpha) \leq \mu/\alpha$$

Remark 9.1. The Chebychev's inequality is a special case of Markov's inequality by considering

$$Y = (X - \mu)^2$$

Note that $A = \{s \in \Omega : |X(s) - E(X)| \geq r\} = \left\{s \in \Omega : (X(s) - E(X))^2 \geq r^2\right\}$
Now, consider the random variable, Y, where $Y(s) = (X(s) - E(X))^2$.
Note that Y is a non-negative random variable.
Thus, we can apply Markov's inequality to it, to get:
$P(A) = P\left(Y \geq r^2\right) \leq \frac{E(Y)}{r^2} = \frac{E\left((X-E(X))^2\right)}{r^2} = \frac{V(X)}{r^2}$.

Theorem 9.3 (Cauchy-Schwartz' inequality)**.** Let X and Y be two random variables having finite second moment. Then

$$[\mathbb{E}[XY]]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

where the equality holds if and only if $P(aX = bY) = 1$ for some $a, b \in \mathbb{R}$.

Theorem 9.4. Let X and Y be two random variables with finite second moment. Then $Y = aX + b$ for some a, b if and only if $|Corr(X, Y)| = 1$.

Lemma 9.1 (Young's inequality)**.** For $p, q > 1$ with $1/p + 1/q = 1$ and two nonnegative real numbers $x, y \geq 0$,

$$xy \leq x^p/p + y^q/q$$

Theorem 9.5 (Hölder's inequality)**.** For $p, q > 1$ with $1/p + 1/q = 1$,

$$\mathbb{E}[|XY|] \leq ||X||_p ||Y||_q$$

when the expectations exist and are finite where $||X||_r = \mathbb{E}[|X|^r]^{1/r}$ for $r > 0$.

Remark 9.2. The Cauchy-Schwartz' inequality is a special case of Hölder's inequality ($p = q = 2$)

Theorem 9.6 (Jensen's inequality)**.** For a convex function φ,

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$$

Theorem 9.7 (Minkowski's inequality)**.** For $p \geq 1$,

$$||X + Y||_p \leq ||X||_p + ||Y||_p$$

10 Conditional Expectation

Definition 10.1. conditional expectation The conditional expectation of Y given $X = x$ is defined by

$$\mathbb{E}[Y|X = x] = \int y \, dcdf_{Y|X}(y|x)$$

Remark 10.1. The conditional expectation $\mathbb{E}[Y|X = x]$ is always a function of x, say $h(x)$. Then denote $h(X) = \mathbb{E}[Y|X]$ as a random variable.

Theorem 10.1. Assume $\mathbb{E}[|Y|] < \infty$. Then

$$\mathbb{E}[Y|X = x] = \int_0^\infty P(Y > z|X = x) \, dz - \int_{-\infty}^0 P(Y < z|X = x) \, dz$$

If Y is discrete, then

$$\mathbb{E}[Y|X = x] = \sum_y y \times pmf_{Y|X}(y|x)$$

If Y is continuous, then

$$\mathbb{E}[Y|X = x] = \int y \times pmf_{Y|X}(y|x) \, dy$$

Theorem 10.2 (Properties of conditional expectation)**.** Satisfies

1. $\mathbb{E}[aY + b|X] = a\mathbb{E}[Y|X] + b$
2. If $P(Y \geq 0|X) = 1$, then $\mathbb{E}[Y|X] \geq 0$
3. $\mathbb{E}[Y + Z|X] = \mathbb{E}[Y|X] + \mathbb{E}[Z|X]$
4. for constant random variable 1, $\mathbb{E}[1|X] = 1$
5. for convex function ϕ, $\varphi(\mathbb{E}[Y|X]) \leq \mathbb{E}[\varphi(Y)|X]$

Theorem 10.3 (Law of Total Expectation)**.**

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

i.e. The expected value of the conditional expected value of Y given X is the same as the expected value of Y. One special case states that if $\{A_i\}_i$ is a finite or countable partition of the sample space, then

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X|A_i]P(A_i)$$

Definition 10.2. conditional variance The conditional variance is given by

$$\mathrm{VAR}\, Y|X = x = \mathbb{E}[(Y - \mathbb{E}[Y|X = x])^2|X = x]$$

Theorem 10.4.

$$\mathrm{VAR}\, Y = \mathbb{E}[\mathrm{VAR}\, Y|X] + \mathrm{VAR}\, \mathbb{E}[Y|X]$$

11 Probability Related Functions

Let X be a random variable.

