Probability

Jingxuan Yang

September 11, 2024

Contents

Preface

Currently, most of the material in this book comes from the lecture notes of Professor Jay Cheng. The proofs of theorems are not included yet, which warrants further efforts in the future.

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 $\varnothing$ 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 subsets of Ω such that*

(i) $\Omega \in \mathcal{A}$,

(ii) $\mathcal{A}$ *is closed under complementation, i.e., if* $A \in \mathcal{A}$, *then* $\Omega \setminus A \in \mathcal{A}$,

(iii) $\mathcal{A}$ *is closed under countable union, i.e., if* $A_n \in \mathcal{A}$ *for* $n = 1, 2, \ldots,$ *then* $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Theorem 1.1 (Properties of σ-algebra)**.** *Suppose $\mathcal{A}$ is a σ-algebra of subsets of a sample space Ω.*

(i) $\varnothing \in \mathcal{A}$,

(ii) $\mathcal{A}$ *is closed under finite union,*

(iii) $\mathcal{A}$ *is closed under countable and finite intersection.*

Proof. abc $\square$

Theorem 1.2 (Intersection of σ-algebras)**.** *Suppose Γ is a nonempty collection of σ-algebras of subsets of a sample space Ω. Then the intersection $\mathcal{B} = \cap_{\mathcal{A} \in \Gamma} \mathcal{A}$ of the σ-algebras in Γ is also a σ-algebra of subsets of Ω.*

Proof. abc $\square$

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 $\mathcal{C}$.*

Proof. abc □

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 denote it as $\sigma(\mathcal{C})$.*

Definition 1.5 (Probability Measure)**.** *Let $\mathcal{A}$ be a σ-algebra of subsets of a sample space Ω, a probability measure $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ on $\mathcal{A}$ is a real-valued function on $\mathcal{A}$ such that*

(i) Nonnegativity: $\mathbb{P}(A) \geqslant 0$, $\forall A \in \mathcal{A}$,

(ii) Normalization: $\mathbb{P}(\Omega) = 1$,

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

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

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

Definition 1.6 (Probability Space)**.** *A probability space is an ordered triple $(\Omega, \mathcal{A}, \mathbb{P})$ consisting of a sample space Ω, a σ-algebra $\mathcal{A}$ of subsets of Ω, and a probability measure $\mathbb{P}$ on $\mathcal{A}$.*

Theorem 1.3 (A Kind of Probability Measure)**.** *Suppose $\Omega = \{\omega_1, \omega_2, \ldots\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ and $\mathbb{P}(A) = \sum_{\omega_i \in \mathcal{A}} P_i$, for all $A \in \mathcal{P}(\Omega)$, where $P_i \geqslant 0$, $\forall i = 1, 2, \ldots$, and $\sum_{i=1}^{\infty} P_i = 1$, then $\mathbb{P}$ is a probability measure on $\mathcal{P}(\Omega)$. A similar result holds if $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$, where $N \geqslant 1$.*

Proof. abc □

Corollary 1.2 (A Kind of Probability Measure (special))**.** *Suppose $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, and $\mathbb{P}(A) = \frac{|A|}{N}$ for all $A \in \mathcal{P}(\Omega)$, then $\mathbb{P}$ is a probability measure on $\mathcal{P}(\Omega)$.*

Proof. abc □

Theorem 1.4 (Classical definition of probability)**.** *Suppose $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ and $\mathbb{P}$ is a probability measure on $\mathcal{P}(\Omega)$ such that $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \cdots = \mathbb{P}(\omega_N)$, then $\mathbb{P}(A) = \frac{|A|}{N}$ for all $A \in \mathcal{P}(\Omega)$.*

Proof. abc □

Theorem 1.5 (Properties of Probability Measure)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space.*

(i) $\mathbb{P}(\varnothing) = 0$.

(ii) $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$. *Therefore,* $0 \leqslant \mathbb{P}(A) \leqslant 1$, *for all* $A \in \mathcal{A}$.

(iii) Finite additivity: If $A_1, A_2, \ldots, A_N$ are pairwise disjoint events in A, then

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

Proof. abc $\square$

Theorem 1.6 (Properties of Probability Measure)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and suppose $A, B \in \mathcal{A}$.*
(1) If $A_1, A_2, \cdots$ are pairwise disjoint events on A and

$$\bigcup_{n=1}^{\infty} A_n = \Omega,$$

then

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}\left(A \cap A_n\right).$$

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

Proof. abc $\square$

Corollary 1.3 (Finite Additivity under Union)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, $A \in \mathcal{A}$, $A_1, A_2, \cdots$ are pairwise disjoint events in $\mathcal{A}$, and*

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = 1,$$

then

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

Proof. abc $\square$

Theorem 1.7 (Inclusion-exclusion Identity)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and suppose $A_1, A_2, \cdots, A_n \in \mathcal{A}$, where $n \geqslant 2$, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \cdot \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right).$$

Proof. abc $\square$

Lemma 1.1 (Generated Pairwise Disjoint)**.** *Suppose $\mathcal{A}$ is a σ-algebra of subsets of a sample space Ω, suppose $A_1, A_2, \cdots \in \mathcal{A}$, $B_1 = A_1$, and*

$$B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i$$

for all $n \geqslant 2$, then $B_1, B_2, \cdots$ are pairwise disjoint events in $\mathcal{A}$,

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

for all $n \geqslant 1$, and

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n.$$

Proof. abc □

Theorem 1.8 (Inclusion-exclusion Inequality)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and suppose $A_1, A_2, \cdots, A_n \in \mathcal{A}$, where $n \geqslant 2$, then*

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

where $1 \leqslant m \leqslant n$.
In particular,

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

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \geqslant \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leqslant i < j \leqslant n} \mathbb{P}\left(A_i \cap A_j\right).$$

Proof. abc □

Theorem 1.9 (Boole's Inequality)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and suppose $A_1, A_2, \cdots \in \mathcal{A}$, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Proof. abc □

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

Definition 1.8 (Limit of Events)**.** *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space.*
(1) The limit $\lim_{n\to\infty} A_n$ *of an increasing sequence* $\{A_1, A_2, \cdots\}$ *of events in* A *is the event that at least one of the events occurs, i.e.,*

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

(2) The limit $\lim_{n\to\infty} A_n$ *of a decreasing sequence* $\{A_1, A_2, \cdots\}$ *of events in* A *is the event that all the events occur, i.e.,*

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Theorem 1.10 (Continuity of Probability Measure)**.** *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space.*
(1) Suppose that $\{A_1, A_2, \cdots\}$ *is an increasing sequence of events in* A*. Then*

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(2) Suppose that $\{A_1, A_2, \cdots\}$ *is a decreasing sequence of events in* A*. Then*

$$\mathbb{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Proof. abc □

Remark 1.1 (Not Necessary)**.** *If* $\mathbb{P}(A) = 0$*, then it is not necessary that* $A = \varnothing$*, e.g.,* $\Omega = (0,1)$ *and* $A = A_\alpha,\ \alpha \in (0,1)$*. If* $\mathbb{P}(A) = 1$*, 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*

$$\mathbb{P} = \frac{\beta - \alpha}{b - a}.$$

Proof. abc □

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* $\mathcal{B}_{(a,b)}$*.*

Theorem 1.12 (Existence of Probability Measure)**.** *For any interval* (a,b)*, there exists a unique probability measure* $\mathbb{P}$ *on* $\mathcal{B}_{(a,b)}$ *s.t.,*

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

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

Proof. abc □

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.*

Proof. abc $\square$

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 $0 \leqslant r \leqslant n$.*

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 $n = n_1 + n_2 + \cdots + n_k$.

Proof. abc $\square$

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)**.**

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

Proof. abc □

Theorem 2.4 (Multinomial Expansion)**.**

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

Proof. abc □

Corollary 2.1 (Binomial Expansion)**.**

$$(x + y)^n = \sum_{i=1}^{n} \binom{n}{i} x^i y^{n-i},\ \forall n \geqslant 0.$$

Theorem 2.5 (Stirling's Formula)**.**

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

Therefore,

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

Proof. abc □

Chapter 3

Conditional Probability and Independence

Definition 3.1 (Conditional Probability)**.** *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space, and* $A, B \in \mathcal{A}$. *The conditional probability of* A *given* B*, denoted* $\mathbb{P}(A|B)$*, is given by*

$$\mathbb{P}(A|B) = \begin{cases} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, & \textit{if } \mathbb{P}(B) > 0, \\ 0, & \textit{if } \mathbb{P}(B) = 0. \end{cases}$$

Remark 3.1 (Property of Conditional Probability)**.**

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

Proof. abc □

Theorem 3.1 (Conditional Probability Space)**.** *Suppose* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space, and suppose* $\mathbb{P}(B) > 0$*, for some* $B \in \mathcal{A}$*. Then the conditional probability function* $\mathbb{P}(\cdot|B) : \mathcal{A} \to \mathbb{R}$ *is a probability measure on* $\mathcal{A}$*, and hence* $(\Omega, \mathcal{A}, \mathbb{P}(\cdot|B))$ *is a probability space.*

Proof. abc □

Theorem 3.2 (Reduction of Probability Space)**.** *Suppose* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space, and suppose* $\mathbb{P}(B) > 0$*, for some* $B \in \mathcal{A}$*. Let* $\mathcal{A}_B : \{A \in \mathcal{A} : A \subseteq B\}$ *and* $P_B(A) = \mathbb{P}(A|B)$ *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.*

Proof. abc □

Remark 3.2 (Conversion of Reduced and Conditional Probability Space)**.** *Note that* $\mathbb{P}(A|B) = \mathbb{P}(A \cap B|B) = P_B(A \cap B), \forall A \in \mathcal{A}$*. And* $\mathbb{P}(A|B) = P_B(A)$*, if* $A \in \mathcal{A}$ *and* $A \subseteq B$*.*

Proof. abc □

Theorem 3.3 (Law of Multiplication)**.** *Suppose* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space, and* $A_1, A_2, \ldots, A_n \in \mathcal{A}$*. Then*

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

Proof. abc □

Theorem 3.4 (Law of Total Probability (infinite))**.** *Suppose* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space, and suppose* $B_1, B_2, \cdots \in \mathcal{A}$ *are pairwise disjoint and* $\bigcup_{n=1}^{\infty} B_n = \Omega$*. Then,*
(1) $\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \cdot \mathbb{P}(A|B_n), \forall A \in \mathcal{A}$.
(2) $\mathbb{P}(A|B) = \sum_{n=1}^{\infty} \mathbb{P}(B_n|B) \cdot \mathbb{P}(A|B \cap B_n), \forall A, B \in \mathcal{A}$.

Proof. abc □

Corollary 3.1 (Law of Total Probability (finite))**.** *Suppose* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space, and suppose* $B_1, B_2, \cdots B_n \in \mathcal{A}$ *are pairwise disjoint and* $\bigcup_{i=1}^{n} B_i = \Omega$*. Then,*
(1) $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(B_i) \cdot \mathbb{P}(A|B_i), \forall A \in \mathcal{A}$.
(2) $\mathbb{P}(A|B) = \sum_{i=1}^{n} \mathbb{P}(B_i|B) \cdot \mathbb{P}(A|B \cap B_i), \forall A, B \in \mathcal{A}$.

Proof. abc □

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

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

Proof. abc □

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

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

Proof. abc □

Theorem 3.6 (Properties of Conditional Probability)**.** *Suppose* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space, and suppose* $A, B \in \mathcal{A}$*.*

(1) $\mathbb{P}(A|B) > \mathbb{P}(A) \Leftrightarrow \mathbb{P}(A \cap B) > \mathbb{P}(A) \cdot \mathbb{P}(B) \Leftrightarrow \mathbb{P}(B|A) > \mathbb{P}(B)$
(2) $\mathbb{P}(A|B) < \mathbb{P}(A), \mathbb{P}(B) > 0 \Leftrightarrow \mathbb{P}\left(A \cap B\right) < \mathbb{P}(A) \cdot \mathbb{P}(B)$
$\Leftrightarrow \mathbb{P}(B|A) < \mathbb{P}(B), \mathbb{P}(A) > 0$
(3) $\mathbb{P}(A|B) = \mathbb{P}(A) \to \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$
$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cap \mathbb{P}(B),\ \mathbb{P}(A) = 0$ *or* $\mathbb{P}(B) > 0 \to \mathbb{P}(A|B) = \mathbb{P}(A)$
If $\mathbb{P}(A) = 0$ *or* $\mathbb{P}(B) > 0$*, then* $\mathbb{P}(A|B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Proof. abc □

Definition 3.2 (Independence). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $A, B \in \mathcal{A}$. If $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$, 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 $\mathbb{P}(A|B) > \mathbb{P}(A)$, then A and B are said to be positively correlated, and if $\mathbb{P}(A|B) < \mathbb{P}(A)$, then A and B are said to be negatively correlated.*

Theorem 3.7 (Properties of Independence). *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and suppose $A, B \in \mathcal{A}$.*

(1) If $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$, then $A \perp B,\ \forall B \in \mathcal{A}$.

(2) If $A \subseteq B$ and $A \perp B$, then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 1$.

(3) If A and B are disjoint and $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, then they are dependent.

Proof. abc □

Theorem 3.8 (Independence of Two Events). *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and suppose $A, B \in \mathcal{A}$, and $A \perp B$.*

Then $A^* \perp B^*$, i.e., $\mathbb{P}(A^* \cap B^*) = \mathbb{P}(A^*) \cdot \mathbb{P}(B^*), \forall A^* = A,\ A^c;\ B^* = B,\ B^c$.

