Revision notes - ST2131

Ma Hongqiang

April 18, 2017

Contents

1 Combinatorial Analysis

Theorem 1.1 (Generalised Basic Principle of Counting)**.**
Suppose that r experiments are to be preformed. If

- experiment 1 can result in n_1 possible outcomes;
- experiment 2 can result in n_2 possible outcomes;
- $\cdots$
- experiment r can result in n_r possible outcomes;

then together there are $n_1 n_2 \cdots n_r$ possible outcomes of the r experiments.

1.1 Permutations

Theorem 1.2 (Permutation of distinct objects)**.**
Suppose there are n distinct objects, then the total number of permutations is

$$n!$$

.

Theorem 1.3 (General principle of permutation)**.**
For n objects of which n_1 are alike, n_2 are alike, ..., n_r are alike, there are

$$\frac{n!}{n_1! n_2! \cdots n_r!}$$

different permutations of the n objects.

1.2 Combinations

Theorem 1.4 (General principle of combination)**.**
If there are n distinct objects, of which we choose a group of r items, number of combinations equals

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

1.2.1 Useful Combinatorial Identities

1. For $1 \leq r \leq n, \binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$
2. (**Binomial Theorem**) Let n be a non-negative integer, then

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

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

1.3 Multinomial Coefficients

If $n_1 + n_2 + \cdots + n_r = n$, we define $\binom{n}{n_1, n_2, \cdots, n_r}$ by

$$\binom{n}{n_1, n_2, \cdots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

Thus $\binom{n}{n_1, n_2, \cdots, n_r}$ represents the number of possible divisions of n distinct objects into r distinct groups of respective sizes $n_1, n_2, \cdots, n_r$.

Theorem 1.5 (Multinomial Theorem)**.**

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{(n_1, \ldots, n_r): n_1 + \cdots + n_r = n} \binom{n}{n_1, n_2, \cdots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

1.4 Number of Integer Solutions of Equations

Theorem 1.6. There are $\binom{n-1}{r-1}$ distinct *positive* integer-valued vectors $(x_1, x_2, \ldots, x_r)$ that satisfies the equation

$$x_1 + x_2 + \cdots + x_r = n$$

where $x_i > 0$ for $i = 1, \ldots, r$

Theorem 1.7. There are $\binom{n+r-1}{r-1}$ distinct *non-negative* integer-valued vectors $(x_1, x_2, \ldots, x_r)$ that satisfies the equation

$$x_1 + x_2 + \cdots + x_r = n$$

where $x_i > 0$ for $i = 1, \ldots, r$

2 Axioms of Probability

2.1 Sample Space and Events

The basic objects of probability is an **experiment**: an activity or procedure that produces distinct, well-defined possibilities called **outcomes**.
The **sample space** is the set of all possible outcomes of an experiment, usually denoted by S.
Any subset E of the sample space is an **event**.

2.2 Axions of probability

Probability, denoted by P, is a function on the collection of events satisfying

(i) For any event A,

$$0 \leq P(A) \leq 1$$

(ii) Let S be the sample space, then

$$P(S) = 1$$

(iii) For any sequence of mutually exclusive events $A_1, A_2, \ldots$(that is $A_iA_j = \varnothing$ when $i \neq j$)

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

2.3 Properties of Probability

Theorem 2.1. $P(\varnothing) = 0$.

Theorem 2.2. For any finite sequence of mutually exclusive event $A_1, A_2, \ldots, A_n$,

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

Theorem 2.3. Let A be an event, then

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

Theorem 2.4. If $A \subseteq B$, then

$$P(A) + P(BA^c) = P(B)$$

Theorem 2.5 (Inclusion-exclusion Principle)**.** Let $A_1, A_2, \ldots, A_n$ be any events, then

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) = & \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i_1 \leq i_2 \leq n} P(A_{i_1}A_{i_2}) + \cdots \\ & + (-1)^{r+1} \sum_{1 \leq i_1 \leq \cdots \leq i_r \leq n} P(A_{i_1} \cdots A_{i_r}) \\ & + \cdots + (-1)^{n+1} P(A_1 \cdots A_n) \end{aligned}$$