1. **moment generating function**: $mgf_X(t) = \mathbb{E}[e^{tX}]$
2. **cumulant generating function**: $cgf_X(t) = \log \mathbb{E}[e^{tX}]$
3. **probability generating function**: $pgf_X(t) = \mathbb{E}[z^X]$
4. **characteristic generating function**: $chf_X(t) = \mathbb{E}[e^{itX}]$

where $t \in \mathbb{R}, z > 0$ and $i = \sqrt{-1}$ is the unit imaginary number.

Theorem 11.1 (properties of mgf)**.** As follows

1. $mgf_X(0) = 1$
2. $\mathbb{E}[X^k] = \frac{d^k}{dt_k} mgf_X(0)$ if it exists
3. If $\mathbb{E}[|X|^k] < \infty$, then for $\mu_j = \mathbb{E}[X^j]$ where $j = 1, \ldots, k$,
$$mgf_X(t) = 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \ldots + \mu_k \frac{t^k}{k!} + o(|t|^k)$$
4. $mgf_{aX+b}(t) = e^{bt} mgf_X(at)$
5. If X and Y are independent, then
$$mgf_{X,Y}(s,t) = mgf_X(s) mgf_Y(t)$$

Theorem 11.2 (properties of cgf)**.** As follows

1. $cgf_X(0) = 0$
2. If X and Y are independent, then
$$cgf_{X,Y}(s,t) = cgf_X(s) + cgf_Y(t)$$

Theorem 11.3 (properties of pgf)**.** As follows

1. $pgf_X(1) = 1$
2. $\mathbb{E}[X(X-1)\ldots(X-k+1)] = \frac{d^k}{dz^k} pgf_X(1)$ if it exists.
3. If X and Y are independent, then
$$pgf_{X,Y}(s,t) = pgf_X(s) + pgf_Y(t)$$

Theorem 11.4 (properties of chf)**.** As follows

1. $chf_X(0) = 1$
2. $\mathbb{E}[X^k] = (i)^{-k}\frac{d^k}{dt^k}chf_X(0)$ if it exists
3. If $\mathbb{E}[|X|^k] < \infty$, then for $\mu_j = \mathbb{E}[X^j]$ where $j = 1, \ldots, k$,
$$chf_X(t) = 1 + i\mu_1 t - \mu_2\frac{t^2}{2!} + \ldots + i^k\mu_k\frac{t^k}{k!} + o(|t|^k)$$
4. $chf_{aX+b} = e^{ibt}chf_X(at)$
5. If X and Y are independent, then
$$chf_{X,Y}(s,t) = chf_X(s)chf_Y(t)$$
6. $|chf_X(t)| \leq 1$ for all t
7. chf is uniformly continuous
8. for any $t_1, \ldots, t_n \in \mathbb{R}$ and $z_1, \ldots, z_n \in \mathbb{C}$,
$$\sum_{j,k} chf_X(t_j - t_k)z_j\bar{z}_k \geq 0$$

Theorem 11.5. If two random variables X and Y have the same moment generating functions in an open neighbourhood of 0, that is, $(-a,b)$ for $a,b > 0$, then X and Y are identically distributed.

Theorem 11.6. If a function $\varphi : \mathbb{R} \to \mathbb{C}$ satisfies 5 - 8 in Theorem 11.4, then there exists a random variable having φ as its characteristic function.

