2017-2018

Probability

Instructor: Jay Cheng

Typing: Yang Jing Xuan

2018/1/6

Contents

Chapter.1 Axioms of Probability

Definition 1.1 Sample Space

The **sample space** Ω of an experiment is the set of **all possible outcomes** of the experiment.

Definition 1.2 Event

An **event** of an experiment is a **subset** of the sample space Ω of the experiment.
We call **Ω** the **certain** event and **Φ** the **impossible** event of the experiment.
We say that an event A **occurs** if the outcome of the experiment **belongs** to A.

Definition 1.3 σ-algebra

A **σ-algebra** $\mathcal{A}$ of subsets of a sample space Ω is **a collection of subset** of Ω s.t.
(1) $\Omega \in \mathcal{A}$
(2) $\mathcal{A}$ is **closed under complementation**, i.e., if $A \in \mathcal{A}$, then $\Omega \backslash A \in \mathcal{A}$
(3) $\mathcal{A}$ is **closed under countable union**, i.e., if $A_n \in \mathcal{A}$ for $n = 1,2, \dots$,
then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Theorem 1.1 Properties of σ-algebra

Suppose $\mathcal{A}$ is a σ-algebra of subsets of a sample space Ω.
(1) $\Phi \in \mathcal{A}$
(2) $\mathcal{A}$ is closed under finite union
(3) $\mathcal{A}$ is closed under countable and finite intersection.

Theorem 1.2 Intersection of σ-algebra

Suppose Γ is a nonempty collection of σ-algebra of subsets of a sample space Ω.
Then the **intersection** $B = \bigcap_{A \in \Gamma} A$ of the σ-algebra in Γ is **also** a σ-algebra of subsets of Ω.

Corollary 1.1 Existence of Smallest σ-algebra

Suppose $\mathcal{C}$ is a **collection of subsets** of a sample space Ω.
Then there exists a **smallest σ-algebra** of subsets of Ω including $\mathcal{C}$.

Definition 1.4 Generated σ-algebra

Let $\mathcal{C}$ be a collection of subsets of a sample space Ω,
we define the σ-algebra of subsets of Ω **generated** by $\mathcal{C}$ as the **smallest σ-algebra** of subsets of Ω including $\mathcal{C}$ and denoted it as **σ(**$\mathcal{C}$**)**.

Definition 1.5 Probability Measure

Let $\mathcal{A}$ be a σ-algebra of subsets of a sample space Ω, a **probability measure** $P: \mathcal{A} \to \mathbb{R}$ on $\mathcal{A}$ is a real-valued function on $\mathcal{A}$ s.t.

(1) **Nonnegativity**: $P(A) \geq 0\ \forall A \in \mathcal{A}$

(2) **Normalization**: $P(\Omega) = 1$

(3) **Countable additivity**: If $A_1, A_2, \ldots$ are pairwise disjoint events in $\mathcal{A}$

then $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A)$.

For an event $A \in \mathcal{A}$, we call $P(A)$ the probability of the event A.

Definition 1.6 Probability Space

A probability space is an **ordered triple** $(\Omega, \mathcal{A}, P)$ consisting of a sample space Ω, a σ-algebra $\mathcal{A}$ of subsets of Ω, and a probability measure P on $\mathcal{A}$.

Theorem 1.3 A Kind of Probability Measure

Suppose $\Omega = \{w_1, w_2, \ldots\}$, $\mathcal{A} \in \mathcal{P}(\Omega)$ and $P(A) = \sum_{w_i \in \mathcal{A}} P_i$ for all $A \in \mathcal{P}(\Omega)$. Where $P_i \geq 0\ \forall i = 1,2, \ldots$ and $\sum_{i=1}^{\infty} P_i = 1$, then P is a **probability measure** on $\mathcal{P}(\Omega)$. A similar result holds if $\Omega = \{w_1, w_2, \ldots, w_N\}$, where $N \geq 1$.

Corollary 1.2 A Kind of Probability Measure (*special*)

Suppose $\Omega = \{w_1, w_2, \ldots, w_N\}$, $\mathcal{A} \in \mathcal{P}(\Omega)$, and $P(A) = \frac{|A|}{N}$ for all $A \in \mathcal{P}(\Omega)$, then P is a **probability measure** on $\mathcal{P}(\Omega)$.

Theorem 1.4 Classical definition of probability

Suppose $\Omega = \{w_1, w_2, \ldots, w_N\}$, $\mathcal{A} \in \mathcal{P}(\Omega)$ and P is a **probability measure** on $\mathcal{P}(\Omega)$ such that $P(\{w_1\}) = P(\{w_2\}) = \cdots = P(\{w_N\})$, then $P(A) = \frac{|A|}{N}$ for all $A \in \mathcal{P}(\Omega)$.

Theorem 1.5 Properties of Probability Measure

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space.

(1) $P(\Phi) = 0$

(2) $P(A) + P(A^c) = 1$. Therefore, $0 \leq P(A) \leq 1$, for all $A \in \mathcal{A}$.

(3) **Finite additivity**: If $A_1, A_2, \ldots, A_N$ are **pairwise disjoint** events in $\mathcal{A}$,

$$\text{then} \qquad P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A).$$

Theorem 1.6 Properties of Probability Measure

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A, B \in \mathcal{A}$.

(1) If $A_1, A_2, \ldots$ are **pairwise disjoint** events on $\mathcal{A}$ and $\bigcup_{n=1}^{\infty} A_n = \Omega$, then $P(A) = \sum_{i=1}^{\infty} P(A \cap A_n)$.

(2) If $B \subseteq A$, then $P(A) = P(A \cap B) + P(A \cap A^c)$ for all $A, B \in \mathcal{A}$.

(3) $P(A \cap B) \leq \min\{P(A), P(B)\} \leq \max\{P(A), P(B)\} \leq P(A \cup B)$.

Corollary 1.3 Finite Additivity under Union
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, $A \in \mathcal{A}$, $A_1, A_2, \ldots$ are **pairwise disjoint** events in $\mathcal{A}$, and $P(\bigcup_{n=1}^{\infty} A_n) = 1$, then

$$P(A) = \sum_{n=1}^{\infty} P\left(A \bigcap A_n\right)$$

Theorem 1.7 Inclusion-exclusion identity
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A_1, A_2, \ldots, A_n \in \mathcal{A}$, where $n \geq 2$, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P\left(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k}\right)$$

Lemma 1.1 Generated Pairwise Disjoint
Suppose $\mathcal{A}$ is a σ-algebra of subsets of a sample space Ω, suppose $A_1, A_2, \ldots \in \mathcal{A}$, $B_1 = A_1$, and $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i$ for all $n \geq 2$,
then $B_1, B_2, \ldots$ are **pairwise disjoint** events in $\mathcal{A}$, $\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$ for all $n \geq 1$, and $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$.

Theorem 1.8 Inclusion-exclusion inequality
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A_1, A_2, \ldots, A_n \in \mathcal{A}$, where $n \geq 2$, then

$$P\left(\bigcup_{i=1}^{n} A_i\right) \begin{cases} \leq \displaystyle\sum_{k=1}^{m} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P\left(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k}\right), \text{ if m is odd} \\ \geq \displaystyle\sum_{k=1}^{m} (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P\left(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k}\right), \text{ if m is even} \end{cases}$$

Where $1 \leq m \leq n$.

In particular,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i),$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P\left(A_i \bigcap A_j\right).$$

Theorem 1.9 Boole's inequality
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A_1, A_2, \ldots \in \mathcal{A}$,
then $P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$.

Definition 1.7 Monotonicity
Let $(\Omega, \mathcal{A}, P)$ be a probability space.
A sequence $\{A_1, A_2, \ldots\}$ of events in $\mathcal{A}$ is **increasing** if $A_1 \subseteq A_2 \subseteq \cdots$
A sequence $\{A_1, A_2, \ldots\}$ of events in $\mathcal{A}$ is **decreasing** if $A_1 \supseteq A_2 \supseteq \cdots$

Definition 1.8 Limit of Events
Let $(\Omega, \mathcal{A}, P)$ be a probability space.

(1) The **limit** $\lim\limits_{n\to\infty} A_n$ of an **increasing** sequence $\{A_1, A_2, \ldots\}$ of events in $\mathcal{A}$ is the event that **at least one** of the events occurs, i.e., $\lim\limits_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

(2) The **limit** $\lim\limits_{n\to\infty} A_n$ of a **decreasing** sequence $\{A_1, A_2, \ldots\}$ of events in $\mathcal{A}$ is the event that **all** the events occur, i.e., $\lim\limits_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Theorem 1.10 Continuity of probability measure
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space.
(1) Suppose that $\{A_1, A_2, \ldots\}$ is an **increasing** sequence of events in $\mathcal{A}$.
Then $P\left(\lim\limits_{n\to\infty} A_n\right) = \lim\limits_{n\to\infty} P(A_n)$.
(2) Suppose that $\{A_1, A_2, \ldots\}$ is a **decreasing** sequence of events in $\mathcal{A}$.
Then $P\left(\lim\limits_{n\to\infty} A_n\right) = \lim\limits_{n\to\infty} P(A_n)$.

Remark 1.1 Not Certain or Impossible
If $P(A) = 0$, then it is **not necessary** that $A = \Phi$,
e.g., $\Omega = (0,1)$ and $A = A_\alpha, \alpha \in (0,1)$.
If $P(A) = 1$, then it is **not necessary** that $A = \Omega$,
e.g., $\Omega = (0,1)$ and $A = A_\alpha^c, \alpha \in (0,1)$.

Definition 1.9 Length
The **length** of the intervals $(a,b), [a,b), (a,b], [a,b]$ are defined to be $(b-a)$.

Definition 1.10 Random
A point is said to be **randomly** selected from an interval (a,b) if **any** subintervals of (a,b) with the same length are **equally likely** to contain the randomly selected point.

Theorem 1.11　Probability of Randomness

The **probability** that a randomly selected point from (a, b) falls in the subinterval (α, β) of (a, b) is

$$\frac{\beta - \alpha}{b - a}$$

Definition 1.11　Borel Algebra

The σ-algebra of subsets of (a, b) generated by the set of all subintervals of (a, b) is called **Borel algebra** associated with (a, b) and is denoted $\mathscr{B}_{(a,b)}$.

Theorem 1.12　Existence of Probability Measure

For **any** interval (a, b), there **exists** a **unique** probability measure P on $\mathscr{B}_{(a,b)}$ s.t.,

$$P\big((a, b)\big) = \frac{\beta - \alpha}{b - a}$$

for all $(\alpha, \beta) \subseteq (a, b)$.

Chapter.2 Combinational Methods

Theorem 2.1 Counting Principle

There are $n_1 \times n_2 \times \cdots \times n_k$ different ways in which we can first choose an element from a set of n_1 elements, then an element from a set of n_2 elements,... , and finally an element from a set of n_k elements.

Definition 2.1 Permutation

An **ordered** arrangement of *r* objects from a set A containing *n* objects is called an *r*-arrangement permutation of A, where $0 \leq r \leq n$.

An *n*-element permutation of A is called a permutation of A.

The **number** of different *r*-permutation **permutations** of A is given by

$$_nP_r = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}.$$

Theorem 2.2 Permutation with Types

The **number** of different (w.r.t. types) **permutations** of *n* objects of ***k* different types** is

$$\frac{n!}{n_1! \times n_2! \times \cdots \times n_k!}$$

where n_1 are alike, n_2 are alike,... , n_k are alike, and $n = n_1 + n_2 + \cdots + n_k$.

Definition 2.2 Combination

An **unordered** arrangement of *r* objects from a set A containing n objects is called an r-element **combination** of A.

The **number** of different *r*-element **combinations** of A is given by

$$_nC_r = \binom{n}{r} = \frac{_nP_r}{r!} = \frac{n!}{(n-r)!\,r!}.$$

Theorem 2.3 Property of Combination

$$\sum_{i=0}^{k} \binom{n+i}{i} = \sum_{i=0}^{k} \binom{n+i}{n} = \binom{n+k+1}{k}$$

Theorem 2.4 Multinomial Expansion

$$(x_1 + x_2 + \cdots + x_k)^n = \sum_{\substack{n_1+n_2+\cdots+n_k=n \\ n_1,n_2,\ldots,n_k \geq 0}} \frac{n!}{n_1! \times n_2! \times \cdots \times n_k!} {x_1}^{n_1} {x_2}^{n_2} \cdots {x_k}^{n_k}, \forall n \geq 0$$

Corollary 2.1 Binomial Expansion

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}, \forall n \geq 0$$

Theorem 2.5 Stirling's Formula

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \exp\left(\frac{1}{12n} - \frac{1}{360n^2}\right) < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \exp\left(\frac{1}{12n}\right), \forall\, n \geq 1$$

Therefore,

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \text{ i.e. }, \lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1$$

Chapter.3 Conditional Probability and Independence

Definition 3.1 Conditional Probability

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $A, B \in \mathcal{A}$. The **conditional probability** of A given B, denoted $P(A|B)$, is given by

$$P(A|B) = \begin{cases} \dfrac{P(A \cap B)}{P(B)}, & \text{if } P(B) > 0 \\ 0 & \text{, if } P(B) = 0 \end{cases}$$

Remark 3.1 Property of Conditional Probability

$$P(A \cap B) = P(B) \cdot P(A|B), \forall A, B \in \mathcal{A}$$

Theorem 3.1 Conditional Probability Space

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $P(B) > 0$, for some $B \in \mathcal{A}$. Then the **conditional** probability function $P(\cdot\,|B): \mathcal{A} \to \mathbb{R}$ is a **probability measure** on $\mathcal{A}$, and hence $(\Omega, \mathcal{A}, P(\cdot\,|B))$ is a **probability space**.

Theorem 3.2 Reduction of Probability Space

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $P(B) > 0$, for some $B \in \mathcal{A}$.
Let $\mathcal{A}_B: \{A \in \mathcal{A}: A \subseteq B\}$ and $P_B(A) = P(A|B)$ for all $A \in \mathcal{A}_B$.
Then $\mathcal{A}_B$ is a **σ-algebra** of subsets of B and P_B is a **probability measure** on $\mathcal{A}_B$, and hence $(B, \mathcal{A}_B, P_B)$ is a **probability space**.

Remark 3.2 Conversion of Reduced and Conditional Probability Space

Note that $P(A|B) = P(A \cap B|B) = P_B(A \cap B),\ \forall A \in \mathcal{A}$.
And $P(A|B) = P_B(A)$, if $A \in \mathcal{A}$ and $A \subseteq B$.

Theorem 3.3 Law of Multiplication

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A_1, A_2, \dots, A_n \in \mathcal{A}$.
Then $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$.

Theorem 3.4 Law of Total Probability (*infinite*)

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $B_1, B_2, \dots \in \mathcal{A}$ are pairwise disjoint and $\bigcup_{n=1}^{\infty} B_n = \Omega$.
Then, (1) $P(A) = \sum_{n=1}^{\infty} P(B_n) \cdot P(A|B_n), \forall A \in \mathcal{A}$
(2) $P(A|B) = \sum_{n=1}^{\infty} P(B_n|B) \cdot P(A|B \cap B_n),\ \forall A, B \in \mathcal{A}$

Corollary 3.1 Law of Total Probability (*finite*)

Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $B_1, B_2, \dots, B_n \in \mathcal{A}$ are pairwise disjoint and $\bigcup_{i=1}^{n} B_i = \Omega$.
Then, (1) $P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A|B_i),\ \forall A \in \mathcal{A}$
(2) $P(A|B) = \sum_{i=1}^{n} P(B_i|B) \cdot P(A|B \cap B_i),\ \forall A, B \in \mathcal{A}$

Theorem 3.5 Bayes' Theorem (*infinite*)
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $B_1, B_2, \ldots \in \mathcal{A}$ are pairwise disjoint and $\bigcup_{n=1}^{\infty} B_n = \Omega$.
Then

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{n=1}^{\infty} P(B_n) \cdot P(A|B_n)}, \forall\, A \in \mathcal{A}, P(A) > 0, k = 1,2, \ldots$$

Corollary 3.2 Bayes' Theorem (*finite*)
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $B_1, B_2, \ldots, B_n \in \mathcal{A}$ are pairwise disjoint and $\bigcup_{i=1}^{n} B_i = \Omega$.
Then

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{i=1}^{n} P(B_i) \cdot P(A|B_i)}, \forall\, A \in \mathcal{A}, P(A) > 0, k = 1,2, \ldots, n$$

Theorem 3.6 Properties of Conditional Probability
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A, B \in \mathcal{A}$.
(1) $P(A|B) > P(A) \Leftrightarrow P(A \cap B) > P(A) \cdot P(B) \Leftrightarrow P(B|A) > P(B)$
(2) $P(A|B) < P(A), P(B) > 0 \Leftrightarrow P(A \cap B) < P(A) \cdot P(B)$
$\Leftrightarrow P(B|A) < P(B), P(A) > 0$
(3) $P(A|B) = P(A) \Rightarrow P(A \cap B) = P(A) \cdot P(B)$
$P(A \cap B) = P(A) \cdot P(B),\ P(A) = 0$ or $P(B) > 0 \Rightarrow P(A|B) = P(A)$
If $P(A) = 0$ or $P(B) > 0$, then $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$

Definition 3.2 Independence
Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $A, B \in \mathcal{A}$.
If $P(A \cap B) = P(A) \cdot P(B)$, then A and B are said to be **independent**, denoted $A \perp B$.
If A and B are not independent, they are said to be **dependent**.
Furthermore, if $P(A|B) > P(A)$, then A and B are said to be **positively** correlated, and if $P(A|B) < P(A)$, then A and B are said to be **negatively** correlated.

Theorem 3.7 Properties of Independence
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A, B \in \mathcal{A}$.
(1) If $P(A) = 0$ or $P(A) = 1$, then $A \perp B\ \ \forall B \in \mathcal{A}$.
(2) If $A \subseteq B$ and $A \perp B$, then either $P(A) = 0$ or $P(B) = 1$.
(3) If A and B are disjoint and $P(A) > 0$, $P(B) > 0$, then they are dependent.

Theorem 3.8 Independence of Two Events
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A, B \in \mathcal{A}$, and $A \perp B$.
Then $A^* \perp B^*$, i.e., $P(A^* \cap B^*) = P(A^*) \cdot P(B^*), \forall\, A^* = A, A^c;\ B^* = B, B^c$.

Corollary 3.3　Conditional Probability with Independence
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A, B \in \mathcal{A}$, and $A \perp B$.
If $P(B) > 0$, then $P(A^*|B) = P(A^*), \forall\, A^* = A, A^c$.
If $P(B) < 1$, then $P(A^*|B^c) = P(A^*), \forall\, A^* = A, A^c$.

Remark 3.3　Conditional Probability with Independence
If $A \perp B$ and $P(B) > 0$, then knowledge about the occurrence of B **does not** change the probability of the occurrence of A^*.
If $A \perp B$ and $P(B) < 1$, then knowledge about the occurrence of B^c **does not** change the probability of the occurrence of A^*.

Definition 3.3　Independent Set
Let $(\Omega, \mathcal{A}, P)$ is a probability space, and $A_1, A_2, \dots, A_n \in \mathcal{A}$, where $n \geq 2$.

If $P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = P\left(A_{i_1}\right)P\left(A_{i_2}\right)\cdots P\left(A_{i_k}\right), \forall\, 2 \leq k \leq n,$

$$\# = \sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1, 1 \leq i_1 < i_2 < \cdots < i_k \leq n, \qquad \# \triangleq \text{number.}$$

Then $A_1, A_2, \dots, A_n$ are said to be independent; otherwise, they are said to be dependent.

Remark 3.4　Sub Independent Set
If $A_1, A_2, \dots, A_n \in \mathcal{A}$ are independent,
then $A_{i_2}, A_{i_2}, \dots, A_{i_k}$ are independent, $\forall\, 2 \leq k \leq n, 1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

Theorem 3.9　Equivalent Statements of Independence
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, $A_1, A_2, \dots, A_n \in \mathcal{A}$, where $n \geq 2$.
The following statements are **equivalent**:
(1) $A_1, A_2, \dots, A_n$ are independent.