2.4 Probability as a Continuous Set Function

Definition 2.1. A sequence of events $\{E_n\}, n \geq 1$ is said to be an **increasing** sequence if

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

whereas it is said to be a **decreasing** sequence if

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

Definition 2.2. If $\{E_n\}, n \geq 1$ is an increasing sequence of events,
define new event, denoted by $\lim_{n\to\infty} E_n$ as

$$\lim_{n\to\infty} E_n = \cup_{i=1}^{\infty} E_i$$

Similarly, if $\{E_n\}, n \geq 1$ is an decreasing sequence of events, define new event, denoted by $\lim_{n\to\infty} E_n$ as

$$\lim_{n\to\infty} E_n = \cap_{i=1}^{\infty} E_i$$

Theorem 2.6. If $\{E_n\}, n \geq 1$ is either an increasing or a decreasing sequence of events, then

$$\lim_{n\to\infty} P(E_n) = P\left(\lim_{n\to\infty} E_n\right)$$

3 Conditional Probability and Independence

3.1 conditional Probabilities

Definition 3.1. Let A and B be two events. Suppose that $P(A) > 0$, the **conditional probability** of B given A is defined as

$$\frac{P(AB)}{P(A)}$$

and is denoted by $P(B|A)$.
Suppose $P(A) > 0$, then $P(AB) = P(A)P(B|A)$.

Theorem 3.1 (General Multiplication Rule)**.**
Let $A_1, A_2, \ldots, A_n$ be n events, then

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1})$$

3.2 Bayes' Formulas

Let A and B be any two events, then

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Definition 3.2. We say that $A_1, A_2, \ldots, A_n$ **partitions** the sample space S if:

- They are **mutually exclusive**, meaning $A_i \cap A_j = \varnothing \; \forall i \neq j$.
- THey are collectively exclusive, meaning $\cup_{i=1}^{n} A_i = S$

Theorem 3.2 (Bayes' First Formula)**.**
Suppose the events $A_1, A_2, \ldots, A_n$ partitions the sample space. Assume further that $P(A_i) > 0$ for $1 \leq i \leq n$. Let B be any event, then

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)$$

Theorem 3.3 (Bayes' Second Formula)**.**
Suppose the events $A_1, A_2, \ldots, A_n$ partitions the sample space. Assume further that $P(A_i) > 0$ for $1 \leq i \leq n$. Let B be any event, then for any $1 \leq i \leq n$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n)}$$

3.3 Independent Events

Definition 3.3. Two events A and B are said to be **independent** if

$$P(AB) = P(A)P(B)$$

They are said to be dependent otherwise.

Theorem 3.4. If A and B are independent, then so are

1. A and B^c
2. A^c and B
3. A^c and B^c

Definition 3.4. Events $A_1, A_2, \ldots, A_n$ are said to be **independent**, if for eveery subcollection of events $A_{i_1}, \ldots, A_{i_r}$, we have

$$P(A_{i_1} \cdots A_{i_r}) = P(A_{i_1})P(A_{i_r})$$

3.4 $P(\cdot|A)$ is a Probability

Theorem 3.5. Let A be an event with $P(A) > 0$. Then the following three conditions hold.

1. For any event B, we have
$$0 \leq P(B|A) \leq 1$$

2. $$P(S|A) = 1$$

3. Let $B_1.B_2, \ldots$ be a sequence of mutually exclusive events, then
$$P(\cup_{k=1}^{\infty} B_k|A) = \sum_{k=1}^{\infty} P(B_k|A)$$

4 Random Variables

4.1 Random Variables

Definition 4.1. A **random variable** X, is a mapping from the sample space to real numbers.

4.2 Discrete Random Variables

Definition 4.2. A random variable is said to be **discrete** if the range of X is either finite or countably infinite.