Definition 11.1. The joint probability/moment/cumulant generating and characteristic functions of X and Y are

1. $mgf_{X,Y}(s,t) = \mathbb{E}[e^{sX+tY}]$
2. $cgf_{X,Y}(s,t) = \log mgf_{X,Y}(s,t)$
3. $pgf_{X,Y}(s,t) = \mathbb{E}[s^X t^Y]$
4. $chf_{X,Y}(s,t) = \mathbb{E}[e^{isX+itY}]$

Theorem 11.7 (Inversion Formula)**.** Let φ be a characteristic function of a random variable X. Then for any a, b,
$$P(a < X < b) + \{P(X = a) + P(X = b)\}/2 = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iat} - e^{-ibt}}{it}\varphi(t)\,dt$$

Theorem 11.8 (Chernoff Bound)**.** Let X be a random variable having moment generating function. For any constant x,
$$P(X \geq x) \leq \inf_{t>0} e^{-xt}mgf_X(t)$$

11.1 Survival Functions

Let X be a non-negative valued random variable.
The **survival** function of X is $S_X(t) = P(X > t)$ or $S_X(t) = 1 - F_X(t)$.
(the probability of surviving longer than time x.
The **hazard** function is

$$h_X(t) = \frac{pdf_X(t)}{S_X(t)} = \frac{pdf_X(t)}{1 - F_X(t)}$$

(measures the risk of event (or death) at time x. The **cumulative hazard** function is

$$H_X(t) = \int_0^t h_X(z)\, dz$$

for $t > 0$.
The **residual** (or future) lifetime given $X > t$ is defined by

$$R_X(t) = X - t$$

The **mean residual lifetime** is the conditional expectation of residual lifetime given $X > t$, that is,

$$\mathbb{E}[R_X(t)|X > t] = \int_0^\infty P(R_X(t) > z|X > t)\, dz = \int_t^\infty \frac{S_X(z)}{S_X(t)} \tag{7}$$

Particularly for $t = 0$ and $S_X(0) = 1$,

$$\mathbb{E}[R_X(0)|X > 0] = \int_0^\infty S_X(z)\, dz = \mathbb{E}[X]$$

12 Stochastic process

Definition 12.1. A **stochastic process** is a collection of time indexed random variables

$$\{X_t : t \in \mathcal{T}\}$$

A collection of σ-field $\mathcal{F} = \{\mathcal{F}_t\}_{t\in\mathcal{T}}$ is called a **filtration** if $\mathcal{F} \subset \mathcal{F}_t$ for any $0 \le s \le t$.
A stochastic process $X = \{X_t\}_{t\in\mathcal{T}}$ is said to be **adapted to the filtration** $\mathcal{F}$ if X_t is $\mathcal{F}_t$-measurable (or $\{X_t \le r\} \in \mathcal{F}_t$ for any real number r).

Definition 12.2 (Martingales)**.** A stochastic process X_n is said to be a (discrete-time) **martingale** if

1. $\mathbb{E}[|X_n|] < \infty$
2. $\mathbb{E}[X_{n+1}|X_0, \ldots, X_n] = X_n$ for all n
3. A stochastic process X_n is said to be **supermartingale** if it satisfies above (1) and

   $$\mathbb{E}[X_{n+1}|X_0, \ldots, X_n] \le X_n$$

   for all n.
4. A stochastic process X_n is said to be **submartingale** if it satisfies above (1) and

   $$\mathbb{E}[X_{n+1}|X_0, \ldots, X_n] \ge X_n$$

   for all n.

Note: the condition $X_0, \ldots, X_n$ is often replaced by $\mathcal{F}$, that is,

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$$

Remark 12.1. A martingale is both supermartingale and submartingale.
If X_n is a submartingale, then $-X_n$ is a supermaringale.