Proof. abc □

Corollary 3.3 (Conditional Probability with Independence). *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and suppose $A, B \in \mathcal{A}$, and $A \perp B$.*

If $\mathbb{P}(B) > 0$, then $\mathbb{P}(A^*|B) = \mathbb{P}(A^*), \forall A^* = A,\ A^c$.
If $\mathbb{P}(B) < 1$, then $\mathbb{P}(A^*|B^c) = \mathbb{P}(A^*), \forall A^* = A,\ A^c$.

Proof. abc □

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

If $A \perp B$ and $\mathbb{P}(B) < 1$, then knowledge about the occurrence of B^c does not change the probability of the occurrence of A^*.

Proof. abc □

Definition 3.3 (Independent Set). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $A_1, A_2, \cdots, A_n \in \mathcal{A}$, where $n \geqslant 2$.*
If $\mathbb{P}\left(A_{i_1} \bigcap A_{i_2} \bigcap \cdots \bigcap A_{i_k}\right) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}), \forall 2 \leqslant k \leqslant n$,

$$\# = \sum_{k=2}^{n} \binom{n}{k} = 2^n - n - 1, 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n, \# := \textit{number}.$$

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

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

Proof. abc $\square$

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

(1) $A_1, A_2, \cdots, A_n$ *are independent.*

(2) $\mathbb{P}\left(A_{i_1}^* \bigcap A_{i_2}^* \bigcap \cdots \bigcap A_{i_k}^*\right) = \mathbb{P}\left(A_{i_1}^*\right) \mathbb{P}\left(A_{i_2}^*\right) \cdots \mathbb{P}\left(A_{i_k}^*\right),\ \forall 2 \leqslant k \leqslant n,$ $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n,\ A_{i_r}^* = A_{i_r}$ *or* $A_{i_r}^c$.

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

Proof. abc $\square$

Definition 3.4 (Independent Set)**.** *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *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* $(\Omega, \mathcal{A}, \mathbb{P})$ *is a probability space, and suppose* $A_1, A_2, \cdots, A_n \in \mathcal{A}$ *are independent. Then*

$$\begin{aligned}
&\mathbb{P}\left[\left(A_{i_1}^* \bigcap A_{i_2}^* \bigcap \cdots \bigcap A_{i_k}^*\right) \bigcap \left(A_{j_1}^* \bigcap A_{j_2}^* \bigcap \cdots \bigcap A_{j_l}^*\right)\right] \\
=&\mathbb{P}\left(A_{i_1}^* \bigcap A_{i_2}^* \bigcap \cdots \bigcap A_{i_k}^*\right) \cdot \mathbb{P}\left(A_{j_1}^* \bigcap A_{j_2}^* \bigcap \cdots \bigcap A_{j_l}^*\right)
\end{aligned}$$

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

In particular, if $\mathbb{P}\left(A_{j_1}^* \bigcap A_{j_2}^* \bigcap \cdots \bigcap A_{j_l}^*\right) > 0$*, for some* $1 \leqslant l \leqslant n-1,\ 1 \leqslant j_1, \cdots, j_l \leqslant n$ *distinct, and* $A_{j_r}^* = A_{j_r}$ *or* $A_{j_r}^c,\ r = 1, 2, \cdots, l$*. Then*

$$\begin{aligned}
&\mathbb{P}\left[\left(A_{i_1}^* \bigcap A_{i_2}^* \bigcap \cdots \bigcap A_{i_k}^*\right) \middle| \left(A_{j_1}^* \bigcap A_{j_2}^* \bigcap \cdots \bigcap A_{j_l}^*\right)\right] \\
=&\mathbb{P}\left(A_{i_1}^* \bigcap A_{i_2}^* \bigcap \cdots \bigcap A_{i_k}^*\right)
\end{aligned}$$

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

Proof. abc $\square$

Chapter 4

Distribution Functions and Discrete Random Variables

4.1 Random Variables

Definition 4.1 (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 4.2 (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* $f^{-1}(B) \in \mathcal{A}_1, \forall B \in \mathcal{A}_2$, *where* $f^{-1}(B) = \{x \in \Omega : f(x) \in B\}$ *is the pre-image of* B *under* f.

Lemma 4.1 (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.

Proof. abc $\square$

Theorem 4.1 (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* $\sigma(\mathcal{C}) \subseteq \{B \subseteq \Omega_2 : f^{-1}(B) \in \mathcal{A}_1\}$.

Proof. abc $\square$

Corollary 4.1 (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* $(\Omega_2, \mathcal{A}_2)$.

Proof. abc $\square$

Theorem 4.2 (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* $(\Omega_3, \mathcal{A}_3)$.

Proof. abc □

Definition 4.3 (Open Set). *A set A in $\mathbb{R}^n$ is called an open set in $\mathbb{R}^n$ if for all* $\boldsymbol{x} \in A, \exists r > 0 \to \mathcal{B}_{\boldsymbol{x}}(r) \subseteq A$, *where* $\mathcal{B}_{\boldsymbol{x}}(r) = \{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < r\}$.

Definition 4.4 (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 4.3 (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* $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$.

Proof. abc □

Definition 4.5 (Cell). *A cell in $\mathbb{R}$ is a finite interval of the form* (a, b), $[a, b)$, $(a, b]$, *or* $[a, b]$ *for some* $a \leqslant b$. *A cell I in $\mathbb{R}^n$, where $n \geqslant 1$, is a Cartesian product of n cells* $I_1, I_2, \cdots, I_n$ *in $\mathbb{R}$, i.e.,* $I = I_1 \times I_2 \times \cdots \times I_n$.

Definition 4.6 (Open Cube). *Let* $\boldsymbol{x} \in \mathbb{R}^n$, $l > 0$, *and* $I_i = \left(x_i - \frac{l}{2}, x_i + \frac{l}{2}\right), \forall 1 \leqslant i \leqslant n$. *The open cube $C_{\boldsymbol{x}}(l)$ in $\mathbb{R}^n$ with center $\boldsymbol{x}$ 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 4.4 (Set from Cells). *Every open set in $\mathbb{R}^n$ is a countable union of open cells in $\mathbb{R}^n$.*

Proof. abc □

Theorem 4.5 (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* $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

Proof. abc □

Theorem 4.6 (Components of Measurable Function). *Suppose $(\Omega, \mathcal{A})$ is a measurable space,* $f = (f_1, f_2, \cdots, f_n)$ *is a function from Ω to $\mathbb{R}^n$. Then f is a measurable function from* $(\Omega, \mathcal{A})$ *to* $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}) \Leftrightarrow f_1, f_2, \cdots, f_n$ *are measurable functions from* $(\Omega, \mathcal{A})$ *to* $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

Proof. abc □

Theorem 4.7 (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* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Proof. abc □

Theorem 4.8 (Limit of Measurable Functions). *Suppose that $f_1, f_2, \cdots$ 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 $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.*

Proof. abc $\square$

Theorem 4.9 (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 $\mathbb{R}$,*

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

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 $i = 1, 2, \cdots, 9$.

Proof. abc $\square$

Theorem 4.10 (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*

$$P_f(B) = P[f^{-1}(B)], \; \forall B \in \mathcal{A}_2$$

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.

Proof. abc $\square$

Remark 4.1 (Conventional Denotation). *(1) The set $f^{-1}(B)$ is conventionally denoted as $f \in B$. Therefore $P_f(B) = P[f^{-1}(B)] = \mathbb{P}(f \in B)$, $\forall B \in \mathcal{A}_2$.*
(2) If $B \in \mathcal{A}_2$, then $f^{-1}(B) = f^{-1}[B \cap f(\Omega_1)]$, and hence $P_f(B) = \mathbb{P}(f \in B) = P[f^{-1}(B)] = P[f^{-1}(B \cap f(\Omega_1)] = P[f \in (B \cap f(\Omega_1))] = P_f(B \cap f(\Omega_1))$.

Proof. abc $\square$

Definition 4.7 (Random Variable). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ 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 $(\Omega, \mathcal{A}, \mathbb{P})$.*
A measurable function $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ from $(\Omega, \mathcal{A})$ to $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is called a random vector (r.vect.) of the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Remark 4.2 (Conventional Denotation of Random Variable). *If X is a r.v. of the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then $P_X(B) = P[X^{-1}(B)] = \mathbb{P}(X \in B) = P[\{w \in \Omega : X(w) \in B\}]$, $\forall B \in \mathcal{B}_{\mathbb{R}}$.*

Proof. abc $\square$

Theorem 4.11 (Additivity of Countable Points)**.** *Suppose $\boldsymbol{X}$ is a r.vect. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and B is a "countable" subset of $\mathbb{R}^n$, then $B \in \mathcal{B}_{\mathbb{R}}$, and*

$$P_{\boldsymbol{X}}(B) = \mathbb{P}(\boldsymbol{X} \in B) = \sum_{\boldsymbol{x} \in B} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \sum_{\boldsymbol{x} \in B} P_{\boldsymbol{X}}(\{\boldsymbol{x}\}).$$

Proof. abc □

4.2 Distribution Functions

Definition 4.8 (Cumulative Distribution Function)**.** *Let X be a r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. 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*

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

Theorem 4.12 (Properties of C.D.F)**.** *Suppose X is a r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
(1) F_X is increasing.
(2) $F_X(+\infty) := \lim_{t \to +\infty} F_X(t) = 1$.
(3) $F_X(-\infty) := \lim_{t \to -\infty} F_X(t) = 0$.
(4) $F_X(t+) = \mathbb{P}(X \leqslant t) = F_X(t)$. $F_X(t)$ is right continuous.
(5) $F_X(t-) = \mathbb{P}(X < t)$.

Proof. abc □

Corollary 4.2 (More Properties of C.D.F)**.** *Suppose X is a r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
(1) $\mathbb{P}(X \leqslant a) = F_X(a)$, $\mathbb{P}(X > a) = 1 - F_X(a)$.
(2) $\mathbb{P}(X < a) = F_X(a-)$, $\mathbb{P}(X \geqslant a) = 1 - F_X(a-)$.
(3) $\mathbb{P}(X = a) = F_X(a) - F_X(a-)$.
(4) $\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a)$, $\quad \mathbb{P}(a \leqslant X \leqslant b) = F_X(b) - F_X(a-)$,
$\mathbb{P}(a < X < b) = F_X(b-) - F_X(a)$, $\quad \mathbb{P}(a \leqslant X < b) = F_X(b-) - F_X(a-)$.

Proof. abc □

Theorem 4.13 (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 $(\Omega, \mathcal{A}, \mathbb{P})$, s.t. the c.d.f. F_X of X is equal to F. We call such function a c.d.f.

Proof. abc □

4.3 Discrete Random Variables

Definition 4.9 (Discrete R.V.). *A r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a discrete r.v. if $X(\Omega) = \{X(w) : w \in \Omega\}$ is countable.*

Definition 4.10 (Probability Mass Function). *Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ s.t. $X(\Omega) = \{x_1, x_2, \cdots\}$. 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 $p_X(x) = P_X(\{X = x\}) = \mathbb{P}(X = x),\ \forall x \in \mathbb{R}$.*

Theorem 4.14 (Properties of P.M.F). *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then,*
(1) $p_X(x) \geqslant 0,\ \forall x \in X(\Omega)$.
(2) $p_X(x) = 0,\ \forall x \in \mathbb{R} \setminus X(\Omega)$.
(3) $\displaystyle\sum_{x \in X(\Omega)} p_X(x) = 1$.
Therefore if $X(\Omega) = \{x_1, x_2, \cdots\}$, then,
(1) $p_X(x_i) \geqslant 0,\ \forall i = 1, 2, \ldots$.
(2) $p_X(x) = 0, \forall x \in \mathbb{R} \setminus \{x_1, x_2, \cdots\}$.
(3) $\displaystyle\sum_{i=1}^{\infty} p_X(x_i) = 1$.

Proof. abc □

Theorem 4.15 (Existence of P.M.F). *Suppose $p : \mathbb{R} \to [0,1]$ is a function s.t.*
(1) $p(x_i) \geqslant 0\ \forall i = 1, 2, \ldots$
(2) $p(x) = 0,\ \forall x \in \mathbb{R} \setminus \{x_1, x_2, \cdots\}$.
(3) $\displaystyle\sum_{i=1}^{\infty} p_X(x_i) = 1$.
for some distinct $x_1, x_2, \cdots \in \mathbb{R}$.
Then there exists a discrete r.v. X of some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ s.t. the p.m.f. p_X of X is equal to p. We call such a function a p.m.f.

Proof. abc □

Theorem 4.16 (Step Distribution Function for Discrete R.V.). *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ s.t. $X(\Omega) = \{x_1, x_2 \cdots\}$, where $x_1 < x_2 < \cdots$. Then the distribution function of X is a step function given by*

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

where

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

Proof. abc □

4.4 Expectations of Discrete Random Variables

Definition 4.11 (Expectation)**.** *Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The expectation (or expected value, or mean) of X is given by*

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

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

Remark 4.3 (Explanations of Expectation)**.** *(1) The expectation $\mathbb{E}[X] = \sum\limits_{x \in X(\Omega)} x \cdot p_X(x)$ is the weighted average of $\{x : x \in X(\Omega)\}$ with weights $\{\mathbb{P}(X = x) : x \in X(\Omega)\}$.*
(2) The expectation $\mathbb{E}[X] = \sum\limits_{x \in X(\Omega)} x \cdot p_X(x)$ is the center of gravity of $\{\mathbb{P}(X = x) : x \in X(\Omega)\}$.