Definition 4.3. Suppose that random variable X is discrete, taking values $x_1, x_2 \ldots$, then the **probability mass function** of X, denoted by p_X, is defined as

$$p_X(x) = \begin{cases} P(X = x) & \text{if } x = x_1, x_2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

Properties of the probability mass function include

1. $p_X(x_i) \geq 0$ for $i = 1, 2, \ldots$;
2. $p_X(x) = 0$ for other values of x;
3. Since X must take on one of the values of x_i, $\sum_{i=1}^{\infty} p_X(x_i) = 1$.

Definition 4.4. The **cumulative distribution function** of X, is defined as

$$F_X : \mathbb{R} \to \mathbb{R}$$

where

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

Remark: SUppose that X is discrete and takes values $x_1, x_2, \ldots$ where $x_1 < x_2 < x_3 < \cdots$. Note then that F is a step function, that is, F is constant in the interval $[x_{i-1}, x_i)$ (F takes value $p(x_1) + \cdots + p(x_{i-1})$), and then take a jump of size $= p(x_i)$.

4.3 Expected Value

Definition 4.5. If X is a discrete random variable having a probability mass function p_X, the **expectation** or **expected value** of X, denoted by $\mathrm{E}(X)$ or μ_X is defined by

$$\mathrm{E}(X) = \sum_x x p_X(x)$$

Definition 4.6 (Bernoulli Random Variable)**.**
Suppose X takes only two values 0 and 1 with

$$P(X = 0) = 1 - p \quad \text{and} \quad P(X = 1) = p$$

We call this random variable a **Bernoulli** random variable of parameter p. And we denote it by $X \sim \mathrm{Be}(p)$.

4.4 Expectation of a Function of a Random Variable

Theorem 4.1. If X is a discrete random variable that takes values x_i, $i \geq 1$, with respective probabilities $p_X(x_i)$, then for any real value function g

$$\begin{aligned}\mathrm{E}[g(x)] &= \sum_i g(x_i)p_X(x_i) \quad \text{or equivalently} \\ &= \sum_x g(x)p_X(x)\end{aligned}$$

Theorem 4.2. Let a and b be constants, then

$$\mathrm{E}[aX + b] = a\mathrm{E}(X) + b$$

Theorem 4.3 (Tail Sum Formula for Expectation)**.**
For *nonnegative integer-valued* random variable X,

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} P(X > k)$$

4.5 Variance and Standard Deviation

Definition 4.7. If X is a random variable with mean μ, then the **variance** of X, denoted by $\mathrm{Var}(X)$, is defined by

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

An alternative formula for variance is:

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

Remark:

1. $\mathrm{Var}(X) \geq 0$
2. $\mathrm{Var}(X) = 0$ if and only if X is a degenerate random variable
3. $\mathrm{E}(X^2) \geq [\mathrm{E}(X)]^2 \geq 0$
4. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$

4.6 Discrete Random Variable arising from Repeated Trials

1. **Bernoulli random variable**, denoted by $\mathrm{Be}(p)$.
   We only perform the Bernoulli Trial once and define

   $$X = \begin{cases} 1 & \text{if it is a success} \\ 0 & \text{if it is a failure} \end{cases}$$

   Here,

   $$P(X = 1) = p, P(X = 0) = 1 - p$$

   and

   $$\mathrm{E}(X) = p, \quad \mathrm{Var}(X) = p(1 - p)$$

2. **Binomial random variable**,denoted by $\text{Bin}(n, p)$
   We perform the experiment (under identical conditions and independently) n times and define

   $$X = \text{number of successes in } n \text{ Bernoulli}(p) \text{ trials}$$

   Therefore, X takes values $0, 1, \ldots, n$. In fact, for $0 \leq k \leq n$,

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

   Here,

   $$\text{E}(X) = np, \qquad \text{Var}(X) = np(1 - p)$$

   Also, a useful fact is

   $$\frac{P(X = k + 1)}{P(X = k)} = \frac{p}{1 - p} \frac{n - k}{k + 1}$$

3. **Geometric random variable**, denoted by $\text{Geom}(p)$
   Define the random variable

   $$X = \text{number of Bernoulli}(p) \text{ trials required to obtain the first success}$$