Definition 12.3 (stopping time)**.** A time valued random variable T is said to be a **stopping time** if the event $\{T \leq n\}$ can be expressed by $X_0, \ldots, X_n$

Example 12.1. The first time T that the stochastic process X_n is bigger than or equal to a constant K is a stopping time by considering

$$\{T = n\} = \{X_1 < K, \ldots, X_{n-1} < K, X_n \geq K\}$$

Theorem 12.1 (Optional Sampling Theorem)**.** Let X_n be a submartingale and T is a stopping time with $P(T \leq k) = 1$. Then

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_k]$$

12.1 Random Walk

Let $X_1, X_2, \ldots$ be a sequence of independent random variables having mean zero and variance 1. Define $S_n = X_1 + \ldots + X_n$

Theorem 12.2. For any $\alpha > 0$,

$$P(\max_{k=1,\ldots,n} |S_k| \geq \alpha) \leq \frac{\mathrm{VAR}\, S_n}{\alpha^2}$$

Theorem 12.3. If X_n is symmetric for each n, then

$$P(\max_{k=1,\ldots,n} |S_k| \geq \alpha) \leq 2P(S_n \geq \alpha)$$

12.2 Poisson Process

A **Poisson process with intensity** λ is a stochastic process $N = \{N_t : t \geq 0\}$ taking values in non-negative integers satisfying
(a) $N_0 = 0$ and $N_s \leq N_t$ if $0 \leq s \leq t$
(b) $P(N_{t+h} = n + m | N_t = n) = \begin{cases} 1 - \lambda h + o(h) & \text{if } m = 0 \\ \lambda h + o(h) & \text{if } m = 1 \\ o(h) & \text{if } m > 1 \end{cases}$
(c) For $0 \leq s < t$, the arrivals $N_t - N_s$ in the interval $(s, t]$ is independent of the arrivals N_s in the interval $(0, s]$.

Theorem 12.4. For any fixed time $t > 0$, $N_t \sim Poisson(\lambda t)$

Theorem 12.5. The interarrival times $X_1, X_2, \ldots$ are independent and identically distributed from exponential with λ

12.3 Reflection principle (Wiener process)

Definition 12.4 (Wiener Process)**.** A continuous-time stochastic process $W(t)$ for $t \geq 0$ with $W(0) = 0$ and such that the increment $W(t) - W(s)$ is Gaussian with mean 0 and variance $t - s$ for any $0 \leq s < t$, and increments for nonoverlapping time intervals are independent.

Remark 12.2. Brownian motion (i.e. random walk with random step sizes) is the most common example of a Wiener process.

Theorem 12.6 (Reflection principle)**.** If $(W(t) : t \geq 0)$ is a Wiener process, and $a > 0$ is a threshold, then

$$P\left(\sup_{0 \leq s \leq t} W(s) \geq a\right) = 2P(W(t) \geq a)$$

Remark 12.3. If the path of a Wiener process $f(t)$ reaches a value $f(s) = a$ at time $t = s$, then the subsequent path after time s has the same distribution as the reflection of the subsequent path about the value a.

13 Mode of Convergence

Definition 13.1. Modes of convergence

- A sequence of random variables X_n converges to X **in distribution** ($X_n \xrightarrow{d} X$) if
$$P(X_n \le x) \to P(X \le x)$$
as $n \to \infty$ for any x with $P(X = x) = 0$.
- A sequence of random variables X_n converges to X **in probability** ($X_n \xrightarrow{p} X$) if
$$P(|X_n - X| > \epsilon) \to 0$$
as $n \to \infty$
- A sequence of random variables X_n converges to X **almost surely** ($X_n \xrightarrow{a.s.} X$) if
$$P(\limsup_{n\to\infty} |X_n - X| = 0) = 1$$
- A sequence of random variables X_n converges to X **in** L^p ($X_n \xrightarrow{L^p} X$) for $p > 0$ if
$$\mathbb{E}[|X_n - X|^p] \to 0$$
as $n \to \infty$

Theorem 13.1. Let X_n and X be discrete random variables with probability mass functions $f_n(x)$ and $f(x)$ satisfying $f_n(x) \to f(x)$ for any x with $f(x) > 0$. Then
$$X_n \longrightarrow X$$
in distribution.