(2) $P\left(A_{i_1}^* \cap A_{i_2}^* \cap \cdots \cap A_{i_k}^*\right) = P\left(A_{i_1}^*\right)\mathrm{P}\left(A_{i_2}^*\right)\cdots \mathrm{P}\left(A_{i_k}^*\right), \forall\, 2 \leq k \leq n,$

$1 \leq i_1 < i_2 < \cdots < i_k \leq n, A_{i_r}^* = A_{i_r}$ or ${A_{i_r}}^c$.

(3) $P\left(A_{i_1}^* \cap A_{i_2}^* \cap \cdots \cap A_{i_n}^*\right) = P\left(A_{i_1}^*\right)P\left(A_{i_2}^*\right)\cdots P\left(A_{i_n}^*\right), \forall A_i^* = A_i, A_i^c, i = 1,2,\dots,n.$

Definition 3.4　Independent Set
Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $A_i \in \mathcal{A}, \forall i \in I$, where I is an index set, then $\{A_i : i \in I\}$ is said to be **independent** if **any finite subset** of $\{A_i : i \in I\}$ is independent; otherwise, it is said to be **dependent**.

Corollary 3.4 Independence under Finite Union
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and suppose $A_1, A_2, \ldots, A_n \in \mathcal{A}$ are independent. Then

$$P\left(\left(A_{i_1}^* \cap A_{i_2}^* \cap \cdots \cap A_{i_k}^*\right) \cap \left(A_{j_1}^* \cap A_{j_2}^* \cap \cdots \cap A_{j_l}^*\right)\right)$$

$$= P\left(A_{i_1}^* \cap A_{i_2}^* \cap \cdots \cap A_{i_k}^*\right) \cdot P\left(A_{j_1}^* \cap A_{j_2}^* \cap \cdots \cap A_{j_l}^*\right)$$

$\forall k, l \geq 1, k + l \leq n, 1 \leq i_1, i_2, \ldots, i_k, j_1, j_2, \ldots, j_l \leq n$ distinct,
and $A_{i_r}^* = A_{i_r}$ or $A_{i_r}^c, r = 1, 2, \ldots, k, A_{j_r}^* = A_{j_r}$ or $A_{j_r}^c, r = 1, 2, \ldots, l.$

In particular,

if $P\left(A_{j_1}^* \cap A_{j_2}^* \cap \cdots \cap A_{j_l}^*\right) > 0$, for some $1 \leq l \leq n-1, 1 \leq j_1, \ldots, j_l \leq n$ distinct,

and $A_{j_r}^* = A_{j_r}$ or $A_{j_r}^c, r = 1, 2, \ldots, l.$

Then $P\left(\left(A_{i_1}^* \cap A_{i_2}^* \cap \cdots \cap A_{i_k}^*\right) | \left(A_{j_1}^* \cap A_{j_2}^* \cap \cdots \cap A_{j_l}^*\right)\right) = P\left(A_{i_1}^* \cap A_{i_2}^* \cap \cdots \cap A_{i_k}^*\right)$

for all $1 \leq k \leq n - l. i_1, i_2, \ldots, i_k \in \{1,2, \ldots, n\} \backslash \{j_1, j_2, \ldots, j_l\}$ distinct,
and $A_{i_r}^* = A_{i_r}$ or $A_{i_r}^c, r = 1, 2, \ldots, k.$

Chapter.4 Distribution Functions and Discrete Random Variables

§ 4.1 Random Variables

Definition 4. 1.1 Measurable Space

A measurable space is an ordered pair $(\Omega,\ \mathcal{A})$ consisting of a sample space Ω and a σ-algebra $\mathcal{A}$ of subsets of Ω.

Definition 4. 1.2 Measurable Function

Let $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$ be measurable spaces.

A function from Ω_1 to Ω_2 is called a measurable function from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ if $f^{-1}(B) \in \mathcal{A}, \forall B \in \mathcal{A}$, where $f^{-1}(B) = \{x \in \Omega: f(x) \in B\}$ is the pre-image of B under *f*.

Lemma 4. 1.1 σ-algebra under Function

Suppose f is a function from Ω_1 to Ω_2.

(1) If $\mathcal{A}_2$ is a **σ-algebra** of subsets of Ω_2, then $\mathcal{A}_1 = \{f^{-1}(B): B \in \mathcal{A}_2\}$ is a **σ-algebra** of subsets of Ω_1.

(2) If $\mathcal{A}_1$ is a **σ-algebra** of subsets of Ω_1, then $\mathcal{A}_2 = \{B \in \Omega_2: f^{-1}(B) \in \mathcal{A}_1\}$ is a **σ-algebra** of subsets of Ω_2.

Theorem 4. 1.1 σ-algebra Including Subset

Suppose $(\Omega_1, \mathcal{A}_1)$ is a measurable space and f is a function from Ω_1 to Ω_2.

If $\mathcal{C} \subseteq \{B \subseteq \Omega_2: f^{-1}(B) \in \mathcal{A}_1\}$, then $\boldsymbol{\sigma}(\mathcal{C}) \subseteq \{B \subseteq \Omega_2: f^{-1}(B) \in \mathcal{A}_1\}$.

Corollary 4.1.1 A Kind of Measurable Function

Suppose $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2)$ are measurable spaces, and f is a function from Ω_1 to Ω_2. Suppose $\mathcal{C} \subseteq \{B \subseteq \Omega_2: f^{-1}(B) \in \mathcal{A}_1\}$ and $\sigma(\mathcal{C}) \supseteq \mathcal{A}_2$.

Then f is a **measurable function** from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$.

Theorem 4.1.2 Composite Measurable Function

Suppose $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2), (\Omega_3, \mathcal{A}_3)$ are measurable spaces,

f is a **measurable function** from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$,

and g is a **measurable function** from $(\Omega_2, \mathcal{A}_2)$ to $(\Omega_3, \mathcal{A}_3)$.

Then $g \circ f$ is a **measurable function** from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_3, \mathcal{A}_3)$.

Definition 4.1.3 Open Set

A set A in $\mathbb{R}^n$ is called an **open set** in $\mathbb{R}^n$ if for all $\underline{x} \in A, \exists r > 0, \Rightarrow \mathcal{B}_{\underline{x}}(r) \subseteq A$,

where $\mathcal{B}_{\underline{x}}(r) = \left\{\underline{y} \in \mathbb{R}^n: \left\|\underline{y} - \underline{x}\right\| < r\right\}$.

Definition 4.1.4 Borel σ-algebra
The σ-algebra generated by the set of all open sets in $\mathbb{R}^n$ is called the **Borel σ-algebra** of subsets of $\mathbb{R}^n$ and is denoted by $\mathcal{B}_{\mathbb{R}^n}$.
We call a set in $\mathcal{B}_{\mathbb{R}^n}$ a **Borel set** in $\mathbb{R}^n$.

Theorem 4.1.3 Measurable Function from Continuity
Suppose f is a **continuous** function from $\mathbb{R}^n$ to $\mathbb{R}^m$.
Then f is a **measurable** function from $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$.

Definition 4.1.5 Cell
A **cell** in $\mathbb{R}$ is a finite interval of the form $(a, b), [a, b), (a, b]$, or $[a, b]$ for some $a \leq b$.
A **cell** I in $\mathbb{R}^n$, where $n \geq 1$, is a Cartesian product of n cells $I_1, I_2, \dots, I_n$ in $\mathbb{R}$, i.e., $I = I_1 \times I_2 \times \cdots \times I_n$.

Definition 4.1.6 Open Cube
Let $\underline{x} \in \mathbb{R}^n, l > 0$, and $I_i = \left(x_i - \frac{l}{2},\ x_i + \frac{l}{2}\right), \forall 1 \leq i \leq n.$
The **open cube** $C_x(l)$ in $\mathbb{R}^n$ with center x and side length l is defined as the **open cell** $I_1 \times I_2 \times \cdots \times I_n\ in\ \mathbb{R}^n$.

Theorem 4.1.4 Set from Cells
Every **open set** in $\mathbb{R}^n$ is a **countable union** of **open cells** in $\mathbb{R}^n$.

Theorem 4.1.5 Measurable Function on Open Cells
Suppose $(\Omega, \mathcal{A})$ is a measurable space and f is a function from Ω to $\mathbb{R}^n$.
Suppose that $f^{-1}(B) \in \mathcal{A}$ for **all open cells** in $\mathbb{R}^n$.
Then f is a **measurable function** from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

Theorem 4.1.6 Components of Measurable Function
Suppose $(\Omega, \mathcal{A})$ is a measurable space, $f = (f_1, f_2, \dots, f_n)$ is a function from Ω to $\mathbb{R}^n$.
Then f is a **measurable function** from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$
$\Leftrightarrow$ $f_1, f_2, \dots, f_n$ are **measurable functions** from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Theorem 4.1.7 Elementary Operation of Measurable Function
Suppose f and g are measurable functions from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, and $c \in \mathbb{R}$.
Then $cf, f^n, |f|, f + g, f \circ g$ are **measurable functions** from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Theorem 4.1.8 Limit of Measurable Functions
Suppose that $f_1, f_2, \dots$ are measurable functions from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $f_n \to f$ as $n \to \infty$, where f is a function from Ω to $\mathbb{R}$.
Then f is **also** a measurable function from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Theorem 4.1.9 Equivalence of Nine Types of Set
Suppose $(\Omega, \mathcal{A})$ is a measurable space and f is a function from Ω to $\mathbb{R}$.
Let $\mathcal{C}_1$ be the set of all open sets in $\mathbb{R}$,

$\mathcal{C}_2 = \{(a,b), a,b \in \mathbb{R}, a \leq b\}$, $\mathcal{C}_3 = \{(a,b], a,b \in \mathbb{R}, a \leq b\}$,
$\mathcal{C}_4 = \{[a,b], a,b \in \mathbb{R}, a \leq b\}$, $\mathcal{C}_5 = \{[a,b), a,b \in \mathbb{R}, a \leq b\}$,
$\mathcal{C}_6 = \{[a,+\infty), a \in \mathbb{R}\}$, $\mathcal{C}_7 = \{(a,+\infty), a \in \mathbb{R}\}$,
$\mathcal{C}_8 = \{(-\infty,a], a \in \mathbb{R}\}$, $\mathcal{C}_9 = \{(-\infty,a), a \in \mathbb{R}\}$.

Then f is a measurable function from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$
if $f^{-1}(B) \in \mathcal{A}\ \forall B \subseteq \mathcal{C}_i$ for **any** $i = 1,2,\cdots,9$.

Theorem 4.1.10 Induced Probability Space under Function
Suppose f is a measurable function from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$.
Suppose P is a probability measure on $\mathcal{A}_1$.
Then the function P_f on $\mathcal{A}_2$ given by

$$P_f(B) = P\left(f^{-1}(B)\right)\ \forall B \in \mathcal{A}_2$$

is a probability measure.
We call $(\Omega_2, \mathcal{A}_2, P_f)$ the probability space **induced** from $(\Omega_1, \mathcal{A}_1, P)$ under f.

Remark 4.1.1 Conventional Denotation
(1) The set $f^{-1}(B)$ is **conventionally** denoted as "$f \in B$".

Therefore $P_f(B) = P\left(f^{-1}(B)\right) = P(f \in B)\ \ \forall B \in \mathcal{A}_2$.

(2) If $B \in \mathcal{A}_2$, then $f^{-1}(B) = f^{-1}(B \cap f(\Omega_1))$, and hence

$$\begin{aligned} P_f(B) &= P(f \in B) = P\left(f^{-1}(B)\right) = P[f^{-1}(B \cap f(\Omega_1))] \\ &= P[f \in (B \cap f(\Omega_1))] = P_f(B \cap f(\Omega_1)) \end{aligned}$$

Definition 4.1.7 Random Variable
Let $(\Omega, \mathcal{A}, P)$ be a probability space.
A measurable function X from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is called a **random variable** (r.v.) of the probability space $(\Omega, \mathcal{A}, P)$.
A measurable function $\underline{X} = (X_1, X_2, \ldots, X_n)$ from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is called a **random vector** (r.vect.) of the probability space $(\Omega, \mathcal{A}, P)$.

Remark 4.1.2 Conventional Denotation of Random Variable
If X is a r.v. of the probability space $(\Omega, \mathcal{A}, P)$,
then $P_X(B) = P\left(X^{-1}(B)\right) = P(X \in B) = P(\{w \in \Omega : X(w) \in B\}), \forall B \in \mathcal{B}_{\mathbb{R}}$.

Theorem 4.1.11 Additivity of Countable Points

Suppose $\underline{X}$ is a r.vect. of a probability space $(\Omega, \mathcal{A}, P)$, and B is a "**countable**" subset of $\mathbb{R}^n$, then $B \in \mathcal{B}_{\mathbb{R}^n}$, and

$$P_{\underline{X}}(B) = P\big(\underline{X} \in B\big) = \sum_{\underline{x} \in B} P\big(\underline{X} = \underline{x}\big) = \sum_{\underline{x} \in B} P_{\underline{X}}\big(\{\underline{x}\}\big).$$

§ 4.2 Distribution Functions

Definition 4.2.1 Cumulative Distribution Function
Let X be a r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
The **cumulative distribution function** (c.d.f) F_X of the r.v. X is a function from $\mathbb{R}$ to $[0,1]$, given by

$$F_X(t) = P_X\big((-\infty, t]\big) = P(X \in (-\infty, t]) = P(X \le t), \forall t \in \mathbb{R}.$$

Theorem 4.2.1 Properties of C.D.F
Suppose X is a r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
(1) F_X is **increasing**.
(2) $F_X(+\infty) \triangleq \lim_{t \to +\infty} F_X(t) = 1$.
(3) $F_X(-\infty) \triangleq \lim_{t \to -\infty} F_X(t) = 0$.
(4) $F_X(t+) = P(X \le t) = F_X(t)$. $F_X(t)$ is **right continuous**.
(5) $F_X(t-) = P(X < t)$.

Corollary 4.2.1 Properties of C.D.F
Suppose X is a r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
(1) $P(X \le a) = F_X(a), P(X > a) = 1 - F_X(a)$.
(2) $P(X < a) = F_X(a-), P(X \ge a) = 1 - F_X(a-)$.
(3) $P(X = a) = F_X(a) - F_X(a-)$.
(4) $P(a < X \le b) = F_X(b) - F_X(a)$, $P(a \le X \le b) = F_X(b) - F_X(a-)$,
$P(a < X < b) = F_X(b-) - F_X(a)$, $P(a \le X < b) = F_X(b-) - F_X(a-)$.

Theorem 4.2.2 Existence of C.D.F
Suppose $F: \mathbb{R} \to [0,1]$ is a function s.t. F is increasing and right continuous,

$$\lim_{t \to +\infty} F_X(t) = 1, \qquad \lim_{t \to -\infty} F_X(t) = 0.$$

Then there **exists** a r.v. X of some probability space $(\Omega, \mathcal{A}, \mathrm{P})$,
s.t. the c.d.f. F_X of X is equal to F.
We call such function a **c.d.f.**

§ 4.3 Discrete Random Variables

Definition 4.3.1 Discrete R.V.
A r.v. of a probability space $(\Omega, \mathcal{A}, P)$ is called a **discrete r.v.**
if $X(\Omega) = \{X(w): w \in \Omega\}$ is **countable**.

Definition 4.3.2 Probability Mass Function
Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ s.t. $X(\Omega) = \{x_1, x_2, \dots\}$.
The **probability mass function** (p.m.f) $p_X: \mathbb{R} \to [0,1]$ of X is a function from $\mathbb{R}$ to $[0,1]$ given by $p_X(x) = P_X(\{X = x\}) = P(X = x), \forall x \in \mathbb{R}$.

Theorem 4.3.1 Properties of P.M.F
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$. Then,
(1) $p_X(x) \geq 0, \forall x \in X(\Omega)$.
(2) $p_X(x) = 0, \forall x \in \mathbb{R} \backslash X(\Omega)$.
(3) $\sum_{x \in X(\Omega)} p_X(x) = 1$.

Therefore if $X(\Omega) = \{x_1, x_2, \dots\}$, then,
(1) $p_X(x_i) \geq 0,\ \forall i = 1,2, \dots$.
(2) $p_X(x) = 0,\ \forall x \in \mathbb{R} \backslash \{x_1, x_2, \dots\}$.
(3) $\sum_{i=1}^{\infty} p_X(x_i) = 1$.

Theorem 4.3.2 Existence of P.M.F
Suppose $p: \mathbb{R} \to [0,1]$ is a function s.t.
(1) $p(x_i) \geq 0,\ \forall i = 1,2, \dots$
(2) $p(x) = 0,\ \forall x \in \mathbb{R} \backslash \{x_1, x_2, \dots\}$.
(3) $\sum_{i=1}^{\infty} p(x_i) = 1$.
for some distinct $x_1, x_2, \dots \in \mathbb{R}$.
Then there **exists** a discrete r.v. X of some probability space $(\Omega, \mathcal{A}, P)$ s.t.
the p.m.f. p_X of X is equal to p.
We call such a function a p.m.f.

Theorem 4.3.3 Step Distribution Function for Discrete R.V.
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ s.t. $X(\Omega) = \{x_1, x_2, \dots\}$,
where $x_1 < x_2 < \cdots$.
Then the distribution function of X is a **step function** given by

$$F_X(t) = \begin{cases} 0 \quad , & \text{if } t < x_1 \\ \sum_{i=1}^{n} p_X(x_i), & \text{if } x_n \leq t \leq x_{n+1},\ n = 1,2, \dots \end{cases} = \sum_{i=1}^{n} p_X(x_i) U(t - x_i),$$

where

$$U(t) = \begin{cases} 1, & \text{if } t \geq 0 \\ 0, & \text{o.w.} \end{cases}$$

§ 4.4 Expectations of Discrete Random Variables

Definition 4.4.1 Expectation
Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
The **expectation (or expected value, or mean)** of X is given by

$$E[X] = \sum_{x \in X(\Omega)} x \cdot P(X = x) = \sum_{x \in X(\Omega)} x \cdot p_X(x)$$

if the sum **converges absolutely**.
And if the sum diverges to $\pm\infty$, $E[X] = \pm\infty$.

Remark 4.4.1 Explanations of Expectation
(1) The expectation $E[X] = \sum_{x \in X(\Omega)} x \cdot p_X(x)$ is the **weighted average** of $\{x: x \in X(\Omega)\}$ with weights $\{P(X = x): x \in X(\Omega)\}$.
(2) The expectation $E[X] = \sum_{x \in X(\Omega)} x \cdot p_X(x)$ is the **center of gravity** of $\{P(X = x): x \in X(\Omega)\}$.

Theorem 4.4.1 Expectation of Constant
Suppose X is a **discrete r.v.** of a probability space $(\Omega, \mathcal{A}, P)$ s.t. X is a **constant** with probability **1**, i.e., $P(X = c) = 1$ for some $c \in \mathbb{R}$.
Then $c \in X(\Omega), P(X = x) = 0, \forall x \in X(\Omega)\backslash\{c\}$, and $E[X] = c$.
In particular, if X is a **constant r.v.** of $(\Omega, \mathcal{A}, P)$, i.e., $X(w) = c, \forall w \in \Omega$, for some $c \in \mathbb{R}$, then $E[X] = c$.

Theorem 4.4.2 Composition of Function and R.V.
Suppose X is a **discrete r.v**. of a probability space $(\Omega, \mathcal{A}, P)$ and g be a **measurable function** from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
Then $g(X) \triangleq g \circ X$ is a discrete r.v. of $(\Omega, \mathcal{A}, P)$ and

$$E[g(X)] = \sum_{x \in X(w)} g(x) P(X = x).$$

Corollary 4.4.1 Linearity of Expectation
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$,
$g_1, g_2, \dots, g_n$ are measurable functions from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,
and $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{R}$,
Then

$$\sum_{i=1}^{n} \alpha_i\, g_i(X)$$

is a discrete r.v. of $(\Omega, \mathcal{A}, P)$ and

$$E\left[\sum_{i=1}^{n} \alpha_i\, g_i(X)\right] = \sum_{i=1}^{n} \alpha_i\, E[g_i(X)].$$

§ 4.5 Variances and Moments of Discrete Random Variables

Definition 4.5.1 Variance and Standard Deviation
Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $E[X]$ exists.
The **variance** of X is given by $Var(X) = E[(X - E[X])^2]$,

and the **standard deviation** of X is given by $\sigma_X = \sqrt{Var(X)}$.