Proof. abc □

Theorem 4.17 (Expectation of Constant)**.** *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ s.t. X is a constant with probability 1, i.e., $\mathbb{P}(X = c) = 1$ for some $c \in \mathbb{R}$. Then $c \in X(\Omega)$, $\mathbb{P}(X = x) = 0$, $\forall x \in X(\Omega) \setminus \{c\}$, and $\mathbb{E}[X] = c$. In particular, if X is a constant r.v. of $(\Omega, \mathcal{A}, \mathbb{P})$, i.e., $X(w) = c$, $\forall w \in \Omega$, for some $c \in \mathbb{R}$, then $\mathbb{E}[X] = c$.*

Proof. abc □

Theorem 4.18 (Composition of Function and R.V.)**.** *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and g be a measurable function from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then $g(X) := g \circ X$ is a discrete r.v. of $(\Omega, \mathcal{A}, \mathbb{P})$ and*

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

Proof. abc □

Corollary 4.3 (Linearity of Expectation)**.** *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, $g_1, g_2, \cdots, g_n$ are measurable functions from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, and $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{R}$, Then*

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

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

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

Proof. abc □

4.5 Variances and Moments of Discrete Random Variables

Definition 4.12 (Variance and Standard Deviation). *Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $\mathbb{E}[X]$ exists. The variance of X is given by*

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

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

Remark 4.4 (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).*

Proof. abc □

Theorem 4.19 (Calculating Variance). *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $\mathbb{E}[X]$ exists. Then $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.*

Proof. abc □

Theorem 4.20 (Minimum Distance with Expectation). *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $\mathbb{E}[X]$ exists. If $\mathbb{E}[X^2] < +\infty$, then $\mathrm{Var}(X) = \min_{a\in\mathbb{R}} \mathbb{E}[(X - a)^2]$.*

Proof. abc □

Theorem 4.21 (With Probability 1). *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
(1) $\mathbb{E}[X^2] \geqslant 0$, "=" holds $\Leftrightarrow X = 0$ with probability 1, i.e., $\mathbb{P}(X = 0) = 1$.
(2) If $\mathbb{E}[X]$ exists, then $\mathrm{Var}(X) \geqslant 0$, "=" holds $\Leftrightarrow X = \mathbb{E}[X]$ with probability 1, i.e., $\mathbb{P}(X = \mathbb{E}[X]) = 1$.

Proof. abc □

Theorem 4.22 (Calculating Linear Combination). *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $\mathbb{E}[X]$ exists. Then $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ and $\sigma_{aX+b} = |a|\sigma_X$, $\forall a, b \in \mathbb{R}$.*

Proof. abc □

Definition 4.13 (Moment and Absolute Moment). *Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and $r, c \in \mathbb{R}$.*

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

If the respective sum converges absolutely.

Theorem 4.23 (Existence of Lower Order Moment)**.** *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $0 < r < s$. If $\mathbb{E}[|X|^s]$ exists, then $\mathbb{E}[|X|^r]$ exists. That is, the existence of a higher order moment of X guarantees the existence of a lower order moment of X.*

Proof. abc □

4.6 Standardized Random Variables

Definition 4.14 (Standardized R.V.)**.** *Let X be a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If $\mathrm{Var}(X)$ exists and $\mathrm{Var}(X) \neq 0$, then the standardized r.v. of X is given by*

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

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

Theorem 4.24 (Expectation and Variance of Standardized R.V.)**.** *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathrm{Var}(X)$ exists, $\mathrm{Var}(X) \neq 0$. Then $\mathbb{E}[X^*] = 0$ and $\mathrm{Var}(X^*) = 1$.*

Proof. abc □

Theorem 4.25 (Independence of Units)**.** *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathrm{Var}(X)$ exists, $\mathrm{Var}(X) \neq 0$. Then the standardized r.v. of X is independent of the units in which X is measured.*

Proof. abc □

Remark 4.5 (Standardization for Comparison)**.** *Standardization can be useful when comparing r.v.'s with different distributions.*

Proof. abc □

Chapter 5

Special Discrete Distributions

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

Definition 5.1 (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.

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

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

Definition 5.2 (Bernoulli R.V.)**.** *A discrete r.v.* X *of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *is called a Bernoulli r.v. with parameter* p *where* $0 < p < 1$*, denoted* $X \sim$ *Bernoulli*(p)*, if its p.m.f is given by*

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

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

Theorem 5.1 (Expectation and Variance of Bernoulli R.V.)**.** *Suppose* $X \sim$ *Bernoulli*(p)*, where* $0 < p < 1$*. Then*

$$\mathbb{E}[X] = p, \qquad \mathrm{Var}(X) = p(1-p).$$

Proof. abc □

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 $\Omega = \{(\omega_1, \omega_2, \cdots, \omega_n) : \omega_i = s$ or $f, i = 1, 2, \cdots, n\}$ and $\mathbb{P}(\{(\omega_1, \omega_2, \cdots, \omega_n)\}) = p^i(1-p)^{n-i}$, where $i = |\{1 \leqslant j \leqslant n : \omega_j = s\}|$.

Let X be the number of successes in the n Bernoulli trials.

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

Definition 5.3 (Binomial R.V.). *A discrete r.v. X of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a binomial r.v. with parameter n and p where $n \geqslant 1$ and $0 < p < 1$, denoted $X \sim$ binomial(n, p), if its p.m.f is given by*

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

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

Remark 5.1 (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)

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

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

Theorem 5.2 (Expectation and Variance of Binomial R.V.). *Suppose $X \sim$ binomial(n, p), where $n \geqslant 1$ and $0 < p < 1$. Then*

$$\mathbb{E}[X] = np, \qquad \mathrm{Var}(X) = np(1-p).$$

Proof. abc □

Theorem 5.3 (Maximum Point of Binomial Probability). *Suppose $X \sim$ binomial(n, p), where $n \geqslant 1$ and $0 < p < 1$. Then*

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

Proof. abc □

5.2 Poisson R.V.'s

If $X \sim$ binomial(n, p), then $p_X(i) = \binom{n}{i} p^i (1-p)^{n-i}$ is difficult to calculate if n is large. A recursive relation:

$$p_X(0) = (1-p)^n,\ p_X(i) = \frac{n-i+1}{i(1-p)} \cdot p_X(i-1),\ \forall i \geqslant 1.$$

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

$$\begin{aligned} p_X(i) &= \binom{n}{i} p^i (1-p)^{n-i} = \frac{n(n-1)\cdots(n-i+1)}{i!} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1)\cdots(n-i+1)}{n^i} \cdot \frac{1}{\left(1-\frac{\lambda}{n}\right)^i} \cdot \frac{\lambda^i}{i!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n\to\infty} e^{-\lambda} \frac{\lambda^i}{i!}. \end{aligned}$$

Definition 5.4 (Poisson R.V.). *A discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a Poisson r.v. with parameter λ where $0 < \lambda < 1$, denoted $X \sim$ Poisson(λ), if its p.m.f is given by*

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

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

Remark 5.2 (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)

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

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

Theorem 5.4 (Expectation and Variance of Poisson R.V.). *Suppose $X \sim$ Poisson(λ), where $\lambda > 0$. Then $\mathbb{E}[X] = \lambda$ and $\mathrm{Var}(X) = \lambda$.*

Proof. abc □

Theorem 5.5 (Maximum Point of Poisson Probability). *Suppose $X \sim$ Poisson(λ), where $\lambda > 0$. Then*

$$arg \max_{i \geqslant 0} p_X(i) = \begin{cases} \lambda - 1 \textit{ or } \lambda,\ \textit{if } \lambda \in \mathbb{Z} \\ \lfloor \lambda \rfloor, \quad \textit{if } \lambda \notin \mathbb{Z} \end{cases}$$

Proof. abc □

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 $\Omega = \{s, fs, ffs, \cdots\}$.

Let X be the number of Bernoulli trials until the first success occurs,

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

Definition 5.5 (Geometric R.V.). *A discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a geometric r.v. with parameter p where $0 < p < 1$, denoted $X \sim$ geometric(p), if its p.m.f is given by*

$$p_X(i) = \begin{cases} (1-p)^{i-1} \cdot p, \ i = 0, 1, 2 \cdots \\ 0, \quad \textit{o.w.} \end{cases}$$

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

Remark 5.3 (Justification of P.M.F.).

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

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

Theorem 5.6 (Expectation and Variance of Geometric R.V.). *Suppose $X \sim$ geometric(p), where $0 < p < 1$. Then*

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

Proof. abc □

Theorem 5.7 (Memoryless Property). *Suppose X is a discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $X(\Omega) = \{1, 2 \cdots\}$. Then $P[(X > m+n)|(X > m)] = \mathbb{P}(X > n)$, $\forall m, n > 0 \Leftrightarrow X$ is a geometric r.v.*

Proof. abc □

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 $r \geqslant 1$.

Let X be the number of Bernoulli trials until the r^{th} success occurs,

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

Definition 5.6 (Negative Binomial R.V.). *A discrete r.v. of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *is called a negative binomial r.v. with parameters* r *and* p *where* $r \geqslant 1$ *and* $0 < p < 1$, *denoted* $X \sim$ *neg.-binomial*(r, p), *if its p.m.f is given by*

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

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

Remark 5.4 (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)

$$\begin{aligned} \sum_{i=r}^{\infty} (i-1)(i-2)\cdots(i-r+1)\, x^{i-r} &= \frac{\mathrm{d}^{r-1}}{\mathrm{d}x^{r-1}} \left(\sum_{i=1}^{\infty} x^{i-1} \right) \\ &= \frac{\mathrm{d}^{r-1}}{\mathrm{d}x^{r-1}} \left(\frac{1}{1-x} \right) = \frac{(r-1)!}{(1-x)^r} \end{aligned}$$

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

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

Theorem 5.8 (Expectation and Variance of Negative Geometric R.V.). *Suppose* $X \sim$ *neg.-binomial*(r, p), *where* $r \geqslant 1$ *and* $0 < p < 1$. *Then*

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

Proof. abc □

Theorem 5.9 (Maximum Point of Negative Geometric Probability). *Suppose* $X \sim$ *neg.-binomial*(r, p), *where* $r \geqslant 1$ *and* $0 < p < 1$. *Then*

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

Proof. abc □

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

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

Definition 5.7 (Hypergeometric R.V.). *A discrete r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a hypergeometric r.v. with parameter N_1, N_2 and n where $N_1, N_2 \geqslant 1$ and $n \geqslant 1$, denoted $X \sim$ hypergeometric(N_1, N_2, n), if its p.m.f is given by*

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

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

Remark 5.5 (Justification of P.M.F.). *(1) If $n \leqslant \min\{N_1, N_2\} \to a = \max\{n - N_1, 0\} = 0, b = \min\{n, N_1\} = n$.*
(2)

$$\begin{aligned} &(1+x)^{N_1+N_2} = (1+x)^{N_1}(1+x)^{N_2} \\ \to\ &\textit{the coefficient of } x^n \textit{ is } \binom{N_1+N_2}{n} = \sum_{i=a}^{b} \binom{N_1}{i}\binom{N_2}{n-i}, \\ &\textit{where } a = \max\{n - N_1, 0\}, b = \min\{n, N_1\} \\ \to\ &\sum_{i=a}^{b} p_X(i) = \sum_{i=a}^{b} \frac{\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}} = 1. \\ \to\ &p_X(\cdot) \textit{ is a p.m.f.} \end{aligned}$$

Theorem 5.10 (Expectation and Variance of Hypergeometric R.V.). *Suppose $X \sim$ hypergeometric(N_1, N_2, n), where $N_1, N_2 \geqslant 1$ and $1 \leqslant n \leqslant \min\{N_1, N_2\}$. Then*

$$\mathbb{E}[X] = \frac{nN_1}{N_1+N_2}, \ \mathrm{Var}(x) = n \cdot \frac{N_1}{N_1+N_2} \cdot \frac{N_2}{N_1+N_2} \cdot \left(1 - \frac{n-1}{N_1+N_2-1}\right).$$

Proof. abc □

Theorem 5.11 (Binomial Approximation for Hypergeometric). *n balls are drawn with replacement*

$$\begin{aligned} &\to\ X \sim \text{ binomial}\left(n, \frac{N_1}{N_1+N_2}\right) \\ &\to\ \mathbb{E}[X] = n \cdot \frac{N_1}{N_1+N_2}, \ \mathrm{Var}(x) = n \cdot \frac{N_1}{N_1+N_2} \cdot \frac{N_2}{N_1+N_2}. \end{aligned}$$

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

Proof. abc □

Theorem 5.12 (Maximum Point of Hypergeometric Probability)**.** *Suppose* $X \sim \text{hypergeometric}(N_1, N_2, n)$*, where* $N_1, N_2 \geqslant 1$ *and* $1 \leqslant n \leqslant \min\{N_1, N_2\}$*. Then*

$$\begin{aligned}
&arg \max_{0\leqslant i\leqslant n} p_X(i) \\
&= \begin{cases} \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} - 1 \ or \ \dfrac{(n+1)(N_1+1)}{N_1+N_2+2}, \ if \ \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} \in \mathbb{Z} \\ \left\lfloor \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} \right\rfloor, \ if \ \dfrac{(n+1)(N_1+1)}{N_1+N_2+2} \notin \mathbb{Z} \end{cases}
\end{aligned}$$

Proof. abc □

Remark 5.6 (Binomial and Poisson Approximation for Hypergeometric)**.**

$$\begin{aligned}
p_X(i) &= \frac{\binom{N_1}{i}\binom{N_2}{n-i}}{\binom{N_1+N_2}{n}} \\
&= \frac{n!}{i!\,(n-i)!} \cdot \frac{N_1\,(N_1-1)\cdots(N_1-i+1)\,N_2\,(N_2-1)\cdots(N_2-n+i+1)}{(N_1+N_2)\,(N_1+N_2-1)\cdots(N_1+N_2+n-1)}
\end{aligned}$$