   Here X takes values $1, 2, 3, \ldots$ and so on. In fact, for $k \geq 1$,

   $$P(X = k) = pq^{k-1}$$

   And,

   $$\text{E}(X) = \frac{1}{p} \qquad \text{Var}(X) = \frac{1 - p}{p^2}$$

4. **Negative Binomial random variable**, denoted by $\text{NB}r, p$
   Define the random variable

   $$X = \text{number of Bernoulli}(p) \text{ trials required to obtain } r \text{ success.}$$

   Here, X takes values $r, r + 1, \ldots$ and so on. In fact, for $k \geq r$,

   $$P(X = k) = \binom{k - 1}{r - 1} p^r q^{k-r}$$

   And

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

4.7 Poisson Random Variable

A random variable X is said to have a **Poisson** distribution with parameter λ if X takes values $0, 1, 2 \dots$ with probabilities given as:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

And

$$\mathrm{E}(X) = \lambda \qquad \mathrm{Var}(X) = \lambda$$

Poisson distribution of parameter $\lambda := np$ can be used as an approximation for a binomial distribution with parameter (n, p), when n is large and p is small such that np is moderate. (**Poisson Paradigm**) Consider n events, with p_i equal to the probability that event i occurs, $i = 1, \dots, n$. If all the p_i are small and trials are either independent or at most weakly dependent, then the number of these events that occur approximately has a Poisson distribution with mean $\sum_{i=1}^{n} p_i := \lambda$. Another use is Poisson process.

4.8 Hypergeometric Random Variable

Suppose that we have a set of N balls, of which m are red and $N - m$ is blue. We choose n of these balls, *without replacement*, and define X to be the number of red balls in our sample. Then

$$P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

for $x = 0, 1, \dots, N$. A random variable whose probability mass function is given as the above equation for some values of n, N, m is said to be a **hypergeometric** random variable. and is denoted by $\mathrm{H}(n, N, m)$. Here,

$$\mathrm{E}(X) = \frac{nm}{N}, \qquad \mathrm{Var}(X) = \frac{nm}{N}\left[\frac{(n-1)(m-1)}{(N-1)} + 1 - \frac{nm}{N}\right]$$

4.9 Expected Value of Sums of Random Variables

Theorem 4.4.

$$\mathrm{E}[X] = \sum_{s \in S} X(s)p(s)$$

Theorem 4.5. For random variables $X_1, X_2, \dots, X_n$,

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

4.10 Distribution Functions and Probability Mass Function

4.10.1 Properties of distribution function

1. F_X is a nondecreasing function.

2. $\lim_{b\to\infty} F_X(b) = 1$

3. $\lim_{b\to-\infty} F_X(b) = 0$

4. F_X is right continuous. That is for any $b \in \mathbb{R}$

$$\lim_{x\to b^+} F_X(x) = F_X(b)$$

4.10.2 Useful Calculations

1. Calculating probabilities from density function

   (a) $P(a < X \leq b) = F_X(b) - F_X(a)$

   (b) $P(X < b) = \lim_{n\to\infty} F(b - \frac{1}{n})$

   (c) $P(X = a) = F_X(a) - F_X(a^-)$ where $F_X(a^-) = \lim_{x\to a^-} F_X(x)$

2. Calculating probabilities from probability mass function

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

3. Calculate probability mass function from density function

$$p_X(x) = F_X(x) - F_X(x^-)$$

4. Calculate density function from probability mass function

$$F_X(x) = \sum_{y\leq x} p_X(y)$$

5 Continuous Random Variable

5.1 Introduction

Definition 5.1. We say that X is a **continuous** random variable if there exists a non-negative function f_X, defined for all real $x \in \mathbb{R}$, having the property that, for any set B of real numbers,

$$P(X \in B) = \int_B f_X(x)\mathrm{d}x$$

The function f_X is called the **probability density function** of the random variable X. For instance, letting $B = [a, b]$, we have

$$P(a \leq X \leq b) = \int_a^b f_X(x)\mathrm{d}x$$

Definition 5.2. We defined the **distribution function** of X by

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(x)\mathrm{d}x$$

and using the Fundamental Theorem of Calculus,

$$F'_X(x) = f_X(x)$$

Theorem 5.1 (Properties of Distribution Function)**.**

1. *$P(X = x) = 0 \forall x \in \mathbb{R}$.*
2. *F_X is continuous.*
3. *For any $a, b \in \mathbb{R}$, where $a < b$,*