Theorem 13.2 (Relations between modes of convergence)**.** As follows:
(a) $X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{p} X$
(b) $X_n \xrightarrow{L^p} X \implies X_n \xrightarrow{p} X$
(c) $X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$

13.1 L^1 Convergence

Lemma 13.1 (L^1 Convergence)**.** If $Y \ge 0$ and $\mathbb{E}[[]Y] < \infty$, then for any $\epsilon > 0$ there exists $M > 0$ such that
$$\mathbb{E}[Y\mathbb{1}\{Y > M\}] < \epsilon$$

Lemma 13.2. Suppose a random variable Y has a finite absolute expectation, that is, $\mathbb{E}[|Y|] < \infty$. For any $\epsilon > 0$, there exists $\delta > 0$ such that $|\mathbb{E}[Y\mathbb{1}\{A\}]| < \epsilon$ for any event A with $P(A) < \delta$ where $\mathbb{1}\{A\}$ is an indicator function of the event A.

Lemma 13.3. Suppose a random variable Y has a finite absolute expectation, that is, $\mathbb{E}[|Y|] < \infty$ and a sequence A_n of events satisfy $P(A_n) \to 0$. Then
$$\mathbb{E}[Y\mathbb{1}\{A_n\}] \to 0$$

Theorem 13.3 (Dominated Convergence Theorem)**.** Suppose that $X_n \to X$ in probability, $|X_n| \le Y$ and $\mathbb{E}[Y] < \infty$. Then
$$\mathbb{E}[X_n] \to \mathbb{E}[X]$$

Theorem 13.4 (Generalized Dominated Convergence Theorem)**.** If all X, Y, X_n, Y_n have finite absolute expectation, $|X_n| \leq Y_n$ for all n, $X_n \to X$ in probability, $Y_n \to Y$, and $\mathbb{E}[Y_n] \to \mathbb{E}[Y]$, then

$$\mathbb{E}[X_n] \to \mathbb{E}[X]$$

Theorem 13.5 (Monotone Convergence Theorem)**.** Let X_n be non-negative non-decreasing random variables. Suppose $\lim_{n\to\infty} X_n = X$ is finite a.s. Then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$$

Theorem 13.6 (Fatou's lemma)**.** Let $X_1, X_2, \ldots$ be a sequence of non-negative random variables. Then

$$\mathbb{E}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]$$

13.2 Almost Sure Convergence

Theorem 13.7 (Borel-Cantelli lemma)**.** Let $A = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} A_n$ be the event that infinitely many A_n's occur.

1. $P(A) = 0$ if $\sum_n P(A_n) < \infty$
2. $P(A) = 1$ if $\sum_n P(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent.

Theorem 13.8. If for any $\epsilon > 0$, $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$, then $X_n \to X$ almost surely.

Theorem 13.9. If a sequence of random variables X_n converges to X in probability, then there exists a subsequence n_k such that X_{n_k} converges to X almost surely.

Theorem 13.10. A sequence x_n of real numbers converges to x if and only if for any subsequence n_k there exists a further subsequence n_{k_l} such that $x_{n_{k_l}}$

Theorem 13.11. A sequence of random variables X_n converges to X in probability if and only if for any subsequence n_k there exists a further subsequence n_{k_l} such that $X_{n_{k_l}}$ converges to X a.s.

13.3 Convergence in distribution

Theorem 13.12. As follows
(a) If $X_n \xrightarrow{d} c$ where c is a constant, then $X_n \xrightarrow{p} c$.
(b) If $X_n \xrightarrow{p} c$ and $P(|X_n| \leq M) = 1$ for some $M > 0$, then $X_n \xrightarrow{L^p} X$ for any $p > 0$

Theorem 13.13. Let X be a random variable with $P(X = x) = 0$ for all x and F be the distribution function of X. Then $F(X) \sim uniform(0, 1)$ and $F^{-1}(U) \sim X$ for any $U \sim uniform(0, 1)$

Theorem 13.14 (Skorokhod's representation theorem)**.** If $X_n \xrightarrow{d} X$, then there exist random variables $Y, Y_1, Y_2, \ldots$ in a probability space such that
(a) X_n and Y_n have the same distribution as well as X and Y have the same distribution
(b) $Y_n \xrightarrow{a.s.} Y$

Theorem 13.15 (Continuous mapping theorem)**.** Let g be a continuous function.