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

$$\begin{aligned}
p_X(i) &= \binom{n}{i} \cdot \frac{1}{1 \cdot \left(1 - \frac{1}{N_1+N_2}\right)\cdots\left(1 - \frac{n-1}{N_1+N_2}\right)} \\
&\cdot \frac{N_1}{N_1+N_2}\left(\frac{N_1}{N_1+N_2} - \frac{1}{N_1+N_2}\right)\cdots\left(\frac{N_1}{N_1+N_2} - \frac{i-1}{N_1+N_2}\right)\left(\frac{N_2}{N_1+N_2}\right) \\
&\cdot \left(\frac{N_2}{N_1+N_2} - \frac{1}{N_1+N_2}\right)\cdots\left(\frac{N_2}{N_1+N_2} - \frac{n-i-1}{N_1+N_2}\right) \\
&\xrightarrow{N_1,\ N_2 \to \infty} \binom{n}{i} p^i (1-p)^{n-i} \leftarrow \text{binomial}(n, p)
\end{aligned}$$

(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*

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

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

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

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

Chapter 6

Continuous Random Variables

6.1 Probability Density Function

Definition 6.1 (Probability Density Function)**.** *Let X be a r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. 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.*

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

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

Remark 6.1 (Approximation of Probability)**.**

$$\mathbb{P}(a \leqslant X \leqslant a+\delta) = \int_a^{a+\delta} f_X(x)\mathrm{d}x = f_X(a_\delta) \cdot \delta,$$

for some $a_\delta \in [a, \ a+\delta]$.
If f_X is continuous at a

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

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

Theorem 6.1 (C.D.F and Probability from P.D.F.)**.** *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
(1)

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

Therefore, $F_X(x)$ is a continuous function.
(2)

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

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

$$\begin{aligned}&\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(a \leqslant X < b)\\ =&\mathbb{P}(a < X \leqslant b) = \mathbb{P}(a < X < b)\\ =&\int_a^b f_X(x)\mathrm{d}x.\end{aligned}$$

Proof. abc □

Theorem 6.2 (Existence of P.D.F.)**.** *Suppose $f : \mathbb{R} \to [0, \infty)$ is a nonnegative real-valued function s.t.*

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

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

Proof. abc □

Definition 6.2 (Sufficient Conditions of P.D.F.)**.** *A nonnegative real-valued function $f : \mathbb{R} \to [0, \infty)$ s.t.*

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

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

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

Remark 6.2 (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 6.3 (Method of Distribution Functions)**.** *Suppose X is a continuous r.v. of a probability space$(\Omega, \mathcal{A}, \mathbb{P})$.*
If $Y = g(X)$, then

$$\begin{aligned}f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}\left[F_Y(y)\right] = \frac{\mathrm{d}}{\mathrm{d}y}\left[\mathbb{P}(Y \leqslant y)\right] = \frac{\mathrm{d}}{\mathrm{d}y}\left[P[g(x) \leqslant y]\right]\\ &\to \frac{\mathrm{d}}{\mathrm{d}y}\left[X \sim g^{-1}(y)\right] \to \frac{\mathrm{d}}{\mathrm{d}y}\left[F_X\left(g^{-1}(y)\right)\right] \to \frac{\mathrm{d}}{\mathrm{d}y}\left[g^{-1}(y)\right] \cdot f_X\left[g^{-1}(y)\right].\end{aligned}$$

Proof. abc □

Theorem 6.4 (Method of Transformations)**.** *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ 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

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

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

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

Proof. abc □

6.3 Expectations and Variances

Definition 6.3 (Expectation)**.** *Let X be a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ s.t. its p.d.f. is continuous. The expectation (or mean) of X is given by*

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

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

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

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

Remark 6.3 (Necessary and Sufficient Condition of Absolutely Integrable)**.**

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \int_{0}^{\infty} x f_X(x)\mathrm{d}x - \int_{-\infty}^{0} (-x) f_X(x)\mathrm{d}x \\
\to \mathbb{E}[|X|] &= \int_{0}^{\infty} x f_X(x)\mathrm{d}x + \int_{-\infty}^{0} (-x) f_X(x)\mathrm{d}x \\
\therefore \mathbb{E}[|X|] < \infty &\Leftrightarrow \int_{0}^{\infty} x f_X(x)\mathrm{d}x < \infty \text{ and } \int_{-\infty}^{0} (-x) f_X(x)\mathrm{d}x < \infty.
\end{aligned}$$

Theorem 6.5 (Calculation of Expectation)**.** *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then*

$$\begin{aligned}
\mathbb{E}[X] &= \int_{0}^{\infty} \mathbb{P}(x > t)\mathrm{d}t - \int_{0}^{\infty} \mathbb{P}(x \leqslant -t)\mathrm{d}t \\
&= \int_{0}^{\infty} [1 - F_X(t)]\mathrm{d}t - \int_{0}^{\infty} [F_X(-t)]\mathrm{d}t.
\end{aligned}$$

Proof. abc □

Corollary 6.1 (Calculation of r^{th} Moment)**.** *Suppose X is a nonnegative continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and $r > 0$. Then*

$$\mathbb{E}[X^r] = \int_0^\infty rt^{r-1}\mathbb{P}(x > t)\mathrm{d}t = \int_0^\infty rt^{r-1}\left[1 - F_X(t)\right]\mathrm{d}t.$$

In particular,

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(x > t)\mathrm{d}t = \int_0^\infty \left[1 - F_X(t)\right]\mathrm{d}t.$$

Proof. abc □

Theorem 6.6 (Approximation of Expectation)**.** *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then*

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

Therefore,

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

Proof. abc □

Theorem 6.7 (Infinite Zero)**.** *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then,*

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

Proof. abc □

Theorem 6.8 (Expectation of Measurable Function)**.** *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and suppose g is a measurable function from $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ to $(\mathbb{R}, \mathcal{B}_\mathbb{R})$. Then*

$$\mathbb{E}[g(X)] = \int_{-\infty}^\infty g(X) \cdot f_X(x)\mathrm{d}x$$

Proof. abc □

Corollary 6.2 (Expectation of Linear Combination of Measurable Functions)**.** *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. $g_1, g_2, \cdots g_n$ are measurable functions from $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ to $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, and $\alpha_1, \alpha_2, \cdots \alpha_n \in \mathbb{R}$. Then*

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

Proof. abc $\square$

Definition 6.4 (Variance and Standard Deviation). *Let X be a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $\mathbb{E}[X]$ exists. The **variance** of X is given by $\mathrm{Var}(x) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. And the **standard deviation** of X is given by $\sigma_X = \sqrt{\mathrm{Var}(x)} = \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}$.*

Theorem 6.9 (Minimum Distance with Expectation). *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and suppose $\mathbb{E}[X]$ exists. If $\mathbb{E}\left[X^2\right] < +\infty$, then $\mathrm{Var}(x) = \min_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2]$.*

Proof. abc $\square$

Theorem 6.10 (Calculation of Linear Combination). *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and suppose $\mathbb{E}[X]$ exists. Then*

(1)

$$\mathrm{Var}(x) = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$$

(2)

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(x), \quad \sigma_{aX+b} = |a|\, \sigma_X, \ \forall a, b \in \mathbb{R}.$$

Proof. abc $\square$

Definition 6.5 (Moment and Absolute Moment). *Let X be a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and $r, c \in \mathbb{R}$.*

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

If the respective sum converges absolutely.

Theorem 6.11 (Existence of Lower Order Moment). *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $0 < r < s$. If $\mathbb{E}[|X|^s]$ exists, then $\mathbb{E}[|X|^r]$ exists. That is, the existence of a higher order moment of X guarantees the existence of a lower order moment of X.*

Proof. abc $\square$

Theorem 6.12 (Positive Variance). *Suppose X is a continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then*

$$\mathbb{E}\left[(X - a)^2\right] > 0, \ \forall a \in \mathbb{R}.$$

Therefore

$$\mathbb{E}[X] \textit{ exists } \to \mathrm{Var}(X) > 0.$$

Proof. abc $\square$

Chapter 7

Special Continuous Distributions

7.1 Uniform R.V.'s

Definition 7.1 (Uniform R.V.). *A continuous r.v. of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *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 (P.D.F. and C.D.F.). *(1)* $f_X(x) \geqslant 0, \ \forall x \in \mathbb{R}$, *and*

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

$\to f_X(x)$ *is a p.d.f.*
(2)

$$F_X(x) = \begin{cases} 0, & \text{if } x \leqslant \alpha \\ \frac{x-\alpha}{\beta-\alpha}, & \text{if } \alpha < x < \beta \\ 1, & \text{if } x \geqslant \beta \end{cases}$$

Theorem 7.1 (Expectation and Variance of Uniform R.V.). *Suppose* $X \sim U(\alpha, \beta)$, *where* $\alpha, \beta \in \mathbb{R}$ *and* $\alpha < \beta$. *Then*

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

Therefore

$$\mathbb{E}[X] = \frac{\alpha + \beta}{2}, \qquad \mathrm{Var}(x) = \frac{(\beta - \alpha)^2}{12}.$$

Proof. abc $\square$

Remark 7.2 (Expectation and Variance of Discrete "Uniform R.V."). *Suppose* $X \sim \text{Uniform}(1, 2, \cdots, n)$*, where* $n \geqslant 1$*. Then*

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

and

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

Theorem 7.2 (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*

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

Proof. abc □

7.2 Normal (Gaussian) R.V.'s

Definition 7.2 (Normal (Gaussian) R.V.). *A continuous r.v. of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *is called a normal (Gaussian) r.v. with parameters* μ *and* σ^2*, where* $\mu, \sigma \in \mathbb{R}$*,* $\sigma \neq 0$*, denoted* $X \sim \mathcal{N}(\mu, \sigma^2)$*, if its p.d.f. is given by*

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \; -\infty < x < \infty.$$

Remark 7.3 (P.D.F. and C.D.F.). *(1)* $f_X(x) \geqslant 0, \; \forall x \in \mathbb{R}$*, and let* $I = \int_{-\infty}^{\infty} e^{-ax^2} \mathrm{d}x$*.*

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-a\left(x^2+y^2\right)} \mathrm{d}x \mathrm{d}y \\ &\xrightarrow{x=r\cos\theta, y=r\sin\theta} \int_0^{\infty} \int_0^{2\pi} e^{-ar^2} r \mathrm{d}r \mathrm{d}\theta = \frac{\pi}{a} \\ &\to I = \sqrt{\frac{\pi}{a}} \to \int_{-\infty}^{\infty} \sqrt{\frac{a}{\pi}} \cdot e^{-ax^2} \mathrm{d}x = 1 \end{aligned}$$

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

$\to f_X(x)$ *is a p.d.f.*
(2) If $\mu = 0, \; \sigma^2 = 1$*, then* X *is called a standard normal (Gaussian) r.v.*

(3)

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

where

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

Theorem 7.3 (Symmetric about μ)**.** *Suppose* $X \sim \mathcal{N}(\mu, \sigma^2)$.
(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

$$F_X(\mu) = \frac{1}{2}.$$

Proof. abc □

Theorem 7.4 (Linear Generated R.V.)**.** *Suppose* $X \sim \mathcal{N}(\mu, \sigma^2)$*, where* $\mu, \sigma \in \mathbb{R},\ \sigma \neq 0$. *Suppose* $Y = aX + b$*, where* $\alpha, \beta \in \mathbb{R}$ *and* $a \neq 0$. *Then,*

$$Y \sim \mathcal{N}\left(a\mu + b, a^2\sigma^2\right).$$

In particular, if

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

then

$$Y \sim \mathcal{N}(0, 1).$$

Proof. abc □

Definition 7.3 (Gamma Function)**.** *The function* $\Gamma : (0, \infty) \to \mathbb{R}$ *given by*

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

is called the gamma function.

Theorem 7.5 (Properties of Gamma Function). *(1)*

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

(2)

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

(3)

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

Proof. abc □

Theorem 7.6 (Calculation of Moment and Absolute Moment). *Suppose* $X\sim\mathcal{N}(\mu,\sigma^2)$, *where* $\mu,\sigma\in\mathbb{R},\ \ \sigma\neq 0.$
(1)

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

(2)

$$\mathbb{E}\left[(x-\mu)^n\right]=\begin{cases}0, & \textit{if}\, n=2k+1,\quad k\geqslant 0\\ \frac{(2k)!}{2^k\cdot k!}\sigma^{2k} & \textit{if}\ \, n=2k,\qquad k\geqslant 0\end{cases}$$

(3)

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

Proof. abc □

Theorem 7.7 (De Moivre-Laplace Theorem). *Suppose* $X\sim\ \mathrm{binomial}(n,p)$, *where* $n\geqslant 1$ *and* $0<p<1$. *Then*

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

Proof. abc □

Theorem 7.8 (Approximation of $\Phi(x)$).