Remark 4.5.1 Explanation about Variance
The variance of a discrete r.v. measures the **dispersion (or spread)** of its probability masses about its expectation (center of gravity of its probability masses).

Theorem 4.5.1 Calculating Variance
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $E[X]$ exists.
Then $Var(X) = E[X^2] - (E[X])^2$.

Theorem 4.5.2 Minimum Distance with Expectation
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $E[X]$ exists.
If $E[X^2] < +\infty$, then $Var(X) = \min_{a \in \mathbb{R}} E[(X - a)^2]$.

Theorem 4.5.3 With Probability 1
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
(1) $E[X^2] \geq 0$, "=" holds $\Leftrightarrow$ $X = 0$ **with probability 1**, i.e., $P(X = 0) = 1$.
(2) If $E[X]$ exists, then $Var(X) \geq 0$, "=" holds $\Leftrightarrow$ $X = E[X]$ **with probability 1**, i.e., $P(X = E[X]) = 1$.

Theorem 4.5.4 Calculating Linear Combination
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $E[X]$ exists.
Then $Var(aX + b) = a^2 Var(X)$ and $\sigma_{aX+b} = |a|\sigma_X, \forall a, b \in \mathbb{R}$.

Definition 4.5.2 Moment and Absolute Moment
Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$, and $r, c \in \mathbb{R}$.

$$\begin{cases} \text{The } r^{th} \textbf{ moment} \text{ of } X \text{ is given by } E[X^r] \\ \text{The } r^{th} \textbf{ central moment} \text{ of } X \text{ is given by } E[(X - E[X])^r] \\ \text{The } r^{th} \textbf{ moment} \text{ of } c \text{ is given by } E[(X - c)^r] \end{cases}$$
$$\begin{cases} \text{The } r^{th} \textbf{ absolute moment} \text{ of } X \text{ is given by } E[|X|^r] \\ \text{The } r^{th} \textbf{ absolute central moment} \text{ of } X \text{ is given by } E[|X - E[X]|^r] \\ \text{The } r^{th} \textbf{ absolute moment} \text{ of } c \text{ is given by } E[|X - c|^r] \end{cases}$$

If the respective sum **converges absolutely**.

Theorem 4.5.6 Existence of Lower Order Moment

Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $0 < r < s$. If $E[|X|^s]$ exists, then $E[|X|^r]$ exists.

That is, the existence of a higher order moment of X **guarantees the existence** of a lower order moment of X.

§ 4.6 Standardized Random Variables

Definition 4.6.1 Standardized R.V.
Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
If $Var(X)$ exists and $Var(X) \neq 0$, then the **standardized r.v.** of X is given by

$$X^* = \frac{X - E[X]}{\sigma_X}$$

i.e., X^* is the number of **standard deviation units** by which X differs from $E[X]$.

Theorem 4.6.1 Expectation and Variance of Standardized R.V.
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and $Var(X)$ exists, $Var(X) \neq 0$.
Then $E[X^*] = 0$ and $Var(X^*) = 1$.

Theorem 4.6.2 Independence of Units
Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and $Var(X)$ exists, $Var(X) \neq 0$.
Then the standardized r.v. of X is **independent of the units** in which X is measured.

Remark 4.6.1 Standardization for Comparison
Standardization can be useful when **comparing** r.v.'s with different distributions.

Charper.5 Special Discrete Distributions

§ 5.1 Bernoulli R.V.'s and Binomial R.V.'s

Definition 5.1.1 Bernoulli Trial
A **Bernoulli trial** is an experiment that has **only two** outcomes, say success and failure, so that its sample space is given by $\Omega = \{s,\ f\}$.

◎ Let X be the number of success in a Bernoulli trial.

$$\Rightarrow p_X(i) = \begin{cases} 1-p, & \text{if } i = 0 \\ p, & \text{if } i = 1 \\ 0, & \text{o.w.} \end{cases}$$

where $p = P(X = 1) = P(\{s\})$ is the probability of success.

Definition 5.1.2 $X \sim$Bernoulli(p)
A discrete r.v. X of a probability space $(\Omega, \mathcal{A}, P)$ is called a **Bernoulli r.v.** with parameter p where $0 < p < 1$, denoted $X \sim \text{Bernoulli}(p)$, if its p.m.f is given by

$$p_X(i) = \begin{cases} 1-p, & \text{if } i = 0 \\ p, & \text{if } i = 1 \\ 0, & \text{o.w.} \end{cases}$$

Such a p.m.f is called a Bernoulli p.m.f with parameter p.

Theorem 5.1.1 Expectation and Variance of Bernoulli R.V.
Suppose $X \sim \text{Bernoulli}(p)$, where $0 < p < 1$.
Then $E[X] = p$ and $Var(X) = p(1-p)$.

◎ Consider an experiment in which n independent Bernoulli trials with the same probability of success, say p, are performed.
The sample space of the experiment is
$\Omega = \{(w_1, w_2, \dots, w_n) : w_i = s \text{ or } f, i = 1,2,\dots,n\}$
and $P(\{(w_1, w_2, \dots, w_n)\}) = p^i(1-p)^{n-i}$, where $i = |\{1 \le j \le n : w_j = s\}|$.
Let X be the number of successes in the n Bernoulli trials.

$$\Rightarrow p_X(i) = \begin{cases} \binom{n}{i} p^i (1-p)^{n-i}, & \text{if } i = 0,1,2,\dots,n. \\ 0, & \text{o.w.} \end{cases}$$

Definition 5.1.3 X~binomial(n, p)

A discrete r.v. X of a probability space $(\Omega, \mathcal{A}, P)$ is called a **binomial r.v.** with parameter n and p where $n \geq 1$ and $0 < p < 1$, denoted X~binomial(n, p), if its p.m.f is given by

$$p_X(i) = \begin{cases} \binom{n}{i} p^i (1-p)^{n-i}, & \text{if } i = 0,1,2,\ldots,n. \\ 0 \quad , & \text{o.w.} \end{cases}$$

Such a p.m.f is called a binomial p.m.f with parameter n and p.

Remark 5.1.1 Bernoulli R.V. from Binomial R.V.

(1) A Bernoulli r.v. with parameter p **is** a binomial r.v. with parameter 1 and p.

(2)

$$\sum_{i=0}^{n} p_X(i) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = (p + (1-p))^n = 1$$

$\Rightarrow p_X(\cdot)$ is a p.m.f.

Theorem 5.1.2 Expectation and Variance of Binomial R.V.

Suppose X~binomial(n, p), where $n \geq 1$ and $0 < p < 1$.
Then $E[X] = np$ and $Var(X) = np(1-p)$.

Theorem 5.1.3 Maximum Point of Binomial Probability

Suppose X~binomial(n, p), where $n \geq 1$ and $0 < p < 1$.
Then

$$arg \max_{0 \leq i \leq n} p_X(i) = \begin{cases} (n+1)p - 1 \ or \ (n+1)p, & \text{if } (n+1)p \in \mathbb{Z} \\ \lfloor (n+1)p \rfloor, & \text{if } (n+1)p \notin \mathbb{Z} \end{cases}$$

§ 5.2 Poisson R.V.'s

◎ If $X \sim \text{binomial}(n,p) \Rightarrow p_X(i) = \binom{n}{i} p^i (1-p)^{n-i}$ is difficult to calculate if n is large.

◎ A recursive relation: $p_X(0) = (1-p)^n$, $p_X(i) = \frac{(n-i+1)}{i(1-p)} p_X(i-1), \forall i \geq 1$.

◎ An approximation for large n, small p, and moderate np, say $np = \lambda$ for some constant λ.

$$\Rightarrow p_X(i) = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n(n-1)\cdots(n-i+1)}{i!} (\frac{\lambda}{n})^i (1-\frac{\lambda}{n})^{n-i}$$

$$= \frac{n(n-1)\cdots(n-i+1)}{n^i} \cdot \frac{1}{(1-\frac{\lambda}{n})^i} \cdot \frac{\lambda^i}{i!} \cdot (1-\frac{\lambda}{n})^n \xrightarrow{n\to\infty} e^{-\lambda}\frac{\lambda^i}{i!}.$$

Definition 5.2.1 Poisson R.V.

A discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ is called a **Poisson r.v.** with parameter λ where $0 < \lambda < 1$, denoted $X \sim \text{Poisson}(\lambda)$, if its p.m.f is given by

$$p_X(i) = \begin{cases} e^{-\lambda}\frac{\lambda^i}{i!}, & i = 0,1,2\ldots \\ 0, & \text{o.w.} \end{cases}$$

Such a p.m.f is called a Poisson p.m.f with parameter λ.

Remark 5.2.1 Poisson R.V. from Binomial R.V.

(1) A Poisson r.v. with parameter λ is an approximation of a binomial p.m.f. with parameters n and p such that n is large and p is small, and $np = \lambda$.

(2)

$$\sum_{i=0}^{\infty} p_X(i) = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

$\Rightarrow$ $p_X(\cdot)$ is a p.m.f.

Theorem 5.2.1 Expectation and Variance of Poisson R.V.

Suppose $X \sim \text{Poisson}(\lambda)$, where $\lambda > 0$. Then $E[X] = \lambda$ and $Var(X) = \lambda$.

Theorem 5.2.2 Maximum Point of Poisson Probability

Suppose $X \sim \text{Poisson}(\lambda)$, where $\lambda > 0$. Then

$$arg\max_{i\geq 0} p_X(i) = \begin{cases} \lambda - 1 \; or \; \lambda, & \text{if } \lambda \in \mathbb{Z} \\ \lfloor \lambda \rfloor, & \text{if } \lambda \notin \mathbb{Z} \end{cases}$$

§ 5.3 Geometric R.V.'s, Negative Binomial R.V.'s and Hypergeometric R.V.'s

◎ Consider an experiment in which independent Bernoulli trials with the same probability of success, say p, are performed until the first success occurs.
The sample space of the experiment is $\Omega = \{s, fs, ffs, \dots\}$.
Let X be the number of Bernoulli trials until the first success occurs

$$\Rightarrow p_X(i) = \begin{cases} (1-p)^{i-1} \cdot p\,, & i = 0,1,2 \dots \\ 0, & \text{o.w.} \end{cases}$$

Definition 5.3.1 Geometric R.V.

A discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ is called a **geometric r.v.** with parameter p where $0 < p < 1$, denoted $X \sim \text{geometric}(p)$, if its p.m.f is given by

$$p_X(i) = \begin{cases} (1-p)^{i-1} \cdot p\,, & i = 0,1,2 \dots \\ 0, & \text{o.w.} \end{cases}$$

Such a p.m.f is called a geometric p.m.f with parameter p.

Remark 5.3.1 Justification of P.M.F.

$$\sum_{i=1}^{\infty} p_X(i) = \sum_{i=1}^{\infty} (1-p)^{i-1} \cdot p = p \cdot \frac{1}{1-(1-p)} = 1$$

$\Rightarrow$ $p_X(\cdot)$is a p.m.f.

Theorem 5.3.1 Expectation and Variance of Geometric R.V.

Suppose $X \sim \text{geometric}(p)$, where $0 < p < 1$.
Then

$$E[X] = \frac{1}{p}$$

and

$$Var(X) = \frac{1-p}{p^2}.$$

Theorem 5.3.2 Memoryless Property

Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ with $X(\Omega) = \{1,2, \dots\}$.
Then $P\big((X > m+n)\big|(X > m)\big) = P(X > n)\ \forall\, m, n > 0 \Leftrightarrow X$ is a geometric r.v.

◎ Consider an experiment in which independent Bernoulli trials with the same probability of success, say p, are performed until the r^{th} success occurs, where $r \geq 1$.
Let X be the number of Bernoulli trials until the r^{th} success occurs.

$$\Rightarrow p_X(i) = \begin{cases} \binom{i-1}{r-1} p^r (1-p)^{i-r}\,, & i = r, r+1, \dots \\ 0, & \text{o.w.} \end{cases}$$

Definition 5.3.2 Negative Binomial R.V.
A discrete r.v. of a probability space $(\Omega, \mathscr{A}, P)$ is called a **negative binomial r.v.** with parameters r and p where $r \geq 1$ and $0 < p < 1$, denoted $X\sim$neg.-binomial(r, p), if its p.m.f is given by

$$p_X(i) = \begin{cases} \binom{i-1}{r-1} p^r (1-p)^{i-r}, & i = r, r+1, \ldots \\ 0, & \text{o.w.} \end{cases}$$

Such a p.m.f is called a negative binomial p.m.f with parameters r and p.

Remark 5.3.2 Geometric R.V. from Negative Binomial R.V.
(1) A geometric r.v. with parameter p is a negative binomial r.v. with parameters 1 and p.
(2)

$$\sum_{i=r}^{\infty} (i-1)(i-2)\cdots(i-r+1)x^{i-r} = \frac{d^{r-1}}{dx^{r-1}}\left(\sum_{i=1}^{\infty} x^i\right)$$
$$= \frac{d^{r-1}}{dx^{r-1}}\left(\frac{1}{1-x}\right) = \frac{(r-1)!}{(1-x)^r}$$

$$\Rightarrow \sum_{i=r}^{\infty} p_X(i) = \sum_{i=r}^{\infty} \binom{i-1}{r-1} p^r (1-p)^{i-r} = \frac{p^r}{(r-1)!} \cdot \frac{(r-1)!}{(1-(1-p))^r} = 1$$

$\Rightarrow$ $p_X(\cdot)$ is a p.m.f.

Theorem 5.3.3 Expectation and Variance of Negative Geometric R.V.
Suppose $X\sim$neg.-binomial(r, p), where $r \geq 1$ and $0 < p < 1$.
Then

$$E[X] = \frac{r}{p}$$

and

$$Var(X) = \frac{r(1-p)}{p^2}.$$

Theorem 5.3.4 Maximum Point of Negative Geometric Probability
Suppose $X\sim$neg.-binomial(r, p), where $r \geq 1$ and $0 < p < 1$. Then

$$arg\max_{i \geq r} p_X(i) = \begin{cases} 1, & \text{if } r = 1 \\ \frac{r-1}{p} \text{ or } \frac{r-1}{p} + 1, & \text{if } \frac{r-1}{p} \in \mathbb{Z}^+ \\ \left\lfloor \frac{r-1}{p} \right\rfloor + 1, & \text{if } \frac{r-1}{p} \notin \mathbb{Z} \end{cases}$$

◎ A box contains N_1 red balls and N_2 blue balls.
Suppose that n balls are randomly drawn from the box, one by one and without replacement.
Let X be the number of "red" balls drawn

$$\Rightarrow p_X(i) = \begin{cases} \dfrac{\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}}, & i = a, a+1, \dots, b.\ a = \max\{n - N_1, 0\}, b = \min\{n, N_1\} \\ 0, & \text{o.w.} \end{cases}$$

Definition 5.3.3 Hypergeometric R.V.

A discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$ is called a **hypergeometric r.v.** with parameter N_1, N_2 and n where $N_1, N_2 \geq 1$ and $n \geq 1$,
denoted $X\sim$ hypergeometric(N_1, N_2, n), if its p.m.f is given by

$$p_X(i) = \begin{cases} \dfrac{\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}}, & i = a, a+1, \dots, b.\ a = \max\{n - N_1, 0\}, b = \min\{n, N_1\} \\ 0, & \text{o.w.} \end{cases}$$

Such a p.m.f is called a hypergeometric r.v. with parameter N_1, N_2 and n.

Remark 5.3.3 Justification of P.M.F.

(1) If $n \leq \min\{N_1, N_2\} \Rightarrow a = \max\{n - N_1, 0\} = 0, b = \min\{n, N_1\} = n$.
(2) $(1+x)^{N_1+N_2} = (1+x)^{N_1}(1+x)^{N_2}$
$\Rightarrow$ the coefficient of x^n is

$$\binom{N_1+N_2}{n} = \sum_{i=a}^{b} \binom{N_1}{i}\binom{N_2}{n-i},$$

where $a = \max\{n - N_1, 0\}$, $b = \min\{n, N_1\}$.

$$\Rightarrow \sum_{i=a}^{b} p_X(i) = \sum_{i=a}^{b} \frac{\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}} = 1.$$

$\Rightarrow$ $p_X(\cdot)$ is a p.m.f.

Theorem 5.3.5 Expectation and Variance of Hypergeometric R.V.

Suppose $X\sim$ hypergeometric(N_1, N_2, n),
where $N_1, N_2 \geq 1$ and $1 \leq n \leq \min\{N_1, N_2\}$.
Then

$$E[X] = \frac{nN_1}{N_1 + N_2}$$

and

$$Var(X) = n \cdot \frac{N_1}{N_1 + N_2} \cdot \frac{N_2}{N_1 + N_2} \cdot \left(1 - \frac{n-1}{N_1 + N_2 - 1}\right).$$

Remark 5.3.4 Binomial Approximation for Hypergeometric

n balls are drawn with replacement

$\Rightarrow$ $X \sim \text{binomial}\left(n, \frac{N_1}{N_1+N_2}\right)$

$$\Rightarrow \; E[X] = n \cdot \frac{N_1}{N_1+N_2}, \qquad Var(X) = n \cdot \frac{N_1}{N_1+N_2} \cdot \frac{N_2}{N_1+N_2}.$$

Therefore, if $n \ll N_1 + N_2$, then drawing with replacement is a good approximation of drawing without replacement.

Theorem 5.3.6 Maximum Point of Hypergeometric Probability

Suppose $X \sim$ hypergeometric(N_1, N_2, n),
where $N_1, N_2 \geq 1$ and $1 \leq n \leq \min\{N_1, N_2\}$.

Then

$$arg \max_{0 \leq i \leq n} p_X(i)$$

$$= \begin{cases} \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} - 1 \text{ or } \dfrac{(n+1)(N_1+1)}{N_1+N_2+2}, & \text{if } \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} \in \mathbb{Z} \\ \left\lfloor \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} \right\rfloor, & \text{if } \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} \notin \mathbb{Z} \end{cases}$$

Remark 5.3.5 Binomial and Poisson Approximation for Hypergeometric

$$p_X(i) = \frac{\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}}$$

$$= \frac{n!}{i!\,(n-i)!} \cdot \frac{N_1(N_1-1)\cdots(N_1-i+1)N_2(N_2-1)\cdots(N_2-n+i+1)}{(N_1+N_2)(N_1+N_2-1)\cdots(N_1+N_2+n-1)}$$

(1) If $N_1 \to \infty, N_2 \to \infty, \frac{N_1}{N_1+N_2} \to p$, then

$$p_X(i) = \binom{n}{i} \frac{\frac{N_1}{N_1+N_2}\left(\frac{N_1}{N_1+N_2}-\frac{1}{N_1+N_2}\right)\cdots\left(\frac{N_1}{N_1+N_2}-\frac{i-1}{N_1+N_2}\right)\left(\frac{N_2}{N_1+N_2}\right)\left(\frac{N_2}{N_1+N_2}-\frac{1}{N_1+N_2}\right)\cdots\left(\frac{N_2}{N_1+N_2}-\frac{n-i-1}{N_1+N_2}\right)}{1\cdot\left(1-\frac{1}{N_1+N_2}\right)\cdots\left(1-\frac{n-1}{N_1+N_2}\right)}$$

$$\xrightarrow{N_1,N_2\to\infty} \binom{n}{i} p^i (1-p)^{n-i} \leftarrow \text{binomial}(n,p)$$

(2) If $n \to \infty, N_1 \to \infty, N_2 \to \infty, \frac{n}{N_1+N_2} \to 0, \frac{N_1}{N_1+N_2} \to \frac{\lambda}{n}$, then

$$p_X(i) = \frac{1}{i!} \frac{nN_1\cdot(n-1)(N_1-1)\cdots(n-i+1)(N_1-i+1)\cdot(N_1+N_2-N_1)(N_1+N_2-N_1-1)\cdots(N_1+N_2-N_1-n+i+1)}{\frac{(N_1+N_2)!}{(N_1+N_2-n)!}}$$

$$= \frac{1}{i!} \frac{\prod_{j=0}^{i-1}\frac{nN_1-j(n+N_1)+j^2}{N_1+N_2}\prod_{j=0}^{n-i-1}\left(1-\frac{N_1+j}{N_1+N_2}\right)}{\frac{1}{(N_1+N_2)^n}\cdot\frac{\sqrt{2\pi(N_1+N_2)}\left(\frac{N_1+N_2}{e}\right)^{N_1+N_2} e^{a_{N_1+N_2}}}{\sqrt{2\pi(N_1+N_2-n)}\left(\frac{N_1+N_2-n}{e}\right)^{N_1+N_2-n} e^{a_{N_1+N_2-n}}}}$$

where $a_n = \ln\frac{n!}{\sqrt{2\pi n}(\frac{n}{e})^n} \xrightarrow{n\to\infty} 0.$

$$p_X(i) \xrightarrow{n,N_1,N_2\to\infty,\frac{n}{N_1+N_2}\to 0,\frac{N_1}{N_1+N_2}\to\frac{\lambda}{n}} \frac{1}{i!}\lim_{n\to\infty}\frac{\lambda^i\left(1-\frac{\lambda}{n}\right)^{n-i}}{\frac{1}{e^n \lim\limits_{N_1,N_2\to\infty}\left(1-\frac{n}{N_1+N_2}\right)^{N_1+N_2-n}}}$$

$$= \lim_{n\to\infty}\frac{\lambda^i}{i!}\left(1-\frac{\lambda}{n}\right)^{n-i} = e^{-\lambda}\frac{\lambda^i}{i!} \leftarrow \text{Poisson}(\lambda)$$

Chapter.6 Continuous Random Variables

§ 6.1 Probability Density Function

Definition 6.1.1 Probability Density Function
Let X be a r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
X is called an absolutely continuous (or a continuous) r.v. if there exists a nonnegative real-valued function $f_X: \mathbb{R} \to [0, \infty)$ s.t.

$$P(x \in B) = \int_B f_X(x)dx, \forall B \in \mathcal{B}_{\mathbb{R}}.$$

The function f_X is called the **probability density function** (p.d.f.) of X.