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

5.2 Expectation and Variance of Continuous Random Variables

Definition 5.3. Let X be a continuous random variable with probability density function f_X, then

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x$$
$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x^2 - \mathrm{E}(X)^2) f_X(x)\mathrm{d}x$$

Theorem 5.2. If X is a continuous random variable with probability density function f_X, then for any real value function g

1. $$\mathrm{E}[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x)\mathrm{d}x$$

2. Same linearity Property:
$$\mathrm{E}(aX+b) = a\mathrm{E}(X) + b$$

3. Same alternative formula for variance
$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

Theorem 5.3 (Tail sum formula)**.**
Suppose X is a nonnegative continuous random variable, then
$$\mathrm{E}(X) = \int_0^\infty P(X > x)\mathrm{d}x = \int_0^\infty P(X \geq x)\mathrm{d}x$$

Theorem 5.4. We have $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

5.3 Uniform distribution

A random variable X is said to be **uniformly** distributed over the interval (a,b) if its probability density function is given by
$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

We denote this by $X \sim U(0,1)$.
Finding F_X:
$$F_X(x) = \int_{-\infty}^x f_X(y)\mathrm{d}y = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } 0 \leq x < 1 \\ 1, & \text{if } 1 \leq x \end{cases}$$

In general, for $a < b$, a random variable X is uniformly distributed over the interval (a,b) if its probability density function is given by
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

We denote this by $X \sim U(a,b)$. In a similar way,
$$F_X(x) = \int_{-\infty}^x f_X(y)\mathrm{d}y = \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \leq x < b \\ 1, & \text{if } b \leq x \end{cases}$$

It was shown that
$$\mathrm{E}(X) = \frac{a+b}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

5.4 Normal Distribution

A random variable is said to be **normally** distributed with parameters μ and σ^2 if its probability density function is given by

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

We denote this by $X \sim N(\mu, \sigma^2)$. The density function is **bell-shaped**, always positive, symmetric at μ and attains its maximum at $x = \mu$.
A normal random variable is called a **standard normal** random variable when $\mu = 0$ and $\sigma = 1$ and is denoted by $Z \sim N(0, 1)$. Its probability density function is denoted by ϕ and its distribution function by Φ.

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

An observation: Let $Y \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$, then

$$P(a < Y \leq b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - Phi\left(\frac{a-\mu}{\sigma}\right)$$

Theorem 5.5 (Properties of Standard Normal)**.** *1. $P(Z \geq 0) = P(Z \leq 0) = 0.5$.*

2. $-Z \sim N(0, 1)$

3. $P(Z \leq x) = 1 - P(Z > x)$

4. $P(Z \leq -x) = P(Z \geq x)$

5. If $Y \sim N(\mu, \sigma^2)$, then $X = \frac{Y-\mu}{\sigma} \sim N(0, 1)$

6. If $X \sim N(0, 1)$, then $Y = aX + b \sim N(b, a^2)$

Important facts:

1. If $Y \sim N(\mu, \sigma^2)$, then $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \sigma^2$.
2. If $Z \sim N(0, 1)$, then $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

Definition 5.4. The qth quantile of a random variable X is defined as a number z_q so that $P(X \leq z_q) = q$.