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

Proof. abc □

Theorem 7.9 (Expectation of Exponential Function). *Suppose* $X\sim\mathcal{N}(\mu,\sigma^2)$, *where* $\mu,\sigma\in\mathbb{R},\ \sigma\neq 0$, *and* $\alpha\in\mathbb{R}$. *Then*

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

Proof. abc □

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

Definition 7.4 (Gamma R.V., Erlang R.V. and Exponential R.V.). *A continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a gamma r.v. with parameters α and λ, where $\alpha, \lambda > 0$, denoted $X \sim \mathcal{G}(\alpha, \lambda)$, if its p.d.f. is given by*

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

If $\alpha = n$, $n \geqslant 1$, 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 7.4 (Properties of P.D.F.). *(1)*

$$\int_{-\infty}^{\infty} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \mathrm{d}x \xrightarrow{t=\lambda x} \int_0^{\infty} \frac{e^{-t}t^{\alpha-1}}{\Gamma(\alpha)} \mathrm{d}t = \frac{\Gamma(\alpha)}{\Gamma(\alpha)} = 1$$

$\rightarrow f_X(x)$ is a p.d.f.
(2)

$$\begin{aligned} f_X'(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x}\left(-\lambda x^{\alpha-1} + (\alpha-1)\, x^{\alpha-2}\right) \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x} \cdot x^{\alpha-2}\left[-\lambda x + (\alpha-1)\right] \end{aligned}$$

$$\begin{aligned} & f_X''(x) \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x}\left[-\lambda^2 x^{\alpha-1} - \lambda(\alpha-1)\, x^{\alpha-2} - \lambda(\alpha-1)\, x^{\alpha-2} + (\alpha-2)(\alpha-1)\, x^{\alpha-3}\right] \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot e^{-\lambda x} \cdot x^{\alpha-3}\left[(\lambda x - (\alpha-1))^2 - (\alpha-1)\right] \end{aligned}$$

$$\therefore 0 < \alpha \leqslant 1 \;\rightarrow\; f_X'(x) < 0,\ f_X''(x) > 0,\ \forall x > 0.$$

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

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 7.10 (Calculation of C.D.F.). *Suppose $X \sim \mathcal{G}(\alpha, \lambda)$, where $\alpha, \lambda > 0$. Then*

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

where

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

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

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

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

Proof. abc □

Theorem 7.11 (Expectation and Variance of Gamma R.V.). *Suppose $X \sim \mathcal{G}(\alpha, \lambda)$, where $\alpha, \lambda > 0$. Then*

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

where

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

Therefore,

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

Proof. abc □

Theorem 7.12 (Linear Generated Gamma R.V.). *Suppose $X \sim \mathcal{G}(\alpha, \lambda)$, where $\alpha, \lambda > 0$, and $Y = aX$, where $a > 0$. Then*

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

Proof. abc □

Theorem 7.13 (Gamma R.V. from Normal R.V.). *Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$, where $\mu, \sigma \in \mathbb{R}$, $\sigma \neq 0$ and $Y = (X - \mu)^2$. Then*

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

Proof. abc □

Lemma 7.1 (Plus to Multiply Property of Exponential Function). *Suppose $f : [0, +\infty) \to \mathbb{R}$ is right continuous on $[0, +\infty)$ and $f(x + y) = f(x) \cdot f(y)$, $\forall x, y \geqslant 0$. Then there either $f(x) = 0$, $\forall x \geqslant 0$ or $\exists \lambda \in \mathbb{R}$ s.t. $f(x) = e^{-\lambda x}$, $\forall x \geqslant 0$.*

Proof. abc ☐

Theorem 7.14 (Memoryless Property). *Suppose X is a nonnegative continuous r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then $\mathbb{P}(x > s+t | x > s) = \mathbb{P}(x > t),\ \forall s, t > 0 \ \Leftrightarrow X \sim \mathcal{E}(\lambda)$, for some $\lambda > 0$.*

Proof. abc ☐

Remark 7.5 (Analog of Geometric R.V.). *Exponential r.v.'s are the continuous analog of geometric r.v.'s.*

Theorem 7.15 (Geometric R.V. from Exponential R.V.). *Suppose $X \sim \mathcal{E}(\lambda)$ where $\lambda > 0$ and $Y = \lceil X \rceil$. Then $Y \sim \mathrm{geometric}\left(1 - e^{-\lambda}\right)$.*

Proof. abc ☐

Definition 7.5 (Independent Set). *A set of r.v.'s $\{X_i : i \in I\}$ of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called independent, if for any finite subset $\{X_{i_1}, X_{i_2}, \cdots, X_{i_k}\},\ k \geqslant 2$ of $\{X_i : i \in I\}$ the events*

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

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

Definition 7.6 (Continuous R.Vect.). *A r.vect. $\boldsymbol{X} = (X_1,\ X_2,\ \cdots, X_n)$ of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called an absolute continuous r.vect. (or continuous r.vect.) if there exists a nonnegative real-valued function $f_{\boldsymbol{X}} : \mathbb{R}^n \to [0, \infty)$ s.t.*

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2, \cdots, X_n \in B_k) = \int_{B_1} \int_{B_2} \cdots \int_{B_n} f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}x_n \cdots \mathrm{d}x_2 \mathrm{d}x_1$$

for all $B_1, B_2, \cdots, B_n \in \mathcal{B}_{\mathbb{R}}$.
Then the function $f_{\boldsymbol{X}}$ is called the p.d.f. of the r.vect. $\boldsymbol{X}$, or the joint p.d.f. of the r.v.'s $X_1,\ X_2,\ \cdots, X_n$.

Theorem 7.16 (P.D.F. and C.D.F. of Continuous R.Vect.). *Suppose $\boldsymbol{X} = (X_1,\ X_2,\ \cdots, X_n)$ is a continuous r.vect. and*

$$F_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbb{P}(X_1 \leqslant x_1, X_2 \leqslant x_2, \cdots, X_n \leqslant x_n).$$

Then

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

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

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

Proof. abc ☐

Theorem 7.17 (Convolution Theorem). *If $\boldsymbol{X} = (X_1, X_2)$ is a continuous r.vect. and $Y = X_1 + X_2$. Then*

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

Furthermore, if $X_1 \perp X_2$, then

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

Proof. abc □

Definition 7.7 (Beta Function). *The function $B:\ \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ is given by*

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

is called beta function.

Lemma 7.2 (Calculation of Beta Function).

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

Proof. abc □

Theorem 7.18 (Independent Additivity of Gamma R.V.). *Suppose $X_i \sim \mathcal{G}(\alpha_i, \lambda)$ where $\alpha_i, \lambda > 0$, $i = 1, 2, \cdots, n$, $X_1,\ X_2,\ \cdots, X_n$ are independent, and $Y = X_1 + X_2 + \cdots + X_n$. Then*

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

Proof. abc □

Theorem 7.19 (Independent Minimum of Exponential R.V.). *Suppose $X_i \sim \mathcal{E}(\lambda_i)$ where $\lambda_i > 0,\ i = 1, 2, \cdots, n$, and $X_1, X_2, \cdots, X_n$ are independent.*
(1) If $Y = \min\{X_1, X_2, \cdots, X_n\}$, then

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

(2)

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

Proof. abc □

Definition 7.8 (Stochastic Process). *A stochastic process (s.p.)* $\{X(t) : i \in I\}$ *is a collection of r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$. *If* $I = \{0, 1, 2, \ldots\}$ *or* $\{0, \pm 1, \pm 2, \cdots\}$, *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 7.9 (Counting Process and Poisson Process). *Let* $\{T_1, T_2, \cdots\}$ *be a discrete-time S.P. s.t.* $T_i,\ i = 1, 2, \ldots,$ *is the time of occurrence of the* i^{th} *event, and* $0 < T_1 < T_2 < \cdots$.

Let $X_i = T_i - T_{i-1},\ i = 1, 2, \ldots,$ *where* $T_0 = 0$ *be the inter-occurrence time between the* $(i-1)^{th}$ *and the* i^{th} *events, and* $N(t) = |\{i : 0 < T_i \leqslant t\}|$ *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.* $\{T_1, T_2, \cdots\}$.

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

Theorem 7.20 (Necessary and Sufficient Condition of Poisson Process). *Suppose* $\{T_1, T_2, \cdots\}$ *is a S.P. s.t.* $0 < T_1 < T_2 < \cdots$ *and its inter-occurrence times* $X_i = T_i - T_{i-1},\ i = 1, 2, \ldots$ *are i.i.d., where* $T_0 = 0$. *Then* $\{T_1, T_2, \cdots\}$ *is a Poisson process with rate* $\lambda \iff X_i \sim \mathcal{E}(\lambda),\ i = 1, 2, \ldots$.

Proof. abc □

Remark 7.6 (Negative Binomial $\leftrightarrow$ Geometric vs Gamma$\leftrightarrow$ 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* $X_i \sim \mathrm{geometric}(p)$ *is the number of Bernoulli trials between the* $(i-1)^{th}$ *and the* i^{th} *successes, and* $X_1, X_2, \cdots$ *are independent.*

(2) A gamma r.v. $T_n = X_1 + X_2 + \cdots + X_n \sim \mathcal{G}(n, \lambda)$ *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 inter-occurrence time between the* $(i-1)^{th}$ *and the* i^{th} *events, and* $X_1, X_2, \cdots$ *are independent.*

Theorem 7.21 (Merging and Splitting of Poisson Process). *(1) Suppose that* k *independent Poisson processes with rates* $\lambda_1, \lambda_2, \cdots, \lambda_k$ *are merged into a S.P.* $\{T_1, T_2, \cdots\}$. *Then* $\{T_1, T_2, \cdots\}$ *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* $P_i,\ i = 1, 2, \cdots, k$. *Then the S.P.* $\{T_1, T_2, \cdots\}$ *of the times of the occurrences of the type-*i *events is a Poisson process with rate* $\lambda \cdot P_i,\ i = 1, 2, \cdots, k$.

Proof. abc □

7.4 Beta R.V.'s

Definition 7.10 (Beta R.V.). *A continuous r.v.* X *of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *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*

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

where

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

Remark 7.7 (P.D.F. and C.D.F.). *(1)* $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1 \ \rightarrow \ f_X(x)$ *is a p.d.f.*

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

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

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

(4)

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

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

where

$$h(x,\alpha,\beta) = (\alpha+\beta-2)\,(\alpha+\beta-3)\,x^2 - 2\,(\alpha-1)\,(\alpha+\beta-3)\,x + (\alpha-1)\,(\alpha-2)\,,$$

and

$$f(x,\alpha,\beta) = \left(x - \frac{\alpha-1}{\alpha+\beta-2}\right)^2 - \frac{(\alpha-1)\,(\beta-1)}{(\alpha+\beta-2)^2\,(\alpha+\beta-3)}.$$

Theorem 7.22 (Expectation and Variance of Beta R.V.)**.** *Suppose* $X \sim \mathcal{B}(\alpha, \beta)$*, then*

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

Therefore,

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

and

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

Proof. abc □

Theorem 7.23 (Beta R.V. vs Binomial R.V.)**.** *Suppose* $X \sim \mathcal{B}(\alpha, \beta)$*, and* $Y \sim \mathrm{binomial}(\alpha+\beta-1, p)$*, where* $\alpha, \beta \in \mathbb{Z}^+,\ 0 < p < 1$*. Then*

$$\mathbb{P}(X \leqslant p) = \mathbb{P}(Y \geqslant \alpha)$$

and

$$\mathbb{P}(X \geqslant p) = \mathbb{P}(Y \leqslant \alpha - 1).$$

Proof. abc □

Chapter 8

Bivariate and Multivariate Distributions

8.1 Joint Distributions of Two or More R.V.'s

Definition 8.1 (Joint P.M.F. of Multiple R.v.'s)**.** *Let* $X_1,\ X_2,\ \cdots, X_n$ *be discrete r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*. The nonnegative function* $P_X : \mathbb{R}^n \to [0,1]$ *given by*

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = P_{\boldsymbol{X}}(\{\boldsymbol{x}\}) = \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \begin{cases} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}), \ \boldsymbol{x} \in \boldsymbol{X}(\Omega) \\ \quad 0, \quad \boldsymbol{x} \in \mathbb{R}^n \backslash \boldsymbol{X}(\Omega) \end{cases}$$

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

Remark 8.1 (Properties of Joint P.M.F.)**.** *(1)*

$$p_{\boldsymbol{X}}(\boldsymbol{x}) \geqslant 0,\ \forall \boldsymbol{x} \in \boldsymbol{X}(\Omega) \text{ and } p_{\boldsymbol{X}}(\boldsymbol{x}) = 0,\ \forall \boldsymbol{x} \in \mathbb{R}^n \backslash \boldsymbol{X}(\Omega).$$

(2)

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

(3)

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

(4)

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \begin{cases} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}), \ \boldsymbol{x} \in \displaystyle\prod_{i=1}^{n} X_i(\Omega) \\ \quad 0, \qquad \boldsymbol{x} \in \mathbb{R}^n \backslash \displaystyle\prod_{i=1}^{n} X_i(\Omega) \end{cases}$$

Theorem 8.1 (Joint Marginal P.M.F.). *Suppose $X_1, X_2, \cdots, X_n$ are discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then*

$$p_{X_{i_1}, X_{i_2}, \cdots, X_{i_k}}\left(x_{i_1}, x_{i_2}, \cdots, x_{i_k}\right) = \begin{cases} \displaystyle\sum_{\substack{x_i \in X_i(\Omega) \\ i \neq i_1, i_2, \cdots i_k}} p_{X_i}\left(x_i\right), \ \forall i = i_1, i_2, \cdots, i_k \\ 0, \qquad o.w. \end{cases}$$

We call

$$p_{X_{i_1}, X_{i_2}, \cdots, X_{i_k}}\left(x_{i_1}, x_{i_2}, \cdots, x_{i_k}\right)$$