Remark 6.1.1 Approximation of Probability

$$P(a \le X \le a + \delta) = \int_a^{a+\delta} f_X(x)dx = f_X(a_\delta) \cdot \delta,$$

for some $a_\delta \in [a, a + \delta]$.

If f_X is **continuous** at a

$$\Rightarrow \lim_{\delta \to 0} \frac{P(a \le X \le a + \delta)}{\delta} = \lim_{\delta \to 0} f_X(a_\delta) = f_X(a).$$

So $P(a \le X \le a + \delta) \approx f_X(a_\delta) \cdot \delta$, if f_X is continuous at a and δ is very small.

Theorem 6.1.1 C.D.F and Probability from P.D.F.
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
(1)

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

Therefore, $F_X(x)$ is a **continuous** function.

(2)

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

(3) If f_X is continuous at a, then $F_X'(a) = f_X(a)$.
Therefore, if f_X is a continuous function, then $F_X'(x) = f_X(x), \forall x \in \mathbb{R}$.

(4) $P(X = a) = 0, \forall a \in \mathbb{R}$. Therefore,

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b) = \int_a^b f_X(x)dx.$$

Theorem 6.1.2 Existence of P.D.F.

Suppose $f: \mathbb{R} \to [0, \infty)$ is a **nonnegative** real-valued function s.t.

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Then there exists a continuous r.v. X of some probability space $(\Omega, \mathcal{A}, \mathrm{P})$ s.t. the p.d.f. is equal to f.

Definition 6.1.2 Sufficient Conditions of P.D.F.

A **nonnegative** real-valued function $f: \mathbb{R} \to [0, \infty)$ s.t.

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

is called a p.d.f.

The c.d.f. $F: \mathbb{R} \to [0,1]$ associated with f is given by

$$F(t) = \int_{-\infty}^{t} f(x)dx, \forall t \in \mathbb{R}.$$

Remark 6.1.2 Neither Discrete Nor Continuous R.V.

There are r.v.'s that are neither discrete nor continuous,

e.g., $F_X(x) = \alpha F_d(x) + (1-\alpha)F_c(x)$, where $0 < \alpha < 1$.

§ 6.2 The Probability Density Function of A Function of A R.V.

Theorem 6.2.1 Method of Distribution Functions

Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
If $Y = g(X)$, then

$$f_Y(y) = \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}[P(Y \le y)] = \frac{d}{dy}\big[P[g(X) \le y]\big]$$
$$\to \frac{d}{dy}[X \sim g^{-1}(y)] \to \frac{d}{dy}\big[F_X\big(g^{-1}(y)\big)\big] \to \frac{d}{dy}[g^{-1}(y)] \cdot f_X\big(g^{-1}(y)\big).$$

Theorem 6.2.2 Method of Transformations

Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$
such that its p.d.f. is continuous.
Suppose $Y = g(X)$, where g is a measurable function from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

(1) If $g(X)$ is a **discrete** r.v., then

$$P_Y(y) = \int_{x:g(x)=y} f_X(x)dx, \forall y \in g[X(\Omega)].$$

(2) If $g(X)$ is a **continuous** r.v., $g'(x)$ exists,
and $g'(x) \neq 0, \forall x \in g^{-1}(\{y\}): \{x: g(x) = y\}$, where $y \in g[X(\Omega)]$.
Then,

$$f_Y(y) = \sum_{x:g(x)=y} \frac{f_X(x)}{|g'(x)|}.$$

§ 6.3 Expectations and Variances

Definition 6.3.1 Expectation

Let X be a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$ s.t. its p.d.f. is continuous.
The **expectation** (or mean) of X is given by

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

if $x f_X(x)$ is **absolutely integrable**, i.e.,

$$\int_{-\infty}^{\infty} |x f_X(x)| dx < +\infty,$$

and is given by $E[X] = \pm\infty$, if the integration diverges to $\pm\infty$.

Remark 6.3.1 Necessary and Sufficient Condition of Absolutely Integrable

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{0}^{\infty} x f_X(x) dx - \int_{-\infty}^{0} (-x) f_X(x) dx$$

$$\Rightarrow E[|X|] = \int_{0}^{\infty} x f_X(x) dx + \int_{-\infty}^{0} (-x) f_X(x) dx$$

$$\therefore\ E[|X|] < \infty \ \Leftrightarrow \ \int_{0}^{\infty} x f_X(x) dx < \infty \ \text{ and } \int_{-\infty}^{0} (-x) f_X(x) dx < \infty.$$

Theorem 6.3.1 Calculation of Expectation

Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
Then

$$E[X] = \int_{0}^{\infty} P(x > t)\, dt - \int_{0}^{\infty} P(x \leq -t)\, dt = \int_{0}^{\infty} \big(1 - F_X(t)\big) dt - \int_{0}^{\infty} \big(F_X(-t)\big) dt.$$

Corollary 6.3.1 Calculation of r^{th} Moment

Suppose X is a **nonnegative** continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$,
and $r > 0$.
Then

$$E[X^r] = \int_{0}^{\infty} r t^{r-1} P(x > t) dt = \int_{0}^{\infty} r t^{r-1} \big(1 - F_X(t)\big) dt.$$

In particular,

$$E[X] = \int_{0}^{\infty} P(x > t)\, dt = \int_{0}^{\infty} \big(1 - F_X(t)\big) dt.$$

Theorem 6.3.2　Approximation of Expectation
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
Then

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E[|X|] \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

Therefore,

$$E[|X|] < \infty \Leftrightarrow \sum_{n=1}^{\infty} P(|X| \geq n) \leq \infty.$$

Theorem 6.3.3　Infinite Zero
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
Then,

$$E[X] < \infty \Rightarrow \lim_{x\to\infty} x \cdot P(X > x) = \lim_{x\to-\infty} x \cdot P(X \leq x) = 0.$$

Theorem 6.3.4　Expectation of Measurable Function
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$,
and suppose g is a **measurable function** from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
Then

$$E[g(X)] = \int_{-\infty}^{\infty} g(X) \cdot f_X(x) dx.$$

Corollary 6.3.2　Expectation of Linear Combination of Measurable Functions
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
$g_1, g_2, \dots g_n$ are **measurable functions** from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,
and $\alpha_1, \alpha_2, \dots \alpha_n \in \mathbb{R}$.
Then

$$E\left[\sum_{i=1}^{n} \alpha_i\, g_i(X)\right] = \sum_{i=1}^{n} \alpha_i\, E[g_i(X)]$$

Definition 6.3.2　Variance and Standard Deviation
Let X be a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $E[X]$ exists.
The **variance** of X is given by $Var(X) = E[(X - E[X])^2]$.

And the **standard deviation** of X is given by $\sigma_X = \sqrt{Var(X)} = \sqrt{E[(X - E[X])^2]}$.

Theorem 6.3.5　Minimum Distance with Expectation
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$,
and suppose $E[X]$ exists.

If $E[X^2] < +\infty$, then $Var(X) = \min_{a\in\mathbb{R}} E[(X - a)^2]$.

Theorem 6.3.6 Calculation of Linear Combination
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$,
and suppose $E[X]$ exists.
Then
(1)
$$Var(X) = E[X^2] - (E[X])^2$$
(2)
$$Var(aX + b) = a^2 Var(X)$$
and
$$\sigma_{aX+b} = |a|\sigma_X, \forall a, b \in \mathbb{R}.$$

Definition 6.3.3 Moment and Absolute Moment
Let X be a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$, and $r, c \in \mathbb{R}$.

$$\begin{cases} \text{The } r^{th} \textbf{ moment} \text{ of } X \text{ is given by } E[X^r] \\ \text{The } r^{th} \textbf{ central moment} \text{ of } X \text{ is given by } E[(X - E[X])^r] \\ \text{The } r^{th} \textbf{ moment} \text{ of } c \text{ is given by } E[(X - c)^r] \end{cases}$$

$$\begin{cases} \text{The } r^{th} \textbf{ absolute moment} \text{ of } X \text{ is given by } E[|X|^r] \\ \text{The } r^{th} \textbf{ absolute central moment} \text{ of } X \text{ is given by } E[|X - E[X]|^r] \\ \text{The } r^{th} \textbf{ absolute moment} \text{ of } c \text{ is given by } E[|X - c|^r] \end{cases}$$

If the respective sum **converges absolutely**.

Theorem 6.3.7 Existence of Lower Order Moment
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $0 < r < s$.
If $E[|X|^s]$ exists, then $E[|X|^r]$ exists.
That is, the existence of a higher order moment of X **guarantees the existence** of a lower order moment of X.

Theorem 6.3.8 Positive Variance
Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
Then
$$E[(X - a)^2] > 0, \forall a \in \mathbb{R}.$$
Therefore
$$E[X] \text{ exists} \Rightarrow Var(X) > 0.$$

Chapter.7 Special Continuous Distributions

§ 7.1 Uniform R.V.'s

Definition 7.1.1 Uniform R.V.
A continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$ is called a **uniform r.v.** over (α, β), where $\alpha, \beta \in \mathbb{R}$ and $\alpha < \beta$, denoted $X \sim U(\alpha, \beta)$, if its p.d.f. is given by

$$f_X(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \text{if } \alpha < x < \beta \\ 0, & \text{o.w.} \end{cases}$$

Remark 7.1.2 P.D.F. and C.D.F.
(1) $f_X(x) \geq 0, \forall x \in \mathbb{R}$, and

$$\int_{-\infty}^{\infty} f_X(x)\, dx = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} dx = 1$$

$\Rightarrow f_X(x)$ is a p.d.f.

(2)

$$F_X(x) = \begin{cases} 0, & \text{if } x \leq \alpha \\ \dfrac{x - \alpha}{\beta - \alpha}, & \text{if } \alpha < x < \beta \\ 1, & \text{if } x \geq \beta \end{cases}$$

Theorem 7.1.1 Expectation and Variance of Uniform R.V.
Suppose $X \sim U(\alpha, \beta)$, where $\alpha, \beta \in \mathbb{R}$ and $\alpha < \beta$.
Then

$$E[X^n] = \frac{\sum_{i=1}^{n} \alpha^{n-i} \beta^i}{n+1}.$$

Therefore

$$E[X] = \frac{\alpha + \beta}{2}$$

and

$$Var(X) = \frac{(\beta - \alpha)^2}{12}.$$

Remark 7.1.2 Expectation and Variance of Discrete "Uniform R.V."

Suppose $X \sim \text{Uniform}(1,2,\dots,n)$, where $n \geq 1$.

Then

$$E[X] = \frac{n+1}{2}, E[X^2] = \frac{(n+1)(2n+1)}{6}$$

and

$$Var(X) = \frac{n^2-1}{12}.$$

Theorem 7.1.2 Linear Generated R.V.

Suppose $X \sim U(\alpha, \beta)$, where $\alpha, \beta \in \mathbb{R}$ and $\alpha < \beta$.

Suppose $Y = aX + b$, where $\alpha, \beta \in \mathbb{R}$ and $a \neq 0$.

Then

$$Y \sim \begin{cases} U(a\alpha + b, a\beta + b), & \text{if } a > 0 \\ U(a\beta + b, a\alpha + b), & \text{if } a < 0 \end{cases}$$

§ 7.2 Normal (Gaussian) R.V.'s

Definition 7.2.1 Normal (Gaussian) R.V.
A continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$ is called a **normal (Gaussian)** r.v. with parameters μ and σ^2, where $\mu, \sigma \in \mathbb{R}$, $\sigma \neq 0$, denoted $X \sim N(\mu, \sigma^2)$, if its p.d.f. is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty.$$

Remark 7.2.1 P.D.F. and C.D.F.

(1) $f_X(x) \geq 0, \forall x \in \mathbb{R}$, and let $I = \int_{-\infty}^{\infty} e^{-ax^2}\, dx$.

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a(x^2+y^2)} dxdy \\ &\xrightarrow{x=rcos\theta, y=rsin\theta} \int_0^{\infty}\int_0^{2\pi} e^{-ar^2} rdrd\theta = \frac{\pi}{a} \\ &\Rightarrow I = \sqrt{\frac{\pi}{a}} \Rightarrow \int_{-\infty}^{\infty} \sqrt{\frac{a}{\pi}} e^{-ax^2}\, dx = 1 \end{aligned}$$

$$\therefore \int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} = 1$$

$\Rightarrow f_X(x)$ is a p.d.f.

(2) If $\mu = 0$, $\sigma^2 = 1$, then X is called a **standard** normal (Gaussian) r.v.

(3)

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy \\ &\xrightarrow{y=\sigma t+\mu} \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt \\ &= \Phi(\frac{x-\mu}{\sigma}) \end{aligned}$$

where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Theorem 7.2.1 Symmetric about $\boldsymbol{\mu}$
Suppose $X \sim N(\mu, \sigma^2)$.

(1) $f_X(x)$ is **symmetric** about $x = \mu$, with maximum at $x = \mu$, and **inflection** points at $x = \mu \pm \sigma$.

(2) $\Phi(-y) = 1 - \Phi(y), \forall y \in \mathbb{R}$ and $\Phi(0) = 1$.

Therefore, $F_X(\mu - y) = 1 - F_X(\mu + y)$ and $F_X(\mu) = \frac{1}{2}$.

Theorem 7.2.2 Linear Generated R.V.
Suppose $X \sim N(\mu, \sigma^2)$, where $\mu, \sigma \in \mathbb{R}$, $\sigma \neq 0$.
Suppose $Y = aX + b$, where $\alpha, \beta \in \mathbb{R}$ and $a \neq 0$.
Then,

$$Y \sim N(a\mu + b, a^2\sigma^2).$$

In particular, if

$$Y = \frac{x - \mu}{\sigma},$$

then

$$Y \sim N(0,1).$$

Definition 7.2.2 Gamma Function
The function $\Gamma: (0, \infty) \to \mathbb{R}$ given by

$$\Gamma(\alpha) = \int_0^\infty e^{-t}\, t^{\alpha-1} dt, \forall \alpha > 0$$

is called the **gamma function**.

Theorem 7.2.3 Properties of Gamma Function
(1)

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \forall \alpha > 0.$$

(2)

$$\Gamma(n + 1) = n!, \forall n \geq 0.$$

(3)

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!}{2^{2n} n!}\sqrt{\pi}, \forall n \geq 0.$$

Theorem 7.2.4 Calculation of Moment and Absolute Moment
Suppose $X \sim N(\mu, \sigma^2)$, where $\mu, \sigma \in \mathbb{R}$, $\sigma \neq 0$.
(1)

$$E[|x - \mu|^n] = \frac{(2\sigma^2)^{\frac{n}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right) = \begin{cases} \dfrac{2^{k+1} \cdot k!}{\sqrt{2\pi}} \sigma^{2k+1}, \text{if } n = 2k+1, k \geq 0 \\ \dfrac{(2k)!}{2^k \cdot k!} \sigma^{2k} \quad, \text{if } n = 2k \quad, k \geq 0 \end{cases}$$

(2)

$$E[(x - \mu)^n] = \begin{cases} 0 \quad, \text{if } n = 2k+1, \quad k \geq 0 \\ \dfrac{(2k)!}{2^k \cdot k!} \sigma^{2k}, \text{ if } n = 2k, \quad k \geq 0 \end{cases}$$

(3)

$$E[x^n] = \sum_{k=0}^{n} \binom{n}{k} E\left[(x - \mu)^k\right] \cdot \mu^{n-k}.$$

Theorem 7.2.5 De Moivre-Laplace Theorem

Suppose $X \sim \text{binomial}(n, p)$, where $n \geq 1$ and $0 < p < 1$.

Then

$$\lim_{n\to\infty} P\left(a < \frac{X - np}{\sqrt{np(1-p)}} < b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx, \forall a, b \in \mathbb{R}, a < b.$$

Theorem 7.2.6 Approximation of $\boldsymbol{\Phi(x)}$

$$\frac{1}{\sqrt{2\pi}x}\left(1 - \frac{1}{x^2}\right) e^{-\frac{x^2}{2}} < 1 - \Phi(x) < \frac{1}{\sqrt{2\pi}x} \cdot e^{-\frac{x^2}{2}}, \forall x > 0.$$

Theorem 7.2.7 Expectation of Exponential Function

Suppose $X \sim N(\mu, \sigma^2)$, where $\mu, \sigma \in \mathbb{R}$, $\sigma \neq 0$, and $\alpha \in \mathbb{R}$.

Then

$$E[e^{\alpha x}] = e^{\alpha\mu + \frac{1}{2}\alpha^2\sigma^2}.$$

§ 7.3 Gamma R.V.'s, Erlang R.V.'s and Exponential R.V.'s

Definition 7.3.1 Gamma R.V., Erlang R.V. and Exponential R.V.

A continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$ is called a **gamma** r.v. with parameters α and λ, where $\alpha, \lambda > 0$, denoted $X \sim g(\alpha, \lambda)$, if its p.d.f. is given by

$$f_X(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } \alpha > 0 \\ 0, & \text{o.w.} \end{cases}$$

If $\alpha = n, n \geq 1$, then X is called an **Erlang** r.v. with parameters n and λ, denoted $X \sim \mathcal{E}(n, \lambda)$.

If $\alpha = 1$, then X is called an **exponential** r.v. with parameters λ, denoted $X \sim \mathcal{E}(\lambda)$.

Remark 7.3.1 Properties of P.D.F.

(1)

$$\int_{-\infty}^{\infty} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} dx \xrightarrow{t=\lambda x} \int_{0}^{\infty} \frac{e^{-t} t^{\alpha-1}}{\Gamma(\alpha)} dt = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1$$

$\Rightarrow f_X(x)$ is a p.d.f.