1. $X_n \xrightarrow{a.s.} X \implies g(X_n) \xrightarrow{a.s.} g(X)$
2. $X_n \xrightarrow{p} X \implies g(X_n) \xrightarrow{p} g(X)$
3. $X_n \xrightarrow{d} X \implies g(X_n) \xrightarrow{d} g(X)$

Theorem 13.16. $X_n \xrightarrow{d} X$ if and only if $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for any bounded continuous function g.

Theorem 13.17. $X_n \xrightarrow{d} X$ if and only if

$$chf_{X_n}(t) \to chf_X(t)$$

Theorem 13.18. If $X_n \xrightarrow{d} X$, then

$$aX_n + b \xrightarrow{d} aX + b$$

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

Theorem 13.19 (Slutsky's lemma)**.** Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ for a constant c.

1. $X_n + Y_n \xrightarrow{d} X + c$
2. $X_n Y_n \xrightarrow{d} Xc$
3. $X_n / Y_n \xrightarrow{d} X/c$ if $c \neq 0$

14 Law of Large Numbers

Theorem 14.1 (Weak Law of Large Numbers)**.** Let X_n be i.i.d. with $\mathbb{E}[|X_n|] < \infty$. Then

$$\bar{X}_n \xrightarrow{p} \mathbb{E}[X_1]$$

Theorem 14.2 (Strong Law of Large Numbers)**.** Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with $\mathbb{E}[|X_n|] < \infty$. Then

$$\bar{X}_n \xrightarrow{a.s.} \mathbb{E}[X_1]$$

Theorem 14.3. Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with $\mathbb{E}[X_n^2] < \infty$.

$$\bar{X}_n = (X_1 + \ldots + X_n)/n \longrightarrow \mathbb{E}[X_1]$$

almost surely and in L^2.

15 Central Limit Theorem

For $k \approx np$, the binomial probability is approximated by

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(k-np)^2}{2np(1-p)}\right)$$

Theorem 15.1 (Levy's Central Limit Theorem)**.** Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{VAR}\, X_i$. Then

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$$

Theorem 15.2 (Lindeberg-Feller Central Limit Theorem)**.** Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with $\mathbb{E}[X_i] = 0$ and $\sigma_i^2 = \mathrm{VAR}\, X_i^2 < \infty$. Let $s_n^2 = \mathbb{E}[X_1^2] + \ldots + \mathbb{E}[X_n^2]$ The Lindeberg condition

$$\frac{1}{s_n^2 \sum_{k=1}^n \mathbb{E}[X_k^2 \mathbb{1}\{X_k^2 > \epsilon s_n^2\}]} \to 0$$

for any $\epsilon > 0$ holds if and only if

$$(X_1 + \ldots + X_n)/s_n \xrightarrow{d} N(0,1)$$

and

$$\max(\sigma_1^2, \ldots, \sigma_n^2)/s_n^2 \to 0$$

Theorem 15.3 (Lyapounov's condition)**.** Let $X_1, \ldots, X_n$ be i.i.d. r.v.s with $\mathbb{E}[X_i] = 0$ and $\sigma_i^2 = \mathrm{VAR}\, X_i^2 < \infty$ satisfying Lyapounov's condition

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}[|X_k|^{2+\delta}] = 0$$

Then Lindeberg's condition holds. Hence

$$(X_1 + \ldots + X_n)/s_n \xrightarrow{d} N(0,1)$$

Theorem 15.4 (δ-method)**.** Let $X_1, \ldots, X_n$ be i.i.d. r.v.s anf a_n is a sequence of positive real numbers diverging to infinity. If $a_n(X_n - \mu) \xrightarrow{d} Z$ for some r.v. Z and a constant μ, then for any continuously differentiable function g,

$$a_n(g(X_n) - g(\mu)) \xrightarrow{d} g'(\mu)Z$$