5.5 Exponential Distribution

A random variable X is said to be **exponentially** distributed with parameter $\lambda > 0$ if its probability density function is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

The distribution function of X is given by

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

Exponential distribution has memoryless property.

$$P(X > s + t \mid X > s) = P(X > t)$$

Mean and variance of $X \sim \mathrm{Exp}(\lambda)$:

$$\mathrm{E}(X) = \frac{1}{\lambda} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

5.6 Gamma Distribution

A random variable X is said to have a **gamma distribution** with parameters (α, λ), denoted by $X \sim \Gamma(\alpha, \lambda)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha - 1}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where $\lambda > 0$, $\alpha > 0$, and $\Gamma(\alpha)$, called the **gamma function**, is defined by

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha - 1} \mathrm{d}y$$

Remark:

1. $\Gamma(1) = 1$.
2. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
3. For integral values of $\alpha = n$, $\Gamma(n) = (n-1)!$.
4. $\Gamma(1, \lambda) = \mathrm{Exp}(\lambda)$.
5. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

5.7 Beta Distribution

A random variable X is said to have a **beta distribution** with parameter (a, b), denoted by $X \sim \text{Beta}(a, b)$, if its density is given by

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x$$

is known as the **beta function**.
It can be shown that

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

If $X \sim \beta(a, b)$, then

$$\mathrm{E}[X] = \frac{a}{a+b} \quad \text{and} \quad \mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

5.8 Cauchy Distribution

A random variable X is said to have a **Cauchy distribution** with parameter θ, $-\infty < \theta < \infty$, denoted by $X \sim \text{Cauchy}(\theta)$, if its density is given by

$$f(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, \quad -\infty < x < \infty$$

5.9 Approximation of Binomial Random Variables

Theorem 5.6 (De Moivre-Laplace Limit Theorem)**.**
Suppose that $X \sim \text{Bin}(n, p)$. Then for any $a < b$

$$P\left(a \leq \frac{X - np}{\sqrt{npq}} \leq b\right) \to \Phi(b) - \Phi(a)$$

as $n \to \infty$, where $q = 1 - p$.
That is,

$$\text{Bin}(n, p) \approx N(np, npq)$$

Equivalently,

$$\frac{X - np}{\sqrt{npq}} \approx Z$$

where $Z \sim N(0, 1)$. **Remark**: The normal approximation will be generally quite good for values of n satisfying $np(1-p) \geq 10$.

Approximation is further improved if we incorporate **continuity correction**.
If $X \sim \text{Bin}(n, p)$, then

$$
\begin{aligned}
P(X = k) &= P\left(k - \frac{1}{2} \leq X \leq k + \frac{1}{2}\right) \\
P(X \geq k) &= P\left(X \geq k - \frac{1}{2}\right) \\
P(X \leq k) &= P\left(X \leq k + \frac{1}{2}\right)
\end{aligned}
$$

5.10 Distribution of a Function of a Random Variable

Theorem 5.7. Let X be a continuous random variable having probability density function f_X. Suppose that $g(x)$ is a strictly monotonic(increasing or decreasing), differentiable (and thus continuous) function of x. Then the random variable Y defined by $Y = g(x)$ has a probability density function given by

$$
f_Y(y) = \begin{cases} f_X(g^{-1}(y))|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)|, & \text{if } y = g(x) \text{ for some } x \\ 0, & \text{if } y \neq g(x) \text{ for all } x \end{cases}
$$

6 Jointly Distributed Random Variables

6.1 Joint Distribution Functions

Definition 6.1. For any two random variables X and Y defined on the same sample space, we defined the **joint distribution function of** X and Y by

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) \quad \text{for} x, y \in \mathbb{R}$$

The distribution function of X can be obtained from the joint density function of X and Y in the following way:

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y)$$

We call F_X the **marginal distribution function** of X.
Similarly,

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x,y)$$

and F_Y is called the **marginal distribution function** of Y.