(2)

$$f_X'(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x}(-\lambda x^{\alpha-1} + (\alpha-1)x^{\alpha-2})$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x} \cdot x^{\alpha-2}\big(-\lambda x + (\alpha-1)\big)$$

$$f_X''(x)$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x}[-\lambda^2 x^{\alpha-1} - \lambda(\alpha-1)x^{\alpha-2} - \lambda(\alpha-1)x^{\alpha-2} + (\alpha-2)(\alpha-1)x^{\alpha-3}]$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x} \cdot x^{\alpha-3}\left[\big(\lambda x - (\alpha-1)\big)^2 - (\alpha-1)\right]$$

$\therefore 0 < \alpha \leq 1 \Rightarrow f_X'(x) < 0, f_X''(x) > 0, \forall x > 0.$

$$\alpha > 1 \Rightarrow f_X'(x) \begin{cases} > 0 \Leftrightarrow x < \dfrac{\alpha-1}{\lambda} \\ = 0 \Leftrightarrow x = \dfrac{\alpha-1}{\lambda} \\ < 0 \Leftrightarrow x > \dfrac{\alpha-1}{\lambda} \end{cases}$$

and

$$f_X''(x) \begin{cases} > 0 \Leftrightarrow x > \dfrac{\alpha-1}{\lambda} + \dfrac{\sqrt{\alpha-1}}{\lambda} \text{ or } x < \dfrac{\alpha-1}{\lambda} - \dfrac{\sqrt{\alpha-1}}{\lambda} \\ = 0 \Leftrightarrow x = \dfrac{\alpha-1}{\lambda} \pm \dfrac{\sqrt{\alpha-1}}{\lambda} \\ < 0 \Leftrightarrow \dfrac{\alpha-1}{\lambda} - \dfrac{\sqrt{\alpha-1}}{\lambda} < x < \dfrac{\alpha-1}{\lambda} + \dfrac{\sqrt{\alpha-1}}{\lambda} \end{cases}$$

Theorem 7.3.1 Calculation of C.D.F.
Suppose $X \sim g(\alpha, \lambda)$, where $\alpha, \lambda > 0$.
Then

$$F_X(x) = 1 - \frac{\Gamma(\alpha, \lambda x)}{\Gamma(\alpha)},$$

where

$$\Gamma(\alpha, x) = \int_x^\infty e^{-t} t^{\alpha-1} dt$$

is the **incomplete** gamma function.
If $\alpha = n \geq 1$, then

$$F_X(x) = 1 - \sum_{i=0}^{n-1} \frac{e^{-\lambda x}(\lambda x)^i}{i!} = P(N \geq n)$$

where $N \sim \text{Poisson}(n\lambda)$.

Theorem 7.3.2 Expectation and Variance of Gamma R.V.
Suppose $X \sim g(\alpha, \lambda)$, where $\alpha, \lambda > 0$.
Then

$$E[X^n] = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)\lambda^n} = \frac{\binom{n+\alpha-1}{n}}{\lambda^n} = \frac{(\alpha)_n}{\lambda^n}$$

where

$$(\alpha)_n = \binom{n + \alpha - 1}{n} = (n + \alpha - 1) \cdots (\alpha - 1) \cdot \alpha$$

Therefore,

$$E[X] = \frac{\alpha}{\lambda} \text{ and } Var(X) = \frac{\alpha}{\lambda^2}.$$

Theorem 7.3.3 Linear Generated Gamma R.V.
Suppose $X \sim g(\alpha, \lambda)$, where $\alpha, \lambda > 0$, and $Y = aX$, where $a > 0$. Then

$$Y \sim g\left(\alpha, \frac{\lambda}{a}\right).$$

Theorem 7.3.4 Gamma R.V. from Normal R.V.
Suppose $X \sim N(\mu, \sigma^2)$, where $\mu, \sigma \in \mathbb{R}$, $\sigma \neq 0$ and $Y = (X - \mu)^2$. Then

$$Y \sim g\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right).$$

Lemma 7.3.1 Plus to Multiply Property of Exponential Function
Suppose $f: [0, +\infty) \to \mathbb{R}$ is **right continuous** on $[0, +\infty)$
and $f(x + y) = f(x) \cdot f(y), \forall x, y \geq 0$.
Then there either $f(x) = 0, \forall x \geq 0$ or $\exists \lambda \in \mathbb{R}$ s.t. $f(x) = e^{-\lambda x}, \forall x \geq 0$.

Theorem 7.3.5 Memoryless Property
Suppose X is a **nonnegative** continuous r.v. of a probability space $(\Omega, \mathcal{A}, P)$.
Then $P(x > s + t | x > s) = P(x > t)\ \forall s, t > 0 \Leftrightarrow X \sim \mathcal{E}(\lambda)$, for some $\lambda > 0$.

Remark 7.3.2 Analog of Geometric R.V.
Exponential r.v.'s are the **continuous analog** of geometric r.v.'s.

Theorem 7.3.6 Geometric R.V. from Exponential R.V.
Suppose $X \sim \mathcal{E}(\lambda)$ where $\lambda > 0$ and $Y = \lceil X \rceil$. Then $Y \sim \text{geometric}(1 - e^{-\lambda})$.

Definition 7.3.2 Independent Set
A set of r.v.'s $\{X_i : i \in I\}$ of a probability space $(\Omega, \mathcal{A}, P)$ is called **independent**,

if for **any finite subset** $\{X_{i_1}, X_{i_2}, \dots, X_{i_k}\}, k \geq 2$ of $\{X_i : i \in I\}$ the events

$$X_{i_1} \in B_1, X_{i_2} \in B_2, \dots, X_{i_k} \in B_k$$

are independent for all $B_1, B_2, \dots, B_k \in \mathcal{B}_{\mathbb{R}}$.
Otherwise, $\{X_i : i \in I\}$ is called dependent.

Definition 7.3.3 Continuous R.Vect.
A r.vect. $\underline{X} = (X_1, X_2, \dots, X_n)$ of a probability space $(\Omega, \mathcal{A}, P)$ is called an absolute continuous **r.vect.** (or continuous r.vect.) if there exists a nonnegative real-valued function $f_{\underline{X}} : \mathbb{R}^n \to [0, \infty)$ s.t.

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_k) = \int_{B_1} \int_{B_2} \cdots \int_{B_n} f_{\underline{X}}(\underline{x}) dx_n \cdots dx_2 dx_1$$

for all $B_1, B_2, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$.

Then the function $f_{\underline{X}}$ is called the **p.d.f.** of the r.vect. $\underline{X}$,

or the joint p.d.f. of the r.v.'s $X_1, X_2, \dots, X_n$.

Theorem 7.3.7 P.D.F. and C.D.F. of Continuous R.Vect.
Suppose $\underline{X} = (X_1, X_2, \dots, X_n)$ is a continuous r.vect. and

$$F_{\underline{X}}(\underline{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n).$$

Then

$$f_{\underline{X}}(\underline{x}) = \frac{\partial F_{\underline{X}}(\underline{x})}{\partial x_1 \cdots \partial x_n}.$$

Furthermore, if $X_1, X_2, \dots, X_n$ are independent, then

$$f_{\underline{X}}(\underline{x}) = f_{X_1}(x) f_{X_2}(x) \cdots f_{X_n}(x).$$

Theorem 7.3.8 Convolution Theorem
If $\underline{X} = (X_1, X_2)$ is a continuous r.vect. and $Y = X_1 + X_2$. Then

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x, y - x)\, dx.$$

Furthermore, if $X_1 \perp X_2$, then

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(x)\, f_{X_2}(y - x) dx.$$

Definition 7.3.4 Beta Function
The function $B: \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ is given by

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}\, (1 - x)^{\beta-1} dx, \forall \alpha, \beta > 0$$

is called **beta function**.

Lemma 7.3.2 Calculation of Beta Function

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)}, \forall \alpha, \beta > 0.$$

Theorem 7.3.9 Independent Additivity of Gamma R.V.
Suppose $X_i \sim g(\alpha_i, \lambda)$ where $\alpha_i, \lambda > 0, i = 1,2, \dots, n,$ $X_1, X_2, \dots, X_n$ are **independent**, and $Y = X_1 + X_2 + \cdots + X_n$. Then

$$Y \sim g\left(\sum_{i=1}^{n} \alpha_i\, , \lambda\right).$$

Theorem 7.3.10 Independent Minimum of Exponential R.V.
Suppose $X_i \sim \mathcal{E}(\lambda_i)$ where $\lambda_i > 0, i = 1,2, \dots, n$, and $X_1, X_2, \dots, X_n$ are **independent**.
(1) If $Y = \min\{X_1, X_2, \dots, X_n\}$, then

$$Y \sim \mathcal{E}\left(\sum_{i=1}^{n} \lambda_i\right).$$

(2)

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

Definition 7.3.5 Stochastic Process
A **stochastic process** (s.p.) $\{X(t): i \in I\}$ is a collection of r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathrm{P})$.
If $I = \{0,1,2, \dots\}$ or $\{0, \pm 1, \pm 2, \dots\}$, then we call $\{X(t): i \in I\}$ a **discrete-time** S.P.
If $I = [0, \infty)$ or $(-\infty, \infty)$, then we call $\{X(t): i \in I\}$ a **continuous-time** S.P.

Definition 7.3.6 Counting Process and Poisson Process
Let $\{T_1, T_2, ...\}$ be a **discrete-time** S.P. s.t. $T_i, i = 1,2, ...$, is the time of occurrence of the i^{th} event, and $0 < T_1 < T_2 < \cdots$.

Let $X_i = T_i - T_{i-1}, i = 1,2, ...$, where $T_0 = 0$ be the interoccurrence time between the $(i-1)^{th}$ and the i^{th} events,
and $N(t) = |\{i: 0 < T_i \le t\}|$ be the number of events occurring in $(0, t]$,
so that $\{N(t): 0 < t < \infty\}$ is called the **counting process** of the S.P. $\{T_1, T_2, ...\}$.

Then we call $\{T_1, T_2, ...\}$ a **Poisson process** with rate λ,
if $X_1, X_2, ...$ are **independent and identically distributed** (i.i.d.)
and $N(t) \sim$Poisson(λt).

Theorem 7.3.11 Necessary and Sufficient Condition of Poisson Process
Suppose $\{T_1, T_2, ...\}$ is a S.P. s.t. $0 < T_1 < T_2 < \cdots$ and its interoccurrence times $X_i = T_i - T_{i-1}, i = 1,2, ...$ are i.i.d., where $T_0 = 0$.
Then $\{T_1, T_2, ...\}$ is a Poisson process with rate λ $\Leftrightarrow$ $X_i \sim \mathcal{E}(\lambda), i = 1,2, ...$

Remark 7.3.3 Negative Binomial ↔ Geometric vs Gamma ↔ Exponential
(1) A negative binomial r.v. $T_r = X_1 + X_2 + \cdots + X_r \sim$ neg.-binomial(r, p) is the number of i.i.d. Bernoulli trials with the same probability of success p until the r^{th} success occurs, where $X_i \sim$ geometric(p) is the number of Bernoulli trials between the $(i-1)^{th}$ and the i^{th} successes, and $X_1, X_2, ...$ are independent.

(2) A gamma r.v. $T_n = X_1 + X_2 + \cdots + X_n \sim g(n, \lambda)$ is the time of occurrence of the n^{th} event of a Poisson process with rate λ, where $X_i \sim \mathcal{E}(\lambda)$ is the interoccurrence time between the $(i-1)^{th}$ and the i^{th} events, and $X_1, X_2, ...$ are independent.

Theorem 7.3.12 Merging and Splitting of Poisson Process
(1) Suppose that k independent Poisson processes with rates $\lambda_1, \lambda_2, ..., \lambda_k$ are merged into a S.P. $\{T_1, T_2, ...\}$.
Then $\{T_1, T_2, ...\}$ is a **Poisson process** with rate $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_k$.

(2) Suppose that in a Poisson process with rate λ, an event is a type-i event with probability $P_i, i = 1, 2, ..., k$.
Then the S.P. $\{T_1, T_2, ...\}$ of the times of the occurrences of the type-i events is a **Poisson process** with rate $\lambda \cdot P_i, i = 1, 2, ..., k$.

§ 7.4 Beta R.V.'s

Definition 7.4.1 Beta R.V.

A continuous r.v. X of a probability space $(\Omega, \mathcal{A}, P)$ is called a beta r.v. with parameter α and β, where $\alpha, \beta > 0$, denoted $X \sim \mathcal{B}(\alpha, \beta)$, if its p.d.f. is given by

$$f_X(x) = \begin{cases} \dfrac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{if } 0 < x < 1 \\ 0, & \text{o.w.} \end{cases}$$

where

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

Remark 7.4.1 P.D.F. and C.D.F.

(1) $\int_{-\infty}^{\infty} f_X(x)dx = 1 \Rightarrow f_X(x)$ is a p.d.f.

(2) Beta r.v.'s are good **approximations** of r.v.'s that vary between **two limits**.

(3) If $X_1, X_2, \ldots, X_n$ are i.i.d. $\sim U(0,1)$ and $X_{(i)}$ is the i^{th} **smallest** r.v. of $X_1, X_2, \ldots, X_n$ so that $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, then

$$X_{(i)} \sim \mathcal{B}(i, n+1-i).$$

(4)

$$\begin{aligned} f_X'(x) &= \frac{(\alpha-1)x^{\alpha-2}(1-x)^{\beta-1} - (\beta-1)x^{\alpha-1}(1-x)^{\beta-2}}{B(\alpha,\beta)} \\ &= \frac{x^{\alpha-2}(1-x)^{\beta-2}}{B(\alpha,\beta)}[(\alpha-1)-(\alpha+\beta-2)x] \\ &\Rightarrow f_X'(x)\begin{cases} > 0, \Leftrightarrow (\alpha+\beta-2)x < \alpha-1 \\ = 0, \Leftrightarrow (\alpha+\beta-2)x = \alpha-1 \\ < 0, \Leftrightarrow (\alpha+\beta-2)x > \alpha-1 \end{cases} \end{aligned}$$

$$\begin{aligned} &f_X''(x) \\ &= \frac{(\alpha-1)(\alpha-2)x^{\alpha-3}(1-x)^{\beta-1} - (\beta-1)(\beta-2)x^{\alpha-1}(1-x)^{\beta-3}}{B(\alpha,\beta)} \\ &= \frac{x^{\alpha-3}(1-x)^{\beta-3}}{B(\alpha,\beta)}[(\alpha+\beta-2)(\alpha+\beta-3)x^2 - 2(\alpha-1)(\alpha+\beta-3)x + (\alpha-1)(\alpha-2)] \\ &= \begin{cases} \dfrac{x^{\alpha-3}(1-x)^{\beta-3}}{B(\alpha,\beta)}(\alpha+\beta-2)(\alpha+\beta-3)\left[\left(x-\dfrac{\alpha-1}{\alpha+\beta-2}\right)^2 - \dfrac{(\alpha-1)(\beta-1)}{(\alpha+\beta-2)^2(\alpha+\beta-3)}\right], & \alpha+\beta \neq 2,3 \\ \dfrac{x^{\alpha-3}(1-x)^{\beta-3}}{B(\alpha,\beta)}\cdot 2\cdot(\alpha-1)\cdot\left(x-\dfrac{\alpha-2}{2}\right), & \alpha+\beta = 2 \\ \dfrac{x^{\alpha-3}(1-x)^{\beta-3}}{B(\alpha,\beta)}\cdot(\alpha-1)\cdot(\alpha-2), & \alpha+\beta = 3 \end{cases} \end{aligned}$$

Theorem 7.4.1 Expectation and Variance of Beta R.V.
Suppose $X \sim \mathcal{B}(\alpha, \beta)$, then

$$E[X^n] = \frac{(\alpha)_n}{(\alpha+\beta)_n} = \frac{\binom{\alpha+n-1}{n}}{\binom{\alpha+\beta+n-1}{n}}.$$

Therefore,

$$E[X] = \frac{\alpha}{\alpha+\beta}$$

and

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

Theorem 7.4.2 Beta R.V. vs Binomial R.V.
Suppose $X \sim \mathcal{B}(\alpha, \beta)$, and $Y \sim \text{binomial}(\alpha+\beta-1, p)$, where $\alpha, \beta \in \mathbb{Z}^+, 0 < p < 1$. Then

$$P(X \leq p) = P(Y \geq \alpha)$$

and

$$P(X \geq p) = P(Y \leq \alpha - 1).$$

Chapter.8 Bivariate and Multivariate Distributions

§ 8.1 Joint Distributions of Two or More R.v.'s

Definition 8.1.1 Joint P.M.F. of Multiple R.v.'s
Let $X_1, X_2, \dots, X_n$ be discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
The nonnegative function $P_X: \mathbb{R}^n \to [0,1]$ given by

$$p_{\underline{X}}(\underline{x}) = P_{\underline{X}}(\{\underline{x}\}) = P(\underline{X} = \underline{x}) = \begin{cases} P(\underline{X} = \underline{x}), & \underline{x} \in \underline{X}(\Omega) \\ 0, & \underline{x} \in \mathbb{R}^n \setminus \underline{X}(\Omega) \end{cases}$$

is called the **joint p.m.f.** of $X_1, X_2, \dots, X_n$.

Remark 8.1.1 Properties of Joint P.M.F.
(1)

$$p_{\underline{X}}(\underline{x}) \geq 0, \forall \underline{x} \in \underline{X}(\Omega) \text{ and } p_{\underline{X}}(\underline{x}) = 0, \forall \underline{x} \in \mathbb{R}^n \setminus \underline{X}(\Omega).$$

(2)

$$\sum_{\underline{x} \in \underline{X}(\Omega)} p_{\underline{X}}(\underline{x}) = \sum_{\underline{x} \in \underline{X}(\Omega)} P(\underline{X} = \underline{x}) = P(\underline{X} \in \underline{X}(\Omega)) = P(\Omega) = 1$$

(3)

$$\underline{X}(\Omega) \subseteq \prod_{i=1}^{n} X_i(\Omega)$$

(4)

$$p_{\underline{X}}(\underline{x}) = \begin{cases} P(\underline{X} = \underline{x}), & \underline{x} \in \prod_{i=1}^{n} X_i(\Omega) \\ 0, & \underline{x} \in \mathbb{R}^n \setminus \prod_{i=1}^{n} X_i(\Omega) \end{cases}$$

Theorem 8.1.1 Joint Marginal P.M.F.
Suppose $X_1, X_2, \dots, X_n$ are discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
Then

$$p_{X_{i_1}, X_{i_2}, \dots, X_{i_k}}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = \begin{cases} \displaystyle\sum_{\substack{x_i \in X_i(\Omega) \\ i \neq i_1, i_2, \dots, i_k}} p_X(x), & x_i \in X_i(\Omega), \forall i = i_1, i_2, \dots, i_k \\ 0, & \text{o.w.} \end{cases}$$

We call $p_{X_{i_1}, X_{i_2}, \dots, X_{i_k}}(x_{i_1}, x_{i_2}, \dots, x_{i_k})$ the **joint p.m.f. marginalized** over $X_{i_1}, X_{i_2}, \dots, X_{i_k}$. If $k = 1$, we call $p_{X_i}(x_i)$ the **marginal p.m.f.** of X_i.

Theorem 8.1.2 Expectation of Measurable Function
Suppose $X_1, X_2, \dots, X_n$ are **discrete r.v.'s** of a probability space $(\Omega, \mathcal{A}, P)$,
and g is a **measurable function** from $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
Then

$$E[g(\underline{X})] = \sum_{\substack{x_i \in X_i(\Omega) \\ i=1,2,\dots,n}} g(\underline{x}) \cdot p_{\underline{X}}(\underline{x}).$$

Corollary 8.1.1 Expectation of Linear Combined Measurable Function
Suppose $X_1, X_2, \dots, X_n$ are discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$,
and $g_1, g_2, \dots, g_m$ are **measurable functions** from $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
and $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{R}$,
Then $\sum_{k=1}^{m} \alpha_k\, g_k(\underline{X})$ is a discrete r.v. of $(\Omega, \mathcal{A}, \mathrm{P})$ and

$$E\left[\sum_{k=1}^{m} \alpha_k\, g_k(\underline{X})\right] = \sum_{k=1}^{m} \alpha_k\, E[g_k(\underline{X})].$$

Definition 8.1.2 Joint P.D.F.
Let $X_1, X_2, \dots, X_n$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
We say that $X_1, X_2, \dots, X_n$ are **jointly continuous** r.v.'s if there exists a nonnegative function $f_{\underline{X}}\colon \mathbb{R}^n \to [0,1]$ s.t.

$$P(\underline{X} \in B) = \int\int_B \cdots \int f_{\underline{X}}(\underline{x})d\underline{x},\ \forall B \in \mathcal{B}_{\mathbb{R}^n}.$$

The function $f_{\underline{X}}$ is called the **joint p.d.f.** of $X_1, X_2, \cdots, X_n$.