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

Proof. abc □

Theorem 8.2 (Expectation of Measurable Function). *Suppose $X_1,\ X_2, \cdots, X_n$ are discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and g is a measurable function from $(\mathbb{R}^n,\ \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R},\ \mathcal{B}_{\mathbb{R}})$. Then*

$$\mathbb{E}\left[g(\boldsymbol{x})\right] = \sum_{\substack{x_i \in X_i(\Omega) \\ i = 1, 2, \cdots, n}} g(\boldsymbol{x}) \cdot p_{\boldsymbol{X}}(\boldsymbol{x}).$$

Proof. abc □

Corollary 8.1 (Expectation of Linear Combined Measurable Function). *Suppose $X_1,\ X_2, \cdots, X_n$ are discrete r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and $g_1, g_2, \cdots, g_m$ are measurable functions from $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, and $\alpha_1, \alpha_2, \cdots, \alpha_m \in \mathbb{R}$, Then*

$$\sum_{k=1}^{m} \alpha_k \cdot g_k(\boldsymbol{x})$$

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

$$\mathbb{E}\left[\sum_{k=1}^{m} \alpha_k g_k(\boldsymbol{x})\right] = \sum_{k=1}^{m} \alpha_k \mathbb{E}\left[g_k(\boldsymbol{x})\right].$$

Proof. abc □

Definition 8.2 (Joint P.D.F.). *Let $X_1,\ X_2, \cdots, X_n$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We say that $X_1,\ X_2, \cdots, X_n$ are jointly continuous r.v.'s if there exists a nonnegative function $f_{\boldsymbol{X}} : \mathbb{R}^n \to [0, 1]$ s.t.*

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

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

Theorem 8.3 (Joint Marginal P.D.F.)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then $X_{i_1}, X_{i_2},\ \cdots, X_{i_k}$ are also jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with joint p.d.f.*

$$f_{X_{i_1},X_{i_2},\cdots,X_{i_k}}(x_{i_1}, x_{i_2},\ \cdots, x_{i_k}) = \underbrace{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}}_{n-k} f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}x_i$$

where $i \neq i_1, i_2, \cdots i_k$.
We call

$$f_{X_{i_1},X_{i_2},\cdots,X_{i_k}}(x_{i_1}, x_{i_2},\ \cdots, x_{i_k})$$

the joint p.d.f. marginalized over $X_{i_1}, X_{i_2},\ \cdots, X_{i_k}$. If $k = 1$, we call $f_{X_i}(x_i)$ the marginal p.d.f. of X_i.

Proof. abc □

Theorem 8.4 (Expectation of Measurable Function)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and g is a measurable function from $(\mathbb{R}^n,\ \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then*

$$\mathbb{E}\left[g(\boldsymbol{x})\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}x_n\cdots\mathrm{d}x_2\mathrm{d}x_1.$$

Proof. abc □

Remark 8.2 (Properties of Joint P.D.F.)**.** *(1)*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) > 0,\ \forall \boldsymbol{x} \in \mathbb{R}^n.$$

(2)

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

(3)

$$\mathbb{P}(X_1 \in B_1, X_2 \in B_2,\ \cdots, X_n \in B_n) = \int_{B_1}\int_{B_2}\cdots\int_{B_n} f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}x_n\cdots\mathrm{d}x_2\mathrm{d}x_1,$$

$\forall B_i \in \mathcal{B}_{\mathbb{R}^n},\ i = 1, 2,\ \cdots, n.$
(4)

$$\mathbb{P}(\boldsymbol{X} = \mathbf{a}) = \int_{a_1}^{a_1}\int_{a_2}^{a_2}\cdots\int_{a_n}^{a_n} f_{\boldsymbol{X}}(\boldsymbol{x})\mathrm{d}x_n\cdots\mathrm{d}x_2\mathrm{d}x_1 = 0.$$

(5)

$$\begin{aligned}
&\mathbb{P}(a_i \leqslant X_i \leqslant a_i + \delta_i,\ i = 1, 2,\ \cdots, n) \\
&= \int_{a_1}^{a_1+\delta_1} \int_{a_2}^{a_2+\delta_2} \cdots \int_{a_n}^{a_n+\delta_n} f_{\boldsymbol{X}}(\boldsymbol{x}) \mathrm{d}x_n \cdots \mathrm{d}x_2 \mathrm{d}x_1 \\
&= f_{\boldsymbol{X}}(\mathbf{a}_{\boldsymbol{\delta}}) \cdot \delta_1 \cdot \delta_2 \cdots \delta_n \textit{ for some } \mathbf{a}_{\boldsymbol{\delta}} \in \prod_{i=1}^{n} [a_i, a_i + \delta_i] \textit{ if } f_{\boldsymbol{X}}(\boldsymbol{x}) \textit{ is continuous.} \\
&\to \lim_{\boldsymbol{\delta} \to \mathbf{0}} \frac{\mathbb{P}(a_i \leqslant X_i \leqslant a_i + \delta_i,\ i = 1, 2,\ \cdots, n)}{\delta_1 \cdot \delta_2 \cdots \delta_n} = \lim_{\boldsymbol{\delta} \to \mathbf{0}} f_{\boldsymbol{X}}(\mathbf{a}_{\boldsymbol{\delta}}) = f_{\boldsymbol{X}}(\mathbf{a}) \\
&\textit{and } \mathbb{P}(a_i \leqslant X_i \leqslant a_i + \delta_i,\ i = 1, 2,\ \cdots, n) \approx f_{\boldsymbol{X}}(\mathbf{a}) \cdot \delta_1 \cdot \delta_2 \cdots \delta_n.
\end{aligned}$$

Corollary 8.2 (Expectation of Linear Combined Measurable Function)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are* ***jointly continuous*** *r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $g_1,\ g_2,\ \cdots, g_m$ are* ***measurable functions*** *from $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, and $\alpha_1, \alpha_2,\ \cdots, \alpha_m \in \mathbb{R}$, then*

$$\sum_{k=1}^{m} \alpha_k \cdot g_k(\boldsymbol{x})$$

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

$$\mathbb{E}\left[\sum_{k=1}^{m} \alpha_k \cdot g_k(\boldsymbol{x})\right] = \sum_{k=1}^{m} \alpha_k \cdot \mathbb{E}\left[g_k(\boldsymbol{x})\right].$$

Proof. abc □

Definition 8.3 (Joint C.D.F.)**.** *Let $X_1,\ X_2,\ \cdots, X_n$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The* ***joint c.d.f.*** *of $X_1,\ X_2,\ \cdots, X_n$ is given by*

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

Theorem 8.5 (Joint Marginal C.D.F.)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then*

$$\begin{aligned}
&F_{X_{i_1}, X_{i_2}, \cdots X_{i_k}}(x_{i_1}, x_{i_2},\ \cdots, x_{i_k}) \\
=&F_{\boldsymbol{X}}(\infty,\ \cdots, \infty, x_{i_1}, \infty,\ \cdots, \infty, x_{i_2}, \infty,\ \cdots, \infty, x_{i_k}, \infty,\ \cdots, \infty)
\end{aligned}$$

We call

$$F_{X_{i_1}, X_{i_2},\ \cdots, X_{i_k}}(x_{i_1}, x_{i_2},\ \cdots, x_{i_k})$$

the ***joint c.d.f.*** *marginalized over $X_1,\ X_2,\ \cdots, X_n$. If $k = 1$, we call $F_{X_i}(x_i)$ the* ***marginal c.d.f.*** *of X_i.*

Proof. abc □

Theorem 8.6 (Properties of Joint C.D.F.)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
*(1) $F_{\boldsymbol{X}}(\boldsymbol{x})$ is **increasing** and **right continuous** in each argument $x_i, i = 1, 2, \cdots, n$.*
(2) $F_{\boldsymbol{X}}(\boldsymbol{x}) = 0$ if there exists at least one i such that $x_i = -\infty$.
(3) $F_{\boldsymbol{X}}(\infty, \infty, \cdots, \infty) = 1$.
*(4) If $X_1,\ X_2,\ \cdots, X_n$ are **jointly continuous** r.v.'s, then*

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

Proof. abc □

8.2 Independent R.V.'s

Definition 8.4 (Independent Set)**.** *Let $\{X_i,\ i \in I\}$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We say that the r.v.'s $\{X_i,\ i \in I\}$ are **independent** if for any finite subset $\{X_{i_1}, X_{i_2}, \cdots, X_{i_k}\}$ $(k \geqslant 2)$ of $\{X_i,\ i \in I\}$, the events $X_{i_1} \in B_{i_1}, X_{i_2} \in B_{i_2}, \cdots, X_{i_k} \in B_{i_k}$ are independent $\forall B_{i_1}, B_{i_2}, \cdots, B_{i_k} \in \mathcal{B}_{\mathbb{R}}$. Otherwise, the r.v.'s $\{X_i,\ i \in I\}$ are dependent.*

Theorem 8.7 (Equivalent Statements of Independence)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The following three statements are **equivalent**:*
(1) $X_1,\ X_2,\ \cdots, X_n$ are independent.
(2)

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

(3)

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

Proof. abc □

Theorem 8.8 (Necessary and Sufficient Condition of Independence)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
*(1) If $X_1,\ X_2,\ \cdots, X_n$ are **discrete** r.v.'s, then $X_1,\ X_2,\ \cdots, X_n$ are independent*

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

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

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

Proof. abc □

Definition 8.5 (Indicator Function)**.** *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and $A \in \mathcal{A}$. The* ***indicator function*** *I_A of the event A is given by*

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

Theorem 8.9 (Indicator Function is a Discrete Measurable Function)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space. I_A is a* ***discrete r.v.*** *of $(\Omega, \mathcal{A}, \mathbb{P})$ for all $A \in \mathcal{A}$.*

Proof. abc □

Theorem 8.10 (Indicator R.V.'s Indicates Independence)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and $A_1, A_2, \cdots, A_n \in \mathcal{A}$. The events $A_1, A_2, \cdots, A_n$ are* ***independent*** *$\Leftrightarrow$ the* ***indicator r.v.'s*** *$I_{A_1}, I_{A_2}, \cdots, I_{A_n}$ are* ***independent****.*

Proof. abc □

Theorem 8.11 (Expectation of Measurable Functions of Independent R.V.)**.** *Suppose $X_1,\ X_2, \cdots, X_n$ are independent r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and $g_1, g_2, \cdots, g_n$ are measurable functions from $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$. Then $g_1(x_1), g_2(x_2), \cdots, g_n(X_n)$ are independent and*

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

Proof. abc □

Remark 8.3 (Independent Expectations Can't Imply Independence of R.V.'s)**.** *The converse is* ***not true****, i.e.,*

$$\mathbb{E}\left[\prod_{i=1}^{n} g_i(x_i)\right] = \prod_{i=1}^{n} \mathbb{E}[g_i(x_i)] \nRightarrow g_1(x_1),\ g_2(x_2), \cdots, g_n(x_n) \text{ are independent.}$$

8.3 Conditional Distributions

Lemma 8.1 (Properties of Conditional Probability)**.** *Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space, and $A, B, A_1, A_2, \cdots, A_n, B_1, B_2, \cdots, B_n \in \mathcal{A}$.*

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

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

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

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

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

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

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

Proof. abc □

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

Definition 8.6 (P.M.F. and C.D.F. of D-D)**.** *Let* X *and* Y *be discrete r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *and* $y \in \mathbb{R}$*. The conditional p.m.f.* $P_{X|Y}(x|y)$ *of* X *given that* $Y = y$ *is given by*

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

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

$$\begin{aligned} F_{X|Y}(x|y) &= \mathbb{P}(X \leqslant x|Y = y) \\ &= \sum_{t \leqslant X,\ t \in X(\Omega)} \mathbb{P}(X = t|Y = y) \\ &= \sum_{t \leqslant X,\ t \in X(\Omega)} P_{X|Y}(t|y), \forall x \in \mathbb{R}. \end{aligned}$$

Remark 8.4 (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 8.12 (Properties of D-D Conditional Probability)**.** *Suppose* $X, Y, X_1, X_2, \cdots, X_n$ *are discrete r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$*.*
(1) If $y \in \mathbb{R}$ *and* $P_Y(y) \neq 0$*, then* $P_{X|Y}(\cdot|y)$ *is a p.m.f.*
(2) $\forall x \in \mathbb{R}^n$,

$$P_X(x) = P_{X_1}(x_1)\cdot P_{X_2|X_1}(x_2|x_1)\cdots P_{X_n|X_1,X_2,\cdots,X_{n-1}}(x_n|x_1, x_2, \cdots, x_{n-1}).$$

(3) $\forall x \in \mathbb{R}$,

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

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

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

Proof. abc □

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

Definition 8.7 (C.D.F. and P.D.F. of C-C)**.** *Let X and Y be jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $y \in \mathbb{R}$. The conditional c.d.f. $F_{X|Y}(x|y)$ of X given that $Y = y$ is given by*

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

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

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

Remark 8.5 (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 8.13 (Properties of C-C Conditional Probability)**.** *Suppose $X, Y, X_1, X_2, \cdots, X_n$ are jointly continuous r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
(1) If $y \in \mathbb{R}$ and $f_Y(y) \neq 0$, then $f_{X|Y}(\cdot|y)$ is a p.d.f.
(2) $\forall x \in \mathbb{R}^n$,

$$f_X(x) = f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2|x_1) \cdots f_{X_n|X_1, X_2, \cdots X_{n-1}}(x_n|x_1, x_2, \cdots, x_{n-1}).$$

(3)

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

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

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

Proof. abc □

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

Definition 8.8 (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 $(\Omega, \mathcal{A}, \mathbb{P})$.*
The conditional ***c.d.f.*** $F_{X|Y}(\cdot|y)$ *of X given that $Y = y,\ y \in \mathbb{R}$ is given by*

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

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

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

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

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

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

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

Remark 8.6 (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 8.14 (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 $(\Omega, \mathcal{A}, \mathbb{P})$.*
(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)

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

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

(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

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

Proof. abc □

Definition 8.9 (Expectation of Conditional R.V.)**.** *Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $y \in \mathbb{R}$. The conditional expectation $\mathbb{E}[X|Y = y]$ of X given that $Y = y$ is given by*

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

Theorem 8.15 (Expectation of Conditional Measurable Function)**.** *Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$), and g is a measurable function from $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. Then*

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

Proof. abc □

8.4 Transformations of Two R.V.'s

Theorem 8.16 (Transformations of Two R.V.'s). *Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, 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

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

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

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

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

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

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

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

Proof. abc □

Theorem 8.17 (Convolution Theorem). *Suppose X and Y are two independent r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ 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

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

Proof. abc □

8.5 Order Statistics

Definition 8.10 (Order Statistic)**.** *Let* $X_1,\ X_2,\ \cdots,X_n$ *be i.i.d. r.v.'s of a probability space* $(\Omega,\mathcal{A},\mathbb{P})$*. The* i^{th} *order statistic* $X_{(i)},\quad i=1,2,\ \cdots,n$ *of* $X_1,\ X_2,\ \cdots,X_n$ *is defined as the* i^{th} ***smallest*** *value in* $\{X_1,\ X_2,\ \cdots,X_n\}$ *so that* $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$*, namely,* $X_{(i)}(w)=$ *the* i^{th} *smallest value in* $\{X_1(w),\ X_2(w),\ \cdots,X_n(w)\}$ *for all* $w\in\Omega$*. In particular,* $X_{(1)}=\min\{X_1,\ X_2,\ \cdots,X_n\}$ *and* $X_{(n)}=\max\{X_1,\ X_2,\ \cdots,X_n\}$.