Theorem 8.1.3 Joint Marginal P.D.F.
Suppose $X_1, X_2, \dots, X_n$ are **jointly continuous** r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
Then $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ are also jointly continuous r.v.'s of a probability space$(\Omega, \mathcal{A}, P)$ with joint p.d.f.

$$f_{X_{i_1}, X_{i_2}, \dots, X_{i_k}}(x_{i_1}, x_{i_2}, \dots, x_{i_k}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x}) dx_i$$

$i \neq i_1, i_2, \dots i_k$ and the integral has $n-k$ terms.

We call $f_{X_{i_1}, X_{i_2}, \dots, X_{i_k}}(x_{i_1}, x_{i_2}, \dots, x_{i_k})$ the **joint p.d.f.** marginalized over $X_{i_1}, X_{i_2}, \dots, X_{i_k}$.

If $k = 1$, we call $f_{X_i}(x_i)$ the **marginal p.d.f.** of X_i.

Theorem 8.1.4 Expectation of Measurable Function

Suppose $X_1, X_2, \ldots, X_n$ are **jointly continuous** r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$, and g is a measurable function from $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then

$$E\left[g\left(\underline{X}\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g\left(\underline{x}\right) f_{\underline{X}}\left(\underline{x}\right) dx_n \cdots dx_2 dx_1.$$

Remark 8.1.2 Properties of Joint P.D.F.

(1)

$$f_{\underline{X}}\left(\underline{x}\right) > 0, \forall \underline{x} \in \mathbb{R}^n$$

(2)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{\underline{X}}\left(\underline{x}\right) d\underline{x} = P\left(\underline{X} \in \mathbb{R}^n\right) = 1.$$

(3)

$$P(X_1 \in B_1, X_2 \in B_2, \cdots, X_n \in B_n) = \int_{B_1}\int_{B_2}\cdots\int_{B_n} f_{\underline{X}}\left(\underline{x}\right) dx_n \cdots dx_2\, dx_1$$

$\forall B_i \in \mathcal{B}_{\mathbb{R}^n}, i = 1,2, \ldots, n.$

(4)

$$P\left(\underline{X} = \underline{a}\right) = \int_{a_1}^{a_1}\int_{a_2}^{a_2}\cdots\int_{a_n}^{a_n} f_{\underline{X}}\left(\underline{x}\right) dx_n \cdots dx_2\, dx_1 = 0$$

(5)

$$P(a_i \le X_i \le a_i + \delta_i, i = 1,2, \ldots, n)$$

$$= \int_{a_1}^{a_1+\delta_1}\int_{a_2}^{a_2+\delta_2}\cdots\int_{a_n}^{a_n+\delta_n} f_{\underline{X}}\left(\underline{x}\right) dx_n \cdots dx_2\, dx_1$$

$= f_{\underline{X}}\left(\underline{a_{\delta}}\right) \cdot \delta_1 \cdot \delta_2 \cdots \delta_n$ for some $\underline{a_{\underline{\delta}}} \in \prod_{i=1}^{n}[a_i, a_i + \delta_i]$ if $f_{\underline{X}}\left(\underline{x}\right)$ is continuous.

$$\Rightarrow \lim_{\underline{\delta} \to \underline{0}} \frac{P(a_i \le X_i \le a_i + \delta_i, i = 1,2, \ldots, n)}{\delta_1 \cdot \delta_2 \cdots \delta_n} = \lim_{\underline{\delta} \to \underline{0}} f_{\underline{X}}\left(\underline{a_{\delta}}\right) = f_{\underline{X}}\left(\underline{a}\right)$$

and $P(a_i \le X_i \le a_i + \delta_i, i = 1,2, \ldots, n) \approx f_{\underline{X}}\left(\underline{a}\right) \cdot \delta_1 \cdot \delta_2 \cdots \delta_n.$

Corollary 8.1.2 Expectation of Linear Combined Measurable Function
Suppose $X_1, X_2, \dots, X_n$ are **jointly continuous** r.v.'s of a probability space $(\Omega, \mathscr{A}, P)$, and $g_1, g_2, \dots, g_m$ are **measurable functions** from $(\mathbb{R}^n, \mathscr{B}_{\mathbb{R}^n})$ to $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$.
and $\alpha_1, \alpha_2, \dots, \alpha_m \in \mathbb{R}$,
Then

$$\sum\nolimits_{k=1}^{m} \alpha_k \, g_k(\underline{X})$$

is a continuous r.v. of $(\Omega, \mathscr{A}, P)$ and

$$E\left[\sum_{k=1}^{m} \alpha_k \, g_k(\underline{X})\right] = \sum_{k=1}^{m} \alpha_k \, E\big[g_k(\underline{X})\big].$$

Definition 8.1.3 Joint C.D.F.
Let $X_1, X_2, \dots, X_n$ be r.v.'s of a probability space $(\Omega, \mathscr{A}, P)$.
The **joint c.d.f.** of $X_1, X_2, \dots, X_n$ is given by

$$F_{\underline{X}}(\underline{x}) = P(\, X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n), \forall \underline{x} \in \mathbb{R}^n.$$

Theorem 8.1.5 Joint Marginal C.D.F.
Suppose $X_1, X_2, \dots, X_n$ are r.v.'s of a probability space $(\Omega, \mathscr{A}, P)$.
Then

$$F_{X_{i_1}, X_{i_2}, \dots, X_{i_k}}\left(x_{i_1}, x_{i_2}, \dots, x_{i_k}\right) = F_{\underline{X}}\left(\infty, \dots, \infty, x_{i_1}, \infty, \dots, \infty, x_{i_2}, \infty, \dots, \infty, x_{i_k}, \infty, \dots, \infty\right)$$

We call $F_{X_{i_1}, X_{i_2}, \dots, X_{i_k}}\left(x_{i_1}, x_{i_2}, \dots, x_{i_k}\right)$ the **joint c.d.f.** marginalized over $X_1, X_2, \dots, X_n$.

If $k = 1$, we call $F_{X_i}(x_i)$ the **marginal c.d.f.** of X_i.

Theorem 8.1.6 Properties of Joint C.D.F.
Suppose $X_1, X_2, \dots, X_n$ are r.v.'s of a probability space $(\Omega, \mathscr{A}, P)$.

(1) $F_{\underline{X}}(\underline{x})$ is **increasing** and **right continuous** in each argument $x_i, i = 1,2, \dots, n$.

(2) $F_{\underline{X}}(\underline{x}) = 0$ if there exists at least one i such that $x_i = -\infty$.

(3) $F_{\underline{X}}(\infty, \infty, \dots, \infty) = 1$.

(4) If $X_1, X_2, \dots, X_n$ are **jointly continuous** r.v.'s, then

$$f_{\underline{X}}(\underline{x}) = \frac{\partial F_{\underline{X}}(\underline{x})}{\partial x_1 \partial x_2 \cdots \partial x_n}, \forall \underline{x} \in \mathbb{R}^n.$$

§ 8.2 Independent R.V.'s

Definition 8.2.1 Independent Set
Let $X_i, i \in I$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
We say that the r.v.'s $X_i, i \in I$ are **independent**

if for any finite subset $\{X_{i_1}, X_{i_2}, \dots, X_{i_k}\}$ $(k \geq 2)$ of $\{X_i, i \in I\}$,

the events $X_{i_1} \in B_{i_1}, X_{i_2} \in B_{i_2}, \dots, X_{i_k} \in B_{i_k}$ are independent $\forall B_{i_1}, B_{i_2}, \dots, B_{i_k} \in \mathcal{B}_{\mathbb{R}}$.
Otherwise, the r.v.'s $X_i, i \in I$ are dependent.

Theorem 8.2.1 Equivalent Statements of Independence
Suppose $X_1, X_2, \dots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
The following three statements are **equivalent**:

(1) $X_1, X_2, \dots, X_n$ are independent.

(2)

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i), \forall B_1, B_2, \dots, B_n \in \mathcal{B}_{\mathbb{R}}$$

(3)

$$F_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall \underline{x} \in \mathbb{R}^n$$

Theorem 8.2.2 Necessary and Sufficient Condition of Independence
Suppose $X_1, X_2, \dots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
(1) If $X_1, X_2, \dots, X_n$ are **discrete** r.v.'s,
then $X_1, X_2, \dots, X_n$ are independent

$$\Leftrightarrow P_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} P_{X_i}(x_i), \forall \underline{x} \in \mathbb{R}^n$$

(2) If $X_1, X_2, \dots, X_n$ are **jointly continuous** r.v.'s,
then $X_1, X_2, \dots, X_n$ are independent

$$\Leftrightarrow f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f_{X_i}(x_i), \forall \underline{x} \in \mathbb{R}^n$$

Definition 8.2.2 Indicator Function
Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $A \in \mathcal{A}$.
The **indicator function** I_A of the event A is given by

$$I_A(w) = \begin{cases} 1, & \text{if } w \in A \\ 0, & \text{o.w.} \end{cases} \qquad \text{i.e.} \qquad I_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{o.w.} \end{cases}$$

Theorem 8.2.3 Indicator Function is a Discrete Measurable Function
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space.
I_A is a **discrete r.v.** of $(\Omega, \mathcal{A}, P)$ for all $A \in \mathcal{A}$.

Theorem 8.2.4 Indicator R.V.'s Indicates Independence
Suppose $(\Omega, \mathcal{A}, P)$ is a probability space, and $A_1, A_2, \dots, A_n \in \mathcal{A}$.
The events $A_1, A_2, \dots, A_n$ are **independent**
$\Leftrightarrow$ the **indicator r.v.'s** $I_{A_1}, I_{A_2}, \dots, I_{A_n}$ are **independent**.

Theorem 8.2.5 Expectation of Measurable Functions of Independent R.V.
Suppose $X_1, X_2, \dots, X_n$ are independent r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$, and $g_1, g_2, \dots, g_n$ are measurable functions from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.
Then $g_1(X_1), g_2(X_2), \dots, g_n(X_n)$ are independent and

$$E\left[\prod_{i=1}^{n} g_i(X_i)\right] = \prod_{i=1}^{n} E[g_i(X_i)].$$

Remark 8.2.1 Independent Expectations Can't Imply Independence of R.V.'s
The converse is **not true**, i.e.,

$$E\left[\prod_{i=1}^{n} g_i(X_i)\right] = \prod_{i=1}^{n} E[g_i(X_i)] \nRightarrow g_1(X_1), g_2(X_2), \dots, g_n(X_n) \text{ are independent.}$$

§ 8.3 Conditional Distributions

Recall 8.3.1 Properties of Conditional Probability
Suppose $(\Omega, \mathcal{A}, P)$ be a probability space,
and $A, B, A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_n \in \mathcal{A}$.

$$P(A|B) = \begin{cases} \dfrac{P(A \cap B)}{P(B)}, & \text{if } P(B) \neq 0 \\ 0 & , \text{if } P(B) = 0 \end{cases}$$

(1) If $P(B) \neq 0$, then $P(\cdot\,|B)$ regarded as a function on $\mathcal{A}$ is a **probability measure.**

(2) **Multiplication theorem:**

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)\cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

(3) **Total probability theorem:**
If $\{B_n\}_{n=1}^{\infty}$ is a partition of Ω, then

$$P(A) = \sum_{n=1}^{\infty} P(B_n) \cdot P(A|B_n), \forall A \in \mathcal{A}.$$

(4) **Bayes' theorem:**
If $P(A) \neq 0$ and $\{B_n\}_{n=1}^{\infty}$ is a partition of Ω, then

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{n=1}^{\infty} P(B_n) \cdot P(A|B_n)}, \forall\, A \in \mathcal{A}, P(A) > 0, k = 1,2, \dots$$

◎ $P_{X|Y}(x|y)$: X and Y are discrete r.v.'s

Definition 8.3.1 P.M.F. and C.D.F. of D-D
Let X and Y be discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ and $y \in \mathbb{R}$.
The conditional p.m.f. $P_{X|Y}(x|y)$ of X given that $Y = y$ is given by

$$P_{X|Y}(x|y) = \begin{cases} P(X = x|Y = y) = \dfrac{P(X = x, Y = y)}{P(Y = y)} = \dfrac{P_{X,Y}(x,y)}{P_Y(y)}, & P_Y(y) \neq 0, \forall x \in \mathbb{R} \\ 0, & \text{o.w.} \end{cases}$$

The conditional c.d.f. $F_{X|Y}(\cdot\,|y)$ of X given that $Y = y$ is given by

$$F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{\substack{t \leq X \\ t \in X(\Omega)}} P(X = t|Y = y) = \sum_{\substack{t \leq X \\ t \in X(\Omega)}} P_{X|Y}(t|y), \forall x \in \mathbb{R}.$$

Remark 8.3.1 Joint P.M.F.

(1) $P_{X,Y}(x,y) = P_Y(y)\cdot P_{X|Y}(x|y) = P_X(x)\cdot P_{Y|X}(y|x)$.

(2) A similar definition can be made for discrete **random vectors**.

Theorem 8.3.1 Properties of D-D Conditional Probability

Suppose $X, Y, X_1, X_2, \dots, X_n$ are discrete r.v.'s of a probability space $(\Omega, \mathscr{A}, P)$.

(1) If $y \in \mathbb{R}$ and $P_Y(y) \neq 0$, then $P_{X|Y}(\cdot\,|y)$ is a p.m.f.

(2) $P_X(x) = P_{X_1}(x_1)\cdot P_{X_2|X_1}(x_2|x_1)\cdots P_{X_n|X_1,X_2,\dots,X_{n-1}}(x_n|x_1,x_2,\dots,x_{n-1}), \forall x \in \mathbb{R}^n$.

(3)

$$P_X(x) = \sum_{y\in Y(\Omega)} P_Y(y)\cdot P_{X|Y}(x|y), \forall x \in \mathbb{R}.$$

(4) If $x \in \mathbb{R}$ and $P_X(x) \neq 0$, then

$$P_{Y|X}(y|x) = \frac{P_Y(y)\cdot P_{X|Y}(x|y)}{\sum_{y\in Y(\Omega)} P_Y(y)\cdot P_{X|Y}(x|y)}, \forall y \in \mathbb{R}.$$

◎ $f_{X|Y}(x|y)$: X and Y are jointly continuous r.v.'s

Definition 8.3.2 C.D.F. and P.D.F. of C-C

Let X and Y be jointly continuous r.v.'s of a probability space $(\Omega, \mathscr{A}, P)$ and $y \in \mathbb{R}$. The conditional c.d.f. $F_{X|Y}(x|y)$ of X given that $Y = y$ is given by

$$F_{X|Y}(x|y) = \begin{cases} \lim\limits_{\delta\to 0} P(X = x|y \le Y \le y+\delta) = \lim\limits_{\delta\to 0}\dfrac{P(X = x, y \le Y \le y+\delta)}{P(y \le Y \le y+\delta)} \\ \qquad = \lim\limits_{\delta\to 0}\dfrac{[F_{X,Y}(x,y+\delta) - F_{X,Y}(x,y)]/\delta}{[F_Y(y+\delta) - F_Y(y)]/\delta} \\ \qquad = \dfrac{\frac{\partial F_{X,Y}(x,y)}{\partial y}}{f_Y(y)}, f_Y(y) \neq 0, \forall x \in \mathbb{R} \\ 0, \quad \text{o.w.} \end{cases}$$

The conditional p.d.f. $f_{X|Y}(\cdot\,|y)$ of X given that $Y = y$ is given by

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{\partial F_{X,Y}(x,y)}{\partial x} = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}, f_Y(y) \neq 0, \forall x \in \mathbb{R} \\ 0, \quad \text{o.w.} \end{cases}$$

Remark 8.3.2 Joint P.D.F.

(1)

$$f_{X,Y}(x,y) = f_Y(y)\cdot f_{X|Y}(x|y) = f_X(x)\cdot f_{Y|X}(y|x), \forall x, y \in \mathbb{R}$$

(2) A similar definition can be made for jointly continuous **random vectors**.

Theorem 8.3.2 Properties of C-C Conditional Probability

Suppose $X, Y, X_1, X_2, \dots, X_n$ are jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.

(1) If $y \in \mathbb{R}$ and $f_Y(y) \neq 0$, then $f_{X|Y}(\cdot\,|y)$ is a p.d.f.

(2) $f_X(x) = f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2|x_1) \cdots f_{X_n|X_1,X_2,\dots,X_{n-1}}(x_n|x_1, x_2, \dots, x_{n-1}), \forall x \in \mathbb{R}^n$.

(3)

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) \cdot f_{X|Y}(x|y)\,, \forall x \in \mathbb{R}.$$

(4) If $x \in \mathbb{R}$ and $f_X(x) \neq 0$, then

$$f_{Y|X}(y|x) = \frac{f_Y(y) \cdot f_{X|Y}(x|y)}{\int_{-\infty}^{\infty} f_Y(y) \cdot f_{X|Y}(x|y)}, \forall y \in \mathbb{R}.$$

◎ $f_{X|Y}(x|y)$ and $P_{X|Y}(x|y)$: X is a continuous r.v. and Y is a discrete r.v.

Definition 8.3.3 C.D.F., P.D.F. and P.M.F. of C-D and D-C

Let X be a continuous r.v. and Y be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$.

The conditional **c.d.f.** $F_{X|Y}(\cdot\,|y)$ of X given that $Y = y, y \in \mathbb{R}$ is given by

$$F_{X|Y}(x|y) = \begin{cases} P(X \leq x|Y = y), P_Y(y) \neq 0, \forall x \in \mathbb{R} \\ \qquad 0, \qquad\qquad \text{o.w.} \end{cases}$$

The conditional **p.d.f.** $f_{X|Y}(\cdot\,|y)$ of X given that $Y = y, y \in \mathbb{R}$ is given by

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{\partial F_{X,Y}(x,y)}{\partial x} = \lim\limits_{\delta \to 0} \dfrac{F_{X|Y}(x+\delta|y) - F_{X|Y}(x|y)}{\delta} \\ \qquad\qquad = \lim\limits_{\delta \to 0} \dfrac{P(x \leq X \leq x+\delta|Y=y)}{\delta}, P_Y(y) \neq 0, \forall x \in \mathbb{R} \\ \quad 0, \qquad\qquad\qquad\qquad \text{o.w.} \end{cases}$$

The conditional **p.m.f.** $P_{X|Y}(\cdot\,|y)$ of Y given that $X = x, x \in \mathbb{R}$ is given by

$$P_{Y|X}(y|x) = \begin{cases} \lim\limits_{\delta \to 0} P(Y = y|x \leq X \leq x+\delta) = \lim\limits_{\delta \to 0} \dfrac{P(Y=y) \cdot P(x \leq X \leq x+\delta|Y=y)/\delta}{P(x \leq X \leq x+\delta)/\delta} \\ \qquad\qquad = \dfrac{P_Y(y) \cdot f_{X|Y}(x|y)}{f_X(x)}, f_X(x) \neq 0, \forall y \in \mathbb{R} \\ \quad 0, \qquad\qquad\qquad \text{o.w.} \end{cases}$$

The conditional **c.d.f.** $F_{Y|X}(\cdot\,|x)$ of Y given that $X = x, x \in \mathbb{R}$ is given by

$$F_{Y|X}(y|x) = \begin{cases} \sum\limits_{\substack{t \leq X \\ t \in X(\Omega)}} P_{Y,X}(t|x) = \dfrac{\sum_{t \leq X, t \in X(\Omega)} P_Y(t) \cdot f_{X|Y}(x|t)}{f_X(x)}, f_X(x) \neq 0, \forall y \in \mathbb{R} \\ \qquad 0, \qquad\qquad\qquad \text{o.w.} \end{cases}$$

Remark 8.3.3 Calculation of C-D P.D.F. and D-C P.M.F.

(1) $P_Y(y) \cdot f_{X|Y}(x|y) = f_X(x) \cdot P_{Y|X}(y|x), \forall x, y \in \mathbb{R}$

(2) If $y \in \mathbb{R}$ and $P_Y(y) \neq 0$, then

$$f_{X|Y}(x|y) = \frac{f_X(x) \cdot P_{Y|X}(y|x)}{P_Y(y)}, \forall x \in \mathbb{R}.$$

If $x \in \mathbb{R}$ and $f_X(x) \neq 0$, then

$$P_{Y|X}(y|x) = \frac{P_Y(y) \cdot f_{X|Y}(x|y)}{f_X(x)}, \forall y \in \mathbb{R}.$$

Theorem 8.3.3 Properties of C-D and D-C Conditional Probability

Suppose X is a continuous r.v. and Y is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, P)$.

(1) If $y \in \mathbb{R}$ and $P_Y(y) \neq 0$, then $f_{X|Y}(\cdot\,|y)$ is a p.d.f.

If $x \in \mathbb{R}$ and $f_X(x) \neq 0$, then $P_{Y|X}(y|x)$ is a p.m.f.

(2)
$$f_X(x) = \textstyle\sum_{y \in Y(\Omega)} P_Y(y) \cdot f_{X|Y}(x|y), \quad \forall x \in \mathbb{R}.$$
$$P_Y(y) = \int_{-\infty}^{\infty} f_X(x) \cdot P_{Y|X}(y|x), \qquad \forall y \in \mathbb{R}.$$

(3) If $x \in \mathbb{R}$ and $f_X(x) \neq 0$, then

$$P_{Y|X}(y|x) = \frac{P_Y(y) \cdot f_{X|Y}(x|y)}{\sum_{y \in Y(\Omega)} P_Y(y) \cdot f_{X|Y}(x|y)}, \forall y \in \mathbb{R}.$$

If $y \in \mathbb{R}$ and $P_Y(y) \neq 0$, then

$$f_{X|Y}(x|y) = \frac{f_X(x) \cdot P_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) \cdot P_{Y|X}(y|x)}, \qquad \forall x \in \mathbb{R}.$$

Definition 8.3.4 Expectation of Conditional R.V.

Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ and $y \in \mathbb{R}$.

The conditional expectation $E[X|Y = y]$ of X given that $Y = y$ is given by

$$E[X|Y = y] = \begin{cases} \displaystyle\sum_{x \in X(\Omega)} x \cdot P_{X|Y}(x|y), & \text{if } X \text{ is a discrete r.v.} \\ \displaystyle\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)dx, & \text{if } X \text{ is a continuous r.v.} \end{cases}$$

Theorem 8.3.4 Expectation of Conditional Measurable Function

Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$,

and g is a measurable function from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then

$$E[g(X)|Y = y] = \begin{cases} \displaystyle\sum_{x \in X(\Omega)} g(x) \cdot P_{X|Y}(x|y), & \text{if } X \text{ is a discrete r.v.} \\ \displaystyle\int_{-\infty}^{\infty} g(x) \cdot f_{X|Y}(x|y)dx, & \text{if } X \text{ is a continuous r.v.} \end{cases}$$

§ 8.4 Transformations of Two R.V.'s

Theorem 8.4.1 Transformations of Two R.V.'s
Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$, g and h are measurable functions from $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, and $U = g(X,Y)$ and $V = h(X,Y)$.

(1) If X and Y are discrete r.v.'s, then U and V are discrete r.v.'s and

$$P_{U,V}(u,v) = \sum_{(x,y):g(,y)=u,h(x,y)=v} P_{X,Y}(x,y).$$

(2) If X and Y are jointly continuous r.v.'s, U and V are discrete r.v.'s, then

$$P_{U,V}(u,v) = \iint_{\{(x,y):g(x,y)=u,h(x,y)=v\}} f_{X,Y}(x,y)dxdy.$$

(3) If X and Y are jointly continuous r.v.'s, U and V are jointly continuous r.v.'s, and

$$J(x,y) = \begin{vmatrix} \dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y} \\ \dfrac{\partial h(x,y)}{\partial x} & \dfrac{\partial h(x,y)}{\partial y} \end{vmatrix} \neq 0$$

$\forall (x,y) \in \{(x,y): g(x,y) = u, h(x,y) = v\}$,
where $J(x,y)$ is the Jacobian determinant, $(u,v) \in g(X,Y)(\Omega) \times h(X,Y)(\Omega)$, then

$$f_{U,V}(u,v) = \sum_{(x,y):g(,y)=u,h(x,y)=v} \frac{f_{X,Y}(x,y)}{|J(x,y)|}$$

Theorem 8.4.2 Convolution Theorem
Suppose X and Y are two independent r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ and $Z = X + Y$.

(1) If X and Y are discrete r.v.'s, then

$$P_Z(z) = \sum_{x \in X(\Omega)} P_X(x) \cdot P_Y(z-x)$$

(2) If X and Y are jointly continuous r.v.'s, then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)dx.$$

§ 8.5 Order Statistics

Definition 8.5.1 Order Statistic

Let $X_1, X_2, \ldots, X_n$ be i.i.d. r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
The i^{th} order statistic $X_{(i)}$, $i = 1,2,\ldots,n$ of $X_1, X_2, \ldots, X_n$ is defined as the i^{th} **smallest** value in $\{X_1, X_2, \ldots, X_n\}$ so that $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$,
namely, $X_{(i)}(w)$ =the i^{th} smallest value in $\{X_1(w), X_2(w), \ldots, X_n(w)\}$ for all $w \in \Omega$.
In particular, $X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$ and $X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$.

Remark 8.5.1 Without Equal & Not I.I.D.

(1) If $X_1, X_2, \ldots, X_n$ are jointly continuous r.v.'s, then

$$P\left(X_{(i)} = X_{(j)}\right) = 0, \forall i \neq j \;\Rightarrow P\left(X_{(1)} < X_{(2)} < \cdots < X_{(n)}\right) = 1.$$

(2) $X_{(i)}$, $i = 1,2,\ldots,n$ is a function of $X_1, X_2, \ldots, X_n$
$\Rightarrow$ $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are **neither independent nor identically distributed** in general.

Definition 8.5.2 Random Sample

A **random sample** of size n of a probability space $(\Omega, \mathcal{A}, P)$ is a sequence of n i.i.d. r.v.'s $X_1, X_2, \ldots, X_n$ of $(\Omega, \mathcal{A}, P)$.

Definition 8.5.3 Range, Midrange, Median and Mean of Random Sample

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n of a probability space $(\Omega, \mathcal{A}, P)$.
The **sample range** is given by $X_{(1)} + X_{(n)}$.
The **sample midrange** is given by $\frac{X_{(1)} + X_{(n)}}{2}$.
The **sample median** is given by $\begin{cases} X_{(i-1)}, & \text{if } n = 2i+1 \\ \frac{X_{(i)} + X_{(i+1)}}{2}, & \text{if } n = 2i \end{cases}$

The **sample mean** $\overline{X}$ is given by $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

Remark 8.5.1 Forced Decline

If $\exists\ i_j < i_l \Rightarrow x_{i_j} \ge x_{i_l}$, then

$$F_{X_{(i_1)}, X_{(i_2)}, \ldots, X_{(i_k)}}\left(x_{i_1}, \ldots, x_{i_j}, \ldots, x_{i_l}, \ldots, x_{i_k}\right) = F_{X_{(i_1)}, X_{(i_2)}, \ldots, X_{(i_k)}}\left(x_{i_1}, \ldots, x_{i_l}, \ldots, x_{i_l}, \ldots, x_{i_k}\right)$$

and $f_{X_{(i_1)}, X_{(i_2)}, \ldots, X_{(i_k)}}\left(x_{i_1}, x_{i_2}, \ldots, x_{i_k}\right) = 0$.

Theorem 8.5.1 C.D.F. and P.D.F. of Jointly Order R.V.'s

Suppose $X_1, X_2, \dots, X_n$ are i.i.d. jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ with common c.d.f. $F(x)$ and common p.d.f. $f(x)$.

If $1 \le i_1 \le i_2 \le \cdots \le i_k \le n, -\infty < x_{i_1} < x_{i_2} < \cdots < x_{i_k} < \infty$, then

$$F_{X_{(i_1)},X_{(i_2)},\dots,X_{(i_k)}}\left(x_{i_1}, x_{i_2}, \dots, x_{i_k}\right)$$

$$= \sum_{j_k=i_k}^{n} \sum_{j_{k-1}=i_{k-1}}^{j_k} \cdots \sum_{j_1=i_1}^{j_2} \binom{n}{j_k}\binom{j_k}{j_{k-1}} \cdots \binom{j_2}{j_1} \left[F\left(x_{i_1}\right)\right]^{j_1} \left[F\left(x_{i_2}\right) - F\left(x_{i_1}\right)\right]^{j_2-j_1}$$

$$\cdots \left[F\left(x_{i_k}\right) - F\left(x_{i_{k-1}}\right)\right]^{j_k-j_{k-1}} \left[1 - F\left(x_{i_k}\right)\right]^{n-j_k}$$

$$f_{X_{(i_1)},X_{(i_2)},\dots,X_{(i_k)}}\left(x_{i_1}, x_{i_2}, \dots, x_{i_k}\right)$$

$$= \frac{n!}{(i_1-1)!\,(i_2-i_1-1)!\cdots(i_k-i_{k-1}-1)!\,(n-i_k)!} f\left(x_{i_1}\right) f\left(x_{i_2}\right) \cdots f\left(x_{i_k}\right)$$

$$\cdot \left[F\left(x_{i_1}\right)\right]^{i_1-1} \left[F\left(x_{i_2}\right) - F\left(x_{i_1}\right)\right]^{i_2-i_1-1} \cdots \left[F\left(x_{i_k}\right) - F\left(x_{i_{k-1}}\right)\right]^{i_k-i_{k-1}-1} \left[1 - F\left(x_{i_k}\right)\right]^{n-i_k}$$

Corollary 8.5.1 Beta R.V. vs Binomial R.V.

Suppose $X_1, X_2, \dots, X_n$ are i.i.d. r.v.'s $\sim U(0,1)$, then

$$X_{(i)} \sim \mathcal{B}(i, n+1-i), i = 1,2, \dots, n.$$

Proof:

$$\begin{aligned} f_{X_{(i)}}(x) &= \frac{n!}{(i-1)!\,(n-i)!} f(x)[F(x)]^{i-1}[1-F(x)]^{n-i} \\ &= \frac{n!}{(i-1)!\,(n-i)!} 1 \cdot x^{i-1}(1-x)^{n-i} \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n+1-i)} x^{i-1}(1-x)^{(n+1-i)-1} \\ &= \frac{x^{i-1}(1-x)^{(n+1-i)-1}}{B(i, n+1-i)}, 0 < x < 1 \end{aligned}$$

$$\Rightarrow X_{(i)} \sim \mathcal{B}(i, n+1-i)$$

Corollary 8.5.1 Cases One, Two and n Order R.V.'s
Suppose $X_1, X_2, \dots, X_n$ are jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ with continuous c.d.f. $F(x)$ and continuous p.d.f. $f(x)$.

(1) $F_{X_{(i)}}(x) = \sum_{j=i}^{n} \binom{n}{j} [F(x)]^j [1-F(x)]^{n-j}, \infty < x < \infty$

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!\,(n-i)!} f(x)[F(x)]^{i-1}[1-F(x)]^{n-i}, \infty < x < \infty$$

In particular, $F_{X_{(1)}}(x) = 1 - [1-F(x)]^n, \infty < x < \infty,$

$f_{X_{(1)}}(x) = nf(x)[1-F(x)]^{n-1}, \infty < x < \infty.$

And $F_{X_{(n)}}(x) = [F(x)]^n, f_{X_{(1)}}(x) = nf(x)[F(x)]^{n-1}, \infty < x < \infty.$

(2) $F_{X_{(i_1)}, X_{(i_2)}}(x, y)$

$$= \sum_{j_2=i_2}^{n} \sum_{j_1=i_1}^{j_2} \binom{n}{j_2}\binom{j_2}{j_1} [F(x)]^{j_1}[F(y)-F(x)]^{j_2-j_1}[1-F(y)]^{n-j_2}, -\infty < x < y < \infty$$

$$f_{X_{(i_1)}, X_{(i_2)}}(x, y) = \frac{n!}{(i_1-1)!\,(i_2-i_1-1)!\,(n-i_2)!} f(x)f(y)[F(x)]^{j_1}$$

$$\cdot [F(y)-F(x)]^{j_2-j_1}[1-F(y)]^{n-j_2}, -\infty < x < y < \infty$$

(3) $F_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(x_1, x_2, \dots, x_n)$

$$= \sum_{j_{n-1}=i_{n-1}}^{n} \sum_{j_{n-2}=i_{n-2}}^{j_{n-1}} \cdots \sum_{j_1=i_1}^{j_2} \binom{n}{j_{n-1}}\binom{j_{n-1}}{j_{n-2}} \cdots \binom{j_2}{j_1} [F(x_1)]^{j_1}\,[F(x_2)-F(x_1)]^{j_2-j_1}$$

$$\cdots\ [F(x_{n-1})-F(x_{n-2})]^{j_{n-1}-j_{n-2}}\,[F(x_n)-F(x_{n-1})]^{n-j_{n-1}}$$

$f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(x_1, x_2, \dots, x_n)$

$$= n!\, f(x_1)f(x_2)\cdots f(x_n), \quad -\infty < x_1 < x_2 < \cdots < x_n < \infty$$

§ 8.6 Multinomial Distributions

◎ Consider an experiment with k possible outcomes $w_1, w_2, \dots, w_k$.
Let $A_{(i)} = \{w_i\}$ be the event that the outcome is w_i
and let $P_i = P(A_i), i = 1,2, \dots, k$.
Suppose that such an experiment is independently and successively performed n times.
Let $X_i, i = 1,2, \dots, k$ be the number of times that event A_i occurs.
Then $P_{X_1, X_2, \dots, X_k}(x_1, x_2, \cdots, x_k) = P(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k)$
$= \dfrac{n!}{x_1!\, x_2! \cdots x_k!} P_1^{x_1} P_2^{x_2} \cdots P_k^{x_k}, x_1, x_2, \dots, x_k \geq 0$ and $x_1 + x_2 + \cdots + x_k = n$.

Definition 8.6.1 Multinomial Joint R.V.'s
Let $X_1, X_2, \dots, X_k$ be discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
We call $X_1, X_2, \dots, X_k$ multinomial joint r.v.'s with parameters $n, P_1, P_2, \dots, P_k$, where $n \geq 1, P_1, P_2, \dots, P_k \geq 0, P_1 + P_2 + \cdots + P_k = 1$, if the joint p.m.f. is given by

$$P_{\underline{X}}(\underline{x}) = \begin{cases} \dfrac{n!}{x_1!\, x_2! \cdots x_k!} P_1^{x_1} P_2^{x_2} \cdots P_k^{x_k}, & x_1, x_2, \dots, x_k \geq 0 \text{ and } x_1 + x_2 + \cdots + x_k = n \\ 0, & \text{o.w.} \end{cases}$$

Such a joint p.m.f. is called a **multinomial** joint p.m.f. with parameters $n, P_1, P_2, \dots, P_k$.

Remark 8.6.1 Verification of P.M.F.
$P_{\underline{X}}(\underline{x}) \geq 0\ \forall \underline{x} \in \mathbb{R}^n$ and

$$\sum_{\substack{x_1, x_2, \dots, x_k \geq 0 \\ x_1 + x_2 + \cdots + x_k = n}} \frac{n!}{x_1!\, x_2! \cdots x_k!} P_1^{x_1} P_2^{x_2} \cdots P_k^{x_k} = (P_1 + P_2 + \cdots + P_k)^n = 1$$

$\Rightarrow P_{\underline{X}}(\underline{x})$ is a p.m.f.

Theorem 8.6.1 Splitting of Multinomial Joint R.V.'s
Suppose $X_1, X_2, \dots, X_l$ are multinomial r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$, with parameters $n, P_1, P_2, \dots, P_l$, where $n \geq 1, P_1, P_2, \dots, P_k \geq 0, P_1 + P_2 + \cdots + P_k = 1$.

Then $X_{(i_1)}, X_{(i_2)}, \cdots, X_{(i_k)}, n - X_{(i_1)} - X_{(i_2)} - \cdots - X_{(i_k)}$ are multinomial joint r.v.'s with parameters $n, P_{i_1}, P_{i_2}, \dots, P_{i_k}, 1 - P_{i_1} - P_{i_2} - \cdots - P_{i_k}$.

Chapter.9 More Expectations and Variance

§ 9.1 Expected Values of Sums of R.V.'s

Theorem 9.1.1 Expectations of Sum of Finite R.V.'s
Suppose $X_1, X_2, \ldots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$, then

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$$

Theorem 9.1.2 Expectations of Sum of Infinite R.V.'s
Suppose $X_1, X_2, \ldots$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
If $\sum_{i=1}^{\infty} E[X_i] < \infty$ or if X_i is nonnegative for all $i = 1,2, \ldots$, then

$$E\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i=1}^{\infty} E[X_i].$$

Remark 9.1.1 General Expectations of Sum of Infinite R.V.'s
In general,

$$E\left[\sum_{i=1}^{\infty} X_i\right] \neq \sum_{i=1}^{\infty} E[X_i].$$

Corollary 9.1.1 Expectation of Integer-Valued R.V.
Suppose X is an integer-valued r.v. of a probability space $(\Omega, \mathcal{A}, P)$, then

$$E[X] = \sum_{i=1}^{\infty} P(x \geq i) - \sum_{i=1}^{\infty} P(x \leq -i).$$

§ 9.2 Covariance and Correlation Coefficients

Theorem 9.2.1 Cauchy-Schwarz Inequality

Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$,
and suppose $E[X^2]$ and $E[Y^2]$ exists. Then

$$|E[XY]| \leq \sqrt{E[X^2] \cdot E[Y^2]}.$$

"=" $\Leftrightarrow X = 0$ with probability 1 or $Y = 0$ with probability 1
or $Y = aX$ with probability 1, where

$$a = \frac{E[XY]}{E[X^2]}.$$

Remark 9.2.1 Cauchy-Schwarz Equalities

Suppose that $E[X^2] \neq 0$ and $E[Y^2] \neq 0$, then

$E[XY] = \sqrt{E[X^2] \cdot E[Y^2]} \Leftrightarrow Y = aX$ with probability 1, where

$$a = \frac{E[XY]}{E[X^2]} = \sqrt{\frac{E[Y^2]}{E[X^2]}} > 0.$$

$E[XY] = -\sqrt{E[X^2] \cdot E[Y^2]} \Leftrightarrow Y = aX$ with probability 1, where

$$a = \frac{E[XY]}{E[X^2]} = -\sqrt{\frac{E[Y^2]}{E[X^2]}} < 0.$$

Corollary 9.2.1 Variance Larger Than or Equal to Zero

Suppose X is a r.v. of a probability space $(\Omega, \mathcal{A}, P)$ and suppose $E[X^2]$ exists, then

$$|E[X]|^2 \leq E[X^2].$$

Definition 9.2.1 Covariance

Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ with means μ_X and μ_Y, resp.
The covariance $Cov(X,Y)$ (or $\sigma_{X,Y}$) of X and Y is given by

$$Cov(X,Y) = \sigma_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)].$$

We say that X and Y are positively correlated, negatively correlated and uncorrelated if $Cov(X,Y) > 0, Cov(X,Y) < 0$ and $Cov(X,Y) = 0$, resp.

Remark 9.2.2 Covariance of Linear Combination of Two R.V.'s

(1) $Var(X) = E[(X-\mu_X)^2]$ is a measure of the spread or dispersion of X.
$Var(Y) = E[(Y-\mu_Y)^2]$ is a measure of the spread or dispersion of Y.
$Cov(X,Y) = \sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]$ is a measure of the joint spread or dispersion of X and Y.

(2) $Var(aX+bY) = E\left[\left((aX+bY)-(a\mu_X+b\mu_Y)\right)^2\right]$

$= E\left[\left(a(X-\mu_X)+b(Y-\mu_Y)\right)^2\right] = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$

is a measure of the spread or dispersion along the $(ax+by)$-direction.

Theorem 9.2.2 Calculating Covariance

Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.

(1) $Var(X) = Cov(X,X)$.