Remark 8.7 (Without Equal & Not I.I.D.)**.** *(1) If* $X_1,\ X_2,\ \cdots,X_n$ *are jointly continuous r.v.'s, then*

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

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

Definition 8.11 (Random Sample)**.** *A random sample of size* n *of a probability space* $(\Omega,\mathcal{A},\mathbb{P})$ *is a sequence of* n *i.i.d. r.v.'s* $X_1,\ X_2,\ \cdots,X_n$ *of* $(\Omega,\mathcal{A},\mathbb{P})$*.*

Definition 8.12 (Range, Midrange, Median and Mean of Random Sample)**.** *Let* $X_1,\ X_2,\ \cdots,X_n$ *be a random sample of size* n *of a probability space* $(\Omega,\mathcal{A},\mathbb{P})$*.*

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* $\begin{cases} X_{(i-1)}, & \text{if } n=2i+1\\ \dfrac{X_{(i)}+X_{(i+1)}}{2}, & \text{if } n=2i\end{cases}$

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

Remark 8.8 (Forced Decline)**.** *If* $\exists\, i_j<i_l\to x_{i_j}\geqslant x_{i_l}$*, then*

$$\begin{aligned}&F_{X_{(i_1)},X_{(i_2)},\cdots X_{(i_k)}}\left(x_{i_1},\ \cdots,x_{i_j},\ \cdots,x_{i_l},\ \cdots,\ x_{i_k}\right)\\ =&F_{X_{(i_1)},X_{(i_2)},\cdots X_{(i_k)}}\left(x_{i_1},\ \cdots,x_{i_l},\ \cdots,x_{i_l},\ \cdots,\ x_{i_k}\right)\end{aligned}$$

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

Theorem 8.18 (C.D.F. and P.D.F. of Jointly Order R.V.'s)**.** *Suppose* $X_1,\ X_2,\ \cdots,X_n$ *are i.i.d. jointly continuous r.v.'s of a probability space* $(\Omega,\mathcal{A},\mathbb{P})$ *with common c.d.f.* $F(x)$ *and common p.d.f.* $f(x)$*. If* $1\leqslant i_1\leqslant i_2\leqslant\cdots\leqslant i_k\leqslant n,\ -\infty<x_{i_1}<x_{i_2}<\cdots<x_{i_k}<\infty$*, then*

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

and

$$
\begin{aligned}
& f_{X_{(i_1)},X_{(i_2)},\cdots,X_{(i_k)}}\left(x_{i_1},x_{i_2},\,\cdots,\,x_{i_k}\right) \\
= & \frac{n!}{(i_1-1)!\,(i_2-i_1-1)!\cdots(i_k-i_{k-1}-1)!\,(n-i_k)!} \\
& \cdot f\left(x_{i_1}\right) f\left(x_{i_2}\right)\cdots f\left(x_{i_k}\right)\cdot\left[F\left(x_{i_1}\right)\right]^{i_1-1}\left[F\left(x_{i_2}\right)-F\left(x_{i_1}\right)\right]^{i_2-i_1-1} \\
& \cdots\left[F\left(x_{i_k}\right)-F\left(x_{i_{k-1}}\right)\right]^{i_k-i_{k-1}-1}\left[1-F\left(x_{i_k}\right)\right]^{n-i_k}
\end{aligned}
$$

Proof. abc □

Corollary 8.3 (Beta R.V. vs Binomial R.V.). *Suppose* $X_1,\ X_2,\ \cdots,X_n$ *are i.i.d. r.v.'s* $\sim U(0,1)$*, then*

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

Proof.

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

□

Corollary 8.4 (Cases One, Two and n Order R.V.'s). *(1)*

$$
\begin{aligned}
F_{X_{(i)}}(x) &= \sum_{j=i}^{n}\binom{n}{j}\left[F(x)\right]^{j}\left[1-F(x)\right]^{n-j},\ -\infty<x<\infty, \\
f_{X_{(i)}}(x) &= \frac{n!}{(i-1)!\,(n-i)!}f(x)\left[F(x)\right]^{i-1}\left[1-F(x)\right]^{n-i},\ -\infty<x<\infty.
\end{aligned}
$$

In particular,

$$
\begin{aligned}
F_{X_{(1)}}(x) &= 1-\left[1-F(x)\right]^{n},\ -\infty<x<\infty, \\
f_{X_{(1)}}(x) &= n\cdot f(x)\left[1-F(x)\right]^{n-1},\ -\infty<x<\infty,
\end{aligned}
$$

and

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

(2)

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

$$
\begin{aligned}
f_{X_{(i_1)},X_{(i_2)}}(x,y)=&\frac{n!}{(i_1-1)!\,(i_2-i_1-1)!\,(n-i_2)!}f(x)f(y)\,[F(x)]^{j_1}\\
&\cdot[F(y)-F(x)]^{j_2-j_1}[1-F(y)]^{n-j_2},-\infty<x<y<\infty
\end{aligned}
$$

(3)

$$
\begin{aligned}
&F_{X_{(1)},X_{(2)},\cdots,X_{(n)}}(x_1,x_2,\cdots,x_n)\\
=&\sum_{j_{n-1}=i_{n-1}}^{n}\sum_{j_{n-2}=i_{n-2}}^{j_{n-1}}\cdots\sum_{j_1=i_1}^{j_2}\binom{n}{j_{n-1}}\binom{j_{n-1}}{j_{n-2}}\cdots\binom{j_2}{j_1}[F(x_1)]^{j_1}\\
&\cdot[F(x_2)-F(x_1)]^{j_2-j_1}\cdots[F(x_{n-1})-F(x_{n-2})]^{j_{n-1}-j_{n-2}}\;[F(x_n)-F(x_{n-1})]^{n-j_{n-1}}
\end{aligned}
$$

and

$$
\begin{aligned}
&f_{X_{(1)},X_{(2)},\cdots,X_{(n)}}(x_1,x_2,\cdots,x_n)\\
=&n!f(x_1)f(x_2)\cdots f(x_n),\quad -\infty<x_1<x_2<\cdots<x_n<\infty
\end{aligned}
$$

Proof. abc □

8.6 Multinomial Distributions

Consider an experiment with k possible outcomes $\omega_1,\ \omega_2,\ \cdots,\omega_k$. Let $A_{(i)}=\{\omega_i\}$ be the event that the outcome is ω_i and let $P_i=\mathbb{P}(A_i), i=1,2,\ \cdots,k$. Suppose that such an experiment is independently and successively performed n times. Let $X_i,\ i=1,2,\ \cdots,k$ be the number of times that event A_i occurs. Then

$$
\begin{aligned}
&P_{X_1,X_2,\cdots,X_k}(x_1,x_2,\cdots,x_k)\\
=&\mathbb{P}(X_1=x_1,X_2=x_2,\cdots,X_k=x_k)\\
=&\frac{n!}{x_1!x_2!\cdots x_k!}P_1^{x_1}P_2^{x_2}\cdots P_k^{x_k},\ x_1,x_2,\ \cdots,\ x_k\geqslant 0 \text{ and } \sum_{i=1}^{k}x_i=n.
\end{aligned}
$$

Definition 8.13 (Multinomial Joint R.V.'s)**.** *Let $X_1,\ X_2,\ \cdots,X_k$ be discrete r.v.'s of a probability space $(\Omega,\mathcal{A},\mathbb{P})$. We call $X_1,\ X_2,\ \cdots,X_k$ multinomial joint r.v.'s with parameters $n,P_1,P_2,\ \cdots,P_k$, where $n\geqslant 1,\ P_1,P_2,\ \cdots,P_k\geqslant 0,\ P_1+P_2+\cdots+P_k=1$, if the joint p.m.f. is given by*

$$
P_{\boldsymbol{X}}(\boldsymbol{x})=\begin{cases}\dfrac{n!}{x_1!x_2!\cdots x_k!}P_1^{x_1}P_2^{x_2}\cdots P_k^{x_k}, & x_1,x_2,\ \cdots,x_k\geqslant 0 \textit{ and } \displaystyle\sum_{i=1}^{k}x_i=n\\ 0, & \textit{o.w.}\end{cases}
$$

Remark 8.9 (Verification of P.M.F.). *$P_{\boldsymbol{X}}(\boldsymbol{x}) \geqslant 0,\ \forall \boldsymbol{x} \in \mathbb{R}^n$ and*

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

$\to P_{\boldsymbol{X}}(\boldsymbol{x})$ is a p.m.f.

Theorem 8.19 (Splitting of Multinomial Joint R.V.'s). *Suppose $X_1,\ X_2,\ \cdots, X_l$ are multinomial r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with parameters $n, P_1, P_2, \cdots, P_l$, where $n \geqslant 1,\ P_1, P_2, \cdots, P_k \geqslant 0,\ P_1 + P_2 + \cdots + P_k = 1$. 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

$$n, P_{i_1}, P_{i_2}, \cdots, P_{i_k},\ 1 - P_{i_1} - P_{i_2} - \cdots - P_{i_k}.$$

Proof. abc □

Chapter 9

More Expectations and Variance

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

Theorem 9.1 (Expectations of Sum of Finite R.V.'s)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then*

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

Proof. abc ☐

Theorem 9.2 (Expectations of Sum of Infinite R.V.'s)**.** *Suppose $X_1,\ X_2,\ \cdots$ are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If*

$$\sum_{i=1}^{\infty} \mathbb{E}[X_i] < \infty$$

or if X_i is nonnegative for all $i = 1, 2,\ \cdots$, then

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

Proof. abc ☐

Remark 9.1 (General Expectations of Sum of Infinite R.V.'s)**.** *In general,*

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

Corollary 9.1 (Expectation of Integer-Valued R.V.)**.** *Suppose X is an integer-valued r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then*

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

Proof. abc ☐

9.2 Covariance and Correlation Coefficients

Theorem 9.3 (Cauchy-Schwarz Inequality)**.** *Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and suppose $\mathbb{E}[X^2]$ and $\mathbb{E}[Y^2]$ exists. Then*

$$|\mathbb{E}[XY]| \leqslant \sqrt{\mathbb{E}\left[X^2\right] \cdot \mathbb{E}\left[Y^2\right]}.$$

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

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

Proof. abc ☐

Remark 9.2 (Cauchy-Schwarz Equalities)**.** *Suppose that $\mathbb{E}[X^2] \neq 0$ and $\mathbb{E}[Y^2] \neq 0$, then*

$$\mathbb{E}[XY] = \sqrt{\mathbb{E}\left[X^2\right] \cdot \mathbb{E}\left[Y^2\right]} \Leftrightarrow Y = aX$$

with probability 1, where

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

$$\mathbb{E}[XY] = -\sqrt{\mathbb{E}\left[X^2\right] \cdot \mathbb{E}\left[Y^2\right]} \Leftrightarrow Y = aX$$

with probability 1, where

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

Corollary 9.2 (Variance Larger Than or Equal to Zero)**.** *Suppose X is a r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and suppose $\mathbb{E}[X^2]$ exists, then*

$$|\mathbb{E}[X]|^2 \leqslant \mathbb{E}\left[X^2\right].$$

Proof. abc ☐

Definition 9.1 (Covariance)**.** *Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with means μ_X and μ_Y, resp. The covariance $\operatorname{Cov}(X, Y)$ (or $\sigma_{X,Y}$) of X and Y is given by*

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

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

Remark 9.3 (Covariance of Linear Combination of Two R.V.'s)**.** *(1)* $\mathrm{Var}(X) = \mathbb{E}[(X-\mu_X)^2]$ *is a measure of the spread or dispersion of* X.

$\mathrm{Var}(Y) = \mathbb{E}[(Y - \mu_Y)^2]$ *is a measure of the spread or dispersion of* Y.

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

(2)

$$\begin{aligned}\mathrm{Var}(aX + bY) &= \mathbb{E}[[(aX + bY) - (a\mu_X + b\mu_Y)]^2] \\ &= \mathbb{E}[[a(X - \mu_X) + b(Y - \mu_Y)]^2] \\ &= a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)\end{aligned}$$

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

Theorem 9.4 (Calculating Covariance)**.** *Suppose* X *and* Y *are r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$.

(1) $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

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

(3) $|\mathrm{Cov}(X, Y)| \leqslant \sigma_X \cdot \sigma_Y$, *"="* $\Leftrightarrow X = \mu_X$ *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 $\sigma_X \neq 0$ *and* $\sigma_Y \neq 0$*, then*

$$\mathrm{Cov}(X, Y) = \sigma_X \cdot \sigma_Y \Leftrightarrow Y = aX + b$$

with probability 1, where

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

$$\mathrm{Cov}(X, Y) = -\sigma_X \cdot \sigma_Y \Leftrightarrow Y = aX + b$$

with probability 1, where

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

Proof. abc □

Theorem 9.5 (Covariance of Two Linear Combined R.V.'s)**.** *Suppose* $X_1, X_2, \cdots, X_n, Y_1, Y_2, \cdots, Y_m$ *are r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$.

(1)

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

(2)

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

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

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

Proof. abc □

Theorem 9.6 (Independence Implies Uncorrelated)**.** *Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. If $X \perp Y$, then X and Y are uncorrelated, i.e.,*

$$\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$$

Proof. abc □

Remark 9.4 (Uncorrelated Can't Imply Independence)**.** *The inverse is not true, i.e.,*

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

Definition 9.2 (Correlation Coefficient)**.** *Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $0 < \sigma_X^2 < \infty, 0 < \sigma_Y^2 < \infty$. The correlation coefficient between X and Y is given by*

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

Remark 9.5 (Properties of Correlation Coefficient)**.** *(1) $X^* = \frac{X - \mu_X}{\sigma_X}$ 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) $-1 \leqslant \rho_{X,Y} \leqslant 1$.

$\rho_{X,Y} = 1 \Leftrightarrow Y = aX + b$ 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 (Conditional Expectation on R.V.'s)**.** *Let X and Y be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.*
Let $g(Y) = \mathbb{E}[X|Y = y]$, $\forall y \in \mathbb{R}$. We denote $\mathbb{E}[X|Y]$ as the r.v. $g(Y)$. Note that $\mathbb{E}[X|Y]$ is a function of Y.

Theorem 9.7 (Marginal Expectation)**.** *Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then*

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

Proof. abc □

Theorem 9.8 (Marginal Expectation of Measurable Function)**.** *Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then*

$$\mathbb{E}\left[\mathbb{E}\left[X \cdot g(Y)|Y\right]\right] = g(Y)\mathbb{E}\left[X|Y\right].$$

Proof. abc □

Theorem 9.9 (Wald's Equations)**.** *Suppose $X_1,\ X_2,\ \cdots$ are i.i.d. r.v.'s $\sim X$ and N is a positive integer-valued r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and $N \perp \{X_1,\ X_2,\ \cdots\}$.*

(1) If $\mathbb{E}[X] < \infty$ and $\mathbb{E}\left[N\right] < \infty$, then

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

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

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

Proof. abc □

Theorem 9.10 (Law of Total Probability)**.** *Suppose A is an event and X is a r.v. of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then*

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

Proof. abc □

Theorem 9.11 (Conditional Variance on R.V.'s)**.** *Suppose X and Y are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then*

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(x|y)] + \mathrm{Var}(\mathbb{E}[X|Y]).$$

Proof. abc □

9.4 Bivariate Normal (Gaussian) Distribution

Definition 9.4 (Bivariate Normal (Gaussian) R.V.'s)**.** *Let X_1 and X_2 be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We call X_1 and X_2 jointly normal (Gaussian) r.v.'s with parameters*

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} > 0,$$

where ">0" means positive definite, denoted

$$\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

if their joint p.d.f. is given by

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

where

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

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

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

Theorem 9.12 (Explicitly Normal (Gaussian) R.V.)**.** *Suppose X_1 and X_2 are r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and suppose $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.*

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

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

(2)

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

and

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

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

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

and

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

Proof. abc □

Remark 9.6 (Mean Vector and Covariance Matrix). $\boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$ *is called the mean vector of* $\boldsymbol{X}$, *and* $\boldsymbol{\Sigma} = \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}$ *is called the covariance matrix of* $\boldsymbol{X}$.

Lemma 9.1 (Linear Conditional Expectation and Constant Variance). *Suppose* X_1 *and* X_2 *are jointly continuous r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *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 $\mathbb{E}[X_2|X_1 = x_1] = ax_1 + b$ *is a linear function in* x_1, *then*

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

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

$$\mathrm{Var}(X_2|X_1 = x_1) = \left(1-\rho^2\right)\sigma_2^2.$$

Proof. abc □

Theorem 9.13 (Derivation of Jointly Normal R.V.'s). *Suppose* X_1 *and* X_2 *are r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$. *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) $\mathbb{E}[X_2|X_1 = x_1]$ *is a linear function in* X_1, *and* $\mathrm{Var}(X_2|X_1 = x_1) = \sigma^2$ *is a constant.*

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

Proof. abc □

Theorem 9.14 (Independence mutually Implies Uncorrelated). *Suppose* X_1 *and* X_2 *are jointly normal r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$. *Then* X_1 *and* X_2 *are independent* $\Leftrightarrow$ X_1 *and* X_2 *are uncorrelated.*

Proof. abc □

Theorem 9.15 (Linearly Generated Normal R.V.). *Suppose* $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu_X}, \boldsymbol{\Sigma_X})$ *and* $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + b$, *where* $\boldsymbol{A}$ *is* ***nonsingular***, *i.e.,* $|\boldsymbol{A}| \neq 0$. *Then*

$$\boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{A}\boldsymbol{\mu_X} + b, \boldsymbol{A}\boldsymbol{\Sigma_X}\boldsymbol{A}^\top\right).$$

Proof. abc □

Chapter 10

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

10.1 Moment Generating Functions

Definition 10.1 (Moment Generating Function)**.** *The moment generating function (m.g.f.) $M_X(t)$ of a r.v. X is given by $M_X(t) = \mathbb{E}[e^{tx}]$ if $\exists \delta > 0 \to M_X(t)$ is defined for all $t \in (-\delta, \delta)$.*

Theorem 10.1 (Moment Generation)**.** *(1) $\mathbb{E}[X^n] = M_X^{(n)}(0),\ \forall n \geqslant 0$.*

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

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

Proof. abc □

Remark 10.1 (Sufficient Condition for n^{th} Moment to Converge)**.** *If $|M_X(t)| < \infty$ for some $t > 0$, then $|\mathbb{E}[X^n]| < \infty$ for all $n \geqslant 1$. But the converse is not true.*

Theorem 10.2 (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.*

Proof. abc □

10.2 Sums of Independent R.V.'s

Theorem 10.3 (M.G.F. of Sums of Independent R.V.'s)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are **independent** r.v.'s with m.g.f.'s*

$$M_{X_1}(t), M_{X_2}(t),\ \cdots, M_{X_n}(t)$$

respectively. Then the m.g.f. of their ***sum*** $X = X_1 + X_2 + \cdots + X_n$ *is*

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

Proof. abc □

Theorem 10.4 (M.G.F. of Sums of Normal R.V.'s)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are* ***independent*** *r.v.'s and $X_i \sim \mathcal{N}\left(\mu_i,\ \sigma_i^2\right),\ \forall i = 1, 2,\ \cdots, n$ and suppose $a_1, a_2,\ \cdots, a_n \in \mathbb{R}$. If*

$$X = \sum_{i=1}^{n} a_i X_i,$$

then

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

Proof. abc □

Corollary 10.1 (M.G.F. of Sums of I.I.D. Normal R.V.'s)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are* ***i.i.d.*** $\sim \mathcal{N}\left(\mu, \sigma^2\right)$*, then*

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

Proof. abc □

10.3 Markov and Chebyshev Inequalities

Theorem 10.5 (Markov's Inequality)**.** *Suppose X is a nonnegative r.v., then*

$$\mathbb{P}(X \geqslant t) \leqslant \frac{\mathbb{E}[X]}{t},\ \forall t > 0.$$

Proof. abc □

Theorem 10.6 (Chebyshev's Inequality)**.**

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

In particular,

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

Proof. abc □

Remark 10.2 (Not Tight Bounds)**.** *The bounds obtained by Markov and Chebyshev inequalities are usually* ***not very tight****.*

Theorem 10.7 (Zero Absolute Moment)**.**

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

Proof. abc □

Corollary 10.2 (Zero Variance)**.**

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

Proof. abc □

Theorem 10.8 (Chebyshev's Inequality for I.I.D R.V.'s)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are* ***i.i.d.*** *r.v.'s with mean μ and variance $\sigma^2 < \infty$. Let*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

be the sample mean of $X_1,\ X_2,\ \cdots, X_n$. Then

$$\mathbb{P}\left(\left|\bar{X} - \mu\right| \geqslant \epsilon\right) \leqslant \frac{\sigma^2}{n\epsilon^2}.$$

Proof. abc □

Theorem 10.9 (Chebyshev's Inequality for I.I.D. Bernoulli R.V.'s)**.** *Suppose $X_1,\ X_2,\ \cdots, X_n$ are i.i.d. $\sim$ Bernoulli(p). Let*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

be the sample mean of $X_1,\ X_2,\ \cdots, X_n$. Then

$$\mathbb{P}\left(\left|\bar{X} - p\right| \geqslant \epsilon\right) \leqslant \frac{p(1-p)}{n\epsilon^2} \leqslant \frac{1}{4n\epsilon^2}.$$

Proof. abc □

10.4 Laws of Large Numbers (LLN's)

Definition 10.2 (Converge in Probability)**.** *Let $X, X_1,\ X_2, \cdots$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We say that X_n converges to X* ***in probability****, denoted*

$$X_n \xrightarrow{P} X,$$

if

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \epsilon) = 1,\ \forall \epsilon > 0,$$

or

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0,\ \forall \epsilon > 0.$$

Theorem 10.10 (Weak Law of Large Numbers (WLLN))**.** *Suppose $X_1,\ X_2,\ \cdots$ are i.i.d. r.v.'s with mean μ and variance $\sigma^2 < \infty$. Then*

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu,$$

i.e.,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\overline{X_n} - \mu\right| > \epsilon\right) = 0,\ \forall \epsilon > 0.$$

Proof. abc □

Remark 10.3 (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*

$$X_i = \begin{cases} 1, & \textit{if } A \textit{ occurs on the } i^{th} \textit{ repetition,} \\ 0, & \textit{o.w.} \end{cases}$$

Then

$$n(A) = \sum_{i=1}^{n} X_i \ \ \textit{and}\ \ \mathbb{E}[X_i] = 1\cdot\mathbb{P}(A) + 0\cdot\mathbb{P}(A^c) = \mathbb{P}(A).$$

$$\to \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{n(A)}{n} - \mathbb{P}(A)\right| > \epsilon\right) = \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{P}(A)\right| > \epsilon\right) = 0.$$

Therefore, the relative frequency $\frac{n(A)}{n}$ of occurrence of A is very likely close to $\mathbb{P}(A)$ if n is sufficiently large.

Definition 10.3 (Converge Almost Surely)**.** *Let $X, X_1,\ X_2, \cdots$ be r.v.'s of a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We say that X_n converges to X* ***almost surely*** *(a.s.), denoted*

$$X_n \xrightarrow{a.s.} X,$$

if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

Theorem 10.11 (Strong Law of Large Numbers (SLLN))**.** *Suppose $X_1,\ X_2,\ \cdots$ are i.i.d. r.v.'s with mean μ. Then*

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu$$

i.e.,

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{X_n} = \mu\right) = 1.$$

Proof. abc □

Remark 10.4 (Relative Frequency Converges Almost Surely)**.**

$$\mathbb{P}\left(\lim_{n\to\infty}\frac{n(A)}{n}=\mathbb{P}(A)\right)=1 \ \rightarrow \ \lim_{n\to\infty}\frac{n(A)}{n}=\mathbb{P}(A) \textit{ with probability } 1.$$

Theorem 10.12 (Converge Almost Surely Implies Convergence in Probability)**.**

$$\textit{If } X_n \xrightarrow{a.s.} X, \ \textit{ then } \ X_n \xrightarrow{P} X.$$

Proof. abc □

10.5 Central Limit Theorem (CLT)

Theorem 10.13 (Levy Continuity Theorem)**.** *Suppose* $X, X_1, \ X_2, \ \cdots$ *are r.v.'s of a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$.

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.

Proof. abc □

Theorem 10.14 (Central Limit Theorem (CLT))**.** *Suppose* $X_1, \ X_2, \ \cdots, X_n$ *are i.i.d. r.v.'s with mean* μ *and variance* σ^2*. Let*

$$S_n^* = \frac{X_1 + X_2 + \cdots + X_n - \mathbb{E}[S_n]}{\sigma_{S_n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then

$$\lim_{n\to\infty} F_{S_n^*}(X) = \Phi(x),$$

i.e.,

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leqslant x\right) = \Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,\mathrm{d}y.$$

Equivalently,

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leqslant x\right) &= \lim_{n\to\infty}\mathbb{P}\left(\frac{\bar{X}-\mu}{\sqrt{\frac{\mathrm{Var}(X)}{n}}} \leqslant x\right)\\ &= \lim_{n\to\infty}\mathbb{P}\left(\frac{\bar{X}-\mathbb{E}[\bar{X}]}{\sigma_{\bar{X}}} \leqslant x\right)\\ &= \Phi(x).\end{aligned}$$

Proof. abc □