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

(3) $|Cov(X,Y)| \le \sigma_X \cdot \sigma_Y$

"$=$" $\Leftrightarrow X = \mu_X$ with probability 1 or $Y = \mu_Y$ with probability 1 or $Y = aX + b$ with probability 1, where

$$a = \frac{\sigma_{X,Y}}{\sigma_X^2}, b = \mu_Y - \mu_X \cdot \frac{\sigma_{X,Y}}{\sigma_X^2}.$$

If $\sigma_X \ne 0$ and $\sigma_Y \ne 0$, then

$Cov(X,Y) = \sigma_X \cdot \sigma_Y \Leftrightarrow Y = aX + b$ with probability 1, where

$$a = \frac{\sigma_Y}{\sigma_X} > 0, b = \mu_Y - \mu_X \cdot \frac{\sigma_Y}{\sigma_X}.$$

$Cov(X,Y) = -\sigma_X \cdot \sigma_Y \Leftrightarrow Y = aX + b$ with probability 1, where

$$a = -\frac{\sigma_Y}{\sigma_X} < 0, b = \mu_Y + \mu_X \cdot \frac{\sigma_Y}{\sigma_X}.$$

Theorem 9.2.3 Covariance of Two Linear Combined R.V.'s

Suppose $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.

(1) $Cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \, Cov(X_i, Y_j)$.

(2) $Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{1\le i<j\le n} a_i b_j \, Cov(X_i, X_j)$.

In particular, if $X_1, X_2, \ldots, X_n$ are pairwise uncorrelated, then

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i).$$

Theorem 9.2.4 Independence Implies Uncorrelated
Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
If $X \perp Y$, then X and Y are uncorrelated, i.e.,

$$Cov(X,Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0.$$

Remark 9.2.3 Uncorrelated Can't Imply Independence
The inverse is not true, i.e.,

$$Cov(X,Y) = 0 \nRightarrow X \perp Y.$$

Definition 9.2.2 Correlation Coefficient
Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ with $0 < \sigma_X^2 < \infty, 0 < \sigma_Y^2 < \infty$.
The correlation coefficient between X and Y is given by

$$\rho_{X,Y} = Cov(X^*, Y^*) = Cov\left(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\right) = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}.$$

Remark 9.2.4 Properties of Correlation Coefficient
(1) $X^* = \dfrac{X-\mu_X}{\sigma_X}$ is independent of the units in which X is measured.
$\Rightarrow \rho_{X,Y}$ is independent of the units in which X and Y is measured.

(2) $-1 \le \rho_{X,Y} \le 1$.

$\rho_{X,Y} = 1 \Leftrightarrow Y = aX + b$ with probability 1, where

$$a = \frac{\sigma_Y}{\sigma_X} > 0, b = \mu_Y - \mu_X \cdot \frac{\sigma_Y}{\sigma_X}.$$

$\rho_{X,Y} = -1 \Leftrightarrow Y = aX + b$ with probability 1, where

$$a = -\frac{\sigma_Y}{\sigma_X} < 0, b = \mu_Y + \mu_X \cdot \frac{\sigma_Y}{\sigma_X}.$$

§ 9.3 Conditioning on R.V.'s

Definition 9.3.1 Conditional Expectation on R.V.'s
Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
Let $g(y) = E[X|Y = y]$, $\forall y \in \mathbb{R}$.
We denote $E[X|Y]$ as the r.v. $g(Y)$. Note that $E[X|Y]$ is a function of Y.

Theorem 9.3.1 Marginal Expectation
Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
Then $E\big[E[X|Y]\big] = E[X]$.

Theorem 9.3.2 Marginal Expectation of Measurable Function
Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
Then $E\big[E[X \cdot g(Y)|Y]\big] = g(Y)E[X|Y]$.

Theorem 9.3.3 Wald's Equations
Suppose $X_1, X_2, \ldots$ are i.i.d. r.v.'s$\sim X$ and N is a positive integer-valued r.v. of a probability space $(\Omega, \mathcal{A}, P)$, and $N \perp \{X_1, X_2, \ldots\}$.
(1) If $E[X] < \infty$ and $E[N] < \infty$, then

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N] \cdot E[X].$$

(2) If $Var(X) < \infty$ and $Var(N) < \infty$, then

$$Var\left(\sum_{i=1}^{N} X_i\right) = E[N] \cdot Var(X) + (E[X])^2 \cdot Var(N).$$

Theorem 9.3.4 Law of Total Probability
Suppose A is an event and X is a r.v. of a probability space $(\Omega, \mathcal{A}, P)$, then

$$P(A) = \begin{cases} \displaystyle\sum_{x \in X(\Omega)} P(A|X = x) \cdot P_X(x), & \text{if } X \text{ is a discrete r.v.} \\ \displaystyle\int_{-\infty}^{\infty} P(A|X = x) \cdot f_X(x)dx, & \text{if } X \text{ is a continuous r.v.} \end{cases}$$

Theorem 9.3.5 Conditional Variance on R.V.'s
Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$, then

$$Var(X) = E[Var(X|Y)] + Var(E[X|Y]).$$

§ 9.4 Bivariate Normal (Gaussian) Distribution

Definition 9.4.1 Bivariate Normal (Gaussian) R.V.'s

Let X_1 and X_2 be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
We call X_1 and X_2 jointly normal (Gaussian) r.v.'s with parameters
$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} > 0$, where ">0" means positive-definite,

denoted $\underline{X} \sim N\left(\underline{\mu}, \sum\right)$, if their joint p.d.f. is given by

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left[-\frac{1}{2}\left(\underline{x} - \underline{\mu}\right)^T \sum\nolimits^{-1} \left(\underline{x} - \underline{\mu}\right)\right]$$

$$= \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left[-\frac{1}{2}(x_1 - \mu_1,\ x_2 - \mu_2) \frac{1}{|\Sigma|} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right]$$

$$= \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left[-\frac{1}{2}\left(\sigma_{22}(x_1 - \mu_1)^2 - 2\sigma_{12}(x_1 - \mu_1)(x_2 - \mu_2) + \sigma_{11}(x_2 - \mu_2)^2\right)\right]$$

where $|\Sigma| = \det(\Sigma) = \sigma_{11} \cdot \sigma_{22} - \sigma_{12}^2 > 0$.

Such a joint p.d.f. is called a bivariate normal p.d.f. with parameters $\underline{\mu}$ and Σ.

Theorem 9.4.1 Explicitly Normal (Gaussian) R.V.

Suppose X_1 and X_2 are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$,

and suppose $\underline{X} \sim N\left(\underline{\mu}, \sum\right)$.

(1) $X_1 \sim N(\mu_1, \sigma_{11})$ and $X_2 \sim N(\mu_2, \sigma_{22})$. Therefore

$$\mu_1 = \mu_{X_1}, \sigma_{11} = \sigma_{X_1}^2 \triangleq \sigma_1^2, \mu_2 = \mu_{X_2}, \sigma_{22} = \sigma_{X_2}^2 \triangleq \sigma_2^2.$$

(2)

$$X_2|_{X_1 = x_1} \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1), \frac{|\Sigma|}{\sigma_{11}}\right)$$

and

$$X_1|_{X_2 = x_2} \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2), \frac{|\Sigma|}{\sigma_{22}}\right).$$

(3) $\sigma_{12} = \sigma_{X_1, X_2} = \rho_{X_1, X_2} \cdot \sigma_{X_1}\sigma_{X_2} \triangleq \rho \cdot \sigma_1 \sigma_2$.

Therefore

$$X_2|_{X_1 = x_1} \sim N\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right)$$

and

$$X_1|_{X_2 = x_2} \sim N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - \rho^2)\sigma_1^2\right).$$

Remark 9.4.1 Mean Vector and Covariance Matrix

$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ is called the **mean vector** of $\underline{X}$,

and $\sum = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ is called the **covariance matrix** of $\underline{X}$.

Lemma 9.4.1 Linear Conditional Expectation and Constant Variance

Suppose X_1 and X_2 are jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$ with $\mu_{X_1} = \mu_1,\ \mu_{X_2} = \mu_2, \sigma_{X_1}^2 = \sigma_1^2, \sigma_{X_2}^2 = \sigma_2^2, \rho_{X_1,X_2} = \rho$.

(1) If $E[X_2|X_1 = x_1] = ax_1 + b$ is a linear function in x_1, then

$$E[X_2|X_1 = x_1] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1).$$

(2) If $E[X_2|X_1 = x_1] = ax_1 + b$ is a linear function in x_1,
and $Var(X_2|X_1 = x_1) = \sigma^2$ is a constant, then

$$Var(X_2|X_1 = x_1) = (1 - \rho^2)\sigma_2^2.$$

Theorem 9.4.2 Derivation of Jointly Normal R.V.'s

Suppose X_1 and X_2 are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
Suppose

(1) X_1 is a normal r.v.

(2) $X_2|\,X_1 = x_1$ is a normal r.v. for all $x_1 \in \mathbb{R}$.

(3) $E[X_2|X_1 = x_1]$ is a linear function in X_1, and $Var(X_2|X_1 = x_1) = \sigma^2$ is a constant.

Then X_1 and X_2 are **jointly normal** r.v.'s.

Theorem 9.4.3 Independence mutually Implies Uncorrelated

Suppose X_1 and X_2 are jointly normal r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
Then X_1 and X_2 are independent $\Leftrightarrow$ X_1 and X_2 are uncorrelated.

Theorem 9.4.4 Linearly Generated Normal R.V.

Suppose $\underline{X} \sim N\left(\underline{\mu_X}, \sum_{\underline{X}}\right)$ and $\underline{Y} = A\underline{X} + b$, where A is **nonsingular**, i.e., $|A| \neq 0$.

Then

$$\underline{Y} \sim N\left(A\underline{\mu_X} + b, A\sum\nolimits_{\underline{X}} A^T\right).$$

Chapter.10 Sums of Independent R.V.'s and Limit Theorems

§ 10.1 Moment Generating Functions

Definition 10.1.1 Moment Generating Function
The moment generating function (m.g.f.) $M_X(t)$ of a r.v. X is given by
$M_X(t) = E[e^{tx}]$, if $\exists\delta > 0 \Rightarrow M_X(t)$ is defined for all $t \in (-\delta, \delta)$.

Theorem 10.1.1 Moment Generation

(1) $E[X^n] = M_X^{(n)}(0),\ \forall n \geq 0.$

(2) Maclaurin's series for $M_X(t)$:

$$M_X(t) = \sum_{n=0}^{\infty} \frac{M_X^{(n)}(0)}{n!} t^n = \sum_{n=0}^{\infty} \frac{E[X^n]}{n!} t^n.$$

Remark 10.1.1 Sufficient Condition for n^{th} Moment to Converge
If $|M_X(t)| < \infty$ for some $t > 0$, then $|E[X^n]| < \infty$ for all $n \geq 1$.
But the converse is not true.

Theorem 10.1.2 Same M.G.F. Implies Same C.D.F.
If $M_X(t) = M_Y(t)$ for all $t \in (-\delta, \delta)$ for some $\delta > 0$,
then the c.d.f. of X and Y are the same.

X	$E[X]$	$Var(X)$	$M_X(t)$
Bernoulli(p): $p_X(i) = \begin{cases} 1-p \triangleq q, & \text{if } i = 0 \\ p, & \text{if } i = 1 \\ 0, & \text{o.w.} \end{cases}$	p	pq	$pe^t + q, t \in \mathbb{R}$
binomial(n,p): $p_X(i) = \begin{cases} \binom{n}{i} p^i q^{n-i}, & \text{if } i = 0,1,2,\dots,n. \\ 0 \quad , & \text{o.w.} \end{cases}$	np	npq	$(pe^t + q)^n, t \in \mathbb{R}$
geometric(p): $p_X(i) = \begin{cases} (1-p)^{i-1} \cdot p\,, & i = 0,1,2\dots \\ 0, & \text{o.w.} \end{cases}$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}, t < \ln\left(\frac{1}{q}\right)$
neg.-binomial(r,p): $p_X(i) = \begin{cases} \binom{i-1}{r-1} p^r (1-p)^{i-r}\,, & i = r, r+1, \dots \\ 0, & \text{o.w.} \end{cases}$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$\left(\frac{pe^t}{1-qe^t}\right)^r, t < \ln\left(\frac{1}{q}\right)$
Poisson(λ): $p_X(i) = \begin{cases} e^{-\lambda}\frac{\lambda^i}{i!}, & i = 0,1,2\dots \\ 0, & \text{o.w.} \end{cases}$	λ	λ	$e^{\lambda(e^t-1)}, t \in \mathbb{R}$
$U(\alpha,\beta)$: $f_X(x) = \begin{cases} \frac{1}{\beta-\alpha}, & \text{if } \alpha < x < \beta \\ 0, & \text{o.w.} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\begin{cases} \frac{e^{bt}-e^{at}}{(b-a)^t}, & \text{if } t \neq 0 \\ 1, & \text{if } t = 0 \end{cases}$
$g(\alpha,\lambda)$: $f_X(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & \text{if } \alpha > 0 \\ 0, & \text{o.w.} \end{cases}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^\alpha, t < \lambda$
$N(\mu,\sigma^2)$: $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x \in \mathbb{R}$	μ	σ^2	$\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), t \in \mathbb{R}$
$\mathcal{L}(\lambda)$: $f_X(x) = \frac{1}{2\lambda}\exp\left(-\frac{\lvert x\rvert}{\lambda}\right), x \in \mathbb{R}$	0	$2\lambda^2$	$\frac{1}{1-\lambda^2 t^2}, -\frac{1}{\lambda} < t < \frac{1}{\lambda}$

§ 10.2 Sums of Independent R.V.'s

Theorem 10.2.1 M.G.F. of Sums of Independent R.V.'s

Suppose $X_1, X_2, \dots, X_n$ are **independent** r.v.'s with m.g.f.'s $M_{X_1}(t), M_{X_2}(t), \dots, M_{X_n}(t)$ respectively.
Then the m.g.f. of their **sum** $X = X_1 + X_2 + \cdots + X_n$ is

$$M_X(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

Theorem 10.2.2 M.G.F. of Sums of Normal R.V.'s

Suppose $X_1, X_2, \dots, X_n$ are **independent** r.v.'s and $X_i \sim N(\mu_i, \sigma_i^2), \forall i = 1,2, \dots$ and suppose $a_1, a_2, \dots \in \mathbb{R}$.

If $X = \sum_{i=1}^{n} a_i X_i$, then $X \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$.

Corollary 10.2.1 M.G.F. of Sums of I.I.D. Normal R.V.'s

Suppose $X_1, X_2, \dots, X_n$ are **i.i.d.** $\sim N(\mu, \sigma^2)$, then

$$S_n = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2), \text{ and } \overline{X} = \frac{S_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

§ 10.3 Markov and Chebyshev Inequalities

Theorem 10.3.1 Markov's Inequality
Suppose X is a nonnegative r.v., then

$$P(X \geq t) \leq \frac{E[X]}{t}, \forall t > 0.$$

Theorem 10.3.2 Chebyshev's Inequality

$$P(|X - \mu_X| \geq t) \leq \frac{\sigma_X^2}{t^2}, \forall t > 0.$$

In particular,

$$P(|X - \mu_X| \geq k \cdot \sigma_X) \leq \frac{1}{k^2}, \forall k > 0.$$

Remark 10.3.1 Not Tight Bounds
The bounds obtained by Markov and Chebyshev inequalities usually **not very tight**.

Theorem 10.3.3 Zero Absolute Moment

$$E[|X|] = 0 \Leftrightarrow X = 0 \text{ with probability 1.}$$

Corollary 10.3.1 Zero Variance

$$Var(X) = 0 \Leftrightarrow X = 0 \text{ with probability 1.}$$

Theorem 10.3.4 Chebyshev's Inequality for I.I.D R.V.'s
Suppose $X_1, X_2, \dots, X_n$ are **i.i.d.** r.v.'s with mean μ and variance $\sigma^2 < \infty$.

Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the **sample mean** of $X_1, X_2, \dots, X_n$.

Then

$$P(|\overline{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Theorem 10.3.5 Chebyshev's Inequality for I.I.D. Bernoulli R.V.'s
Suppose $X_1, X_2, \dots, X_n$ are i.i.d. $\sim$ Bernoulli(p).

Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean of $X_1, X_2, \dots, X_n$.

Then

$$P(|\overline{X} - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

§ 10.4 Laws of Large Numbers (LLN's)

Definition 10.4.1 Converge in Probability
Let $X, X_1, X_2, \dots$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
We say that X_n converges to X **in probability**, denoted

$$X_n \xrightarrow{P} X,$$

If $\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1, \forall \varepsilon > 0$ or $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0, \forall \varepsilon > 0$.

Theorem 10.4.1 Weak Law of Large Numbers (WLLN)
Suppose $X_1, X_2, \dots$ are i.i.d. r.v.'s with mean μ and variance $\sigma^2 < \infty$.
Then $\overline{X_n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu$, i.e., $\lim_{n\to\infty} P\left(\left|\overline{X_n} - \mu\right| > \varepsilon\right) = 0, \forall \varepsilon > 0$.

Remark 10.4.1 Relative Frequency Converges to Probability in Probability
Let an experiment be repeated independently and let $n(A)$ be the number of times an event A occurs in the first n repetitions of the experiment.
Let $X_i = \begin{cases} 1, & \text{if } A \text{ occurs on the } i^{th} \text{ repetition} \\ 0, & \text{o.w.} \end{cases}$

$\Rightarrow n(A) = \sum_{i=1}^{n} X_i$ and $E[X_i] = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A)$.

$\Rightarrow \lim_{n\to\infty} P\left(\left|\dfrac{n(A)}{n} - P(A)\right| > \varepsilon\right) = \lim_{n\to\infty} P\left(\left|\dfrac{1}{n}\sum_{i=1}^{n} X_i - P(A)\right| > \varepsilon\right) = 0.$

$\therefore$ The **relative frequency** $\dfrac{n(A)}{n}$ of occurrence of A is very likely close to $P(A)$ if n is sufficiently large.

Definition 10.4.2 Converge Almost Surely
Let $X, X_1, X_2, \cdots$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.
We say that X_n converges to X **almost surely** (a.s.), denoted

$$X_n \xrightarrow{\text{a.s.}} X,$$

if $P\left(\lim_{n\to\infty} X_n = X\right) = 1$.

Theorem 10.4.2 Strong Law of Large Numbers (SLLN)

Suppose $X_1, X_2, \ldots$ are i.i.d. r.v.'s with mean μ.

Then $\overline{X_n} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mu$, i.e., $P\left(\lim_{n\to\infty} \overline{X_n} = \mu\right) = 1$.

Remark 10.4.2 Relative Frequency Converges Almost Surely

$$P\left(\lim_{n\to\infty} \frac{n(A)}{n} = P(A)\right) = 1 \Rightarrow \lim_{n\to\infty} \frac{n(A)}{n} = P(A) \text{ with probability 1.}$$

Theorem 10.4.3 Converge Almost Surely Implies Convergence in Probability

If $X_n \xrightarrow{\text{a.s.}} X$, then $X_n \xrightarrow{P} X$.

§ 10.5 Central Limit Theorem (CLT)

Theorem 10.5.1 Levy Continuity Theorem

Suppose $X, X_1, X_2, \ldots$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, P)$.

If $\exists\delta > 0 \Rightarrow \lim_{n\to\infty} M_{X_n}(t) = M_X(t), \forall t \in (-\delta, \delta)$,

then $\lim_{n\to\infty} F_n(x) = F(x)$ if $F(x)$ is continuous at X.

Theorem 10.5.2 Central Limit Theorem (CLT)

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. r.v.'s with mean μ and variance σ^2.

Let $S_n^* = \dfrac{X_1 + X_2 + \cdots + X_n - E[S_n]}{\sigma_{S_n}} = \dfrac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$.

Then

$$\lim_{n\to\infty} F_{S_n^*}(x) = \Phi(x),$$

i.e.,

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \le x\right) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Equivalently,

$$\lim_{n\to\infty} P\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le x\right) = \lim_{n\to\infty} P\left(\frac{\overline{X} - \mu}{\sqrt{\frac{Var(X)}{n}}} \le x\right) = \lim_{n\to\infty} P\left(\frac{\overline{X} - E[\,\overline{X}\,]}{\sigma_{\overline{X}}} \le x\right) = \Phi(x).$$