Theorem 6.1 (Some Useful Calculations)**.**
Let $a, b, a_1 \leq a_2, b_1 \leq b_2$ be real numbers, then

$$P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F_{X,Y}(a,b)$$

$$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F_{X,Y}(a_2,b_2) - F_{X,Y}(a_1,b_2) + F_{X,Y}(a_1,b_1) - F_{X,Y}(a_2,b_1)$$

6.1.1 Jointly Discrete Random Variables

In the case when both X and Y are discrete random variables, we define the **joint probability mass function of** X and Y as:

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

We can recover the probability mass function of X and Y in the following manner:

$$p_X(x) = P(X = x) = \sum_{y \in \mathbb{R}} p_{X,Y}(x,y)$$

$$p_Y(y) = P(Y = y) = \sum_{x \in \mathbb{R}} p_{X,Y}(x,y)$$

We call p_X the **marginal probability mass function** of X and p_Y the **marginal probability mass function** of Y.

Theorem 6.2 (Some useful formulas)**.**

1. $$P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = \sum_{a_1 < X \leq a_2} \sum_{b_1 < Y \leq b_2} p_{X,Y}(x,y)$$

2.

$$F_{X,Y}(a,b) = P(X \le a, Y \le b) = \sum_{X\le a}\sum_{Y\le b} p_{X,Y}(x,y)$$

3.

$$P(X > a, Y > b) = \sum_{X>a}\sum_{Y>b} p_{X,Y}(x,y)$$

6.1.2 Jointly Continuous Random Variables

We say that X and Y are **jointly continuous random variables** if there exists a function (which is denoted by $f_{X,Y}$, called the **jointly probability density function** of X and Y) if for every set $C \subset \mathbb{R}^2$, we have

$$P((X,Y) \in C) = \iint_{(x,y)\in C} f_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y$$

Theorem 6.3 (Some useful formulas)**.** *1. Let $A, B \subset \mathbb{R}^2$, take $C = A \times B$ above*

$$P(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x$$

2. In particular, let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ where $a_1 < a_2$ and $b_1 < b_2$, we have

$$P(a_1 < X \le a_2, b_1 < Y \le b_2) = \int_{a_1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x$$

3. Let $a, b \in \mathbb{R}$, we have

$$F_{X,Y}(a,b) = P(X \le a, Y \le b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x,y)\mathrm{d}y\mathrm{d}x$$

As a result of this,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

Definition 6.2. The marginal probability density function of X is given by

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

Similarly, the marginal probability density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\mathrm{d}x$$

6.2 Independent Random Variables

Two random variables X and Y are said to be **independent** if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \quad \text{for any } A, B \subset \mathbb{R}$$

Theorem 6.4 (For jointly discrete random variables)**.**
The following three statements are equivalent:

1. Random variables X and Y are indepedent.
2. For all $x, y \in \mathbb{R}$, we have
$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$
3. For all $x, y \in \mathbb{R}$, we have
$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Theorem 6.5. Random variables X and Y are independent if and only if there exist functions $g, h : \mathbb{R} \to \mathbb{R}$ such that for all $x, y \in \mathbb{R}$, we have

$$f_{X,Y}(x, y) = g(x)h(y)$$

6.3 Sums of Independent Random Variables

Under the assumption of independence of X and Y, we have

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Then it follows that

$$F_{X,Y}(x) = \int_{-\infty}^{\infty} F_X(x - t)f_Y(t)\mathrm{d}t$$

And

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(x - t)f_Y(t)\mathrm{d}t$$

Theorem 6.6 (Sum of 2 Independent Gamma Random Variables)**.**
Assume that $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$, and X and Y are mutually independent. Then,

$$X + Y \sim \Gamma(\alpha + \beta, \lambda)$$

Theorem 6.7 (Sum of Independent Exponential Random Variables)**.**
Let $X_1, X_2, \ldots, X_n$ be n independent exponential random variables each having parameter λ. Equivalently, $X_i \sim \mathrm{Exp}(\lambda) = \Gamma(1, \lambda)$. Then, $X_1 + X_2 + \cdots + X_n \sim \Gamma(n, \lambda)$.

Theorem 6.8 (Sum of Indepedent Normal Random Variables)**.**
If $X_i \sim N(\mu_i, \sigma_i^2), \forall i = 1, 2, \ldots, n$, then $\sum_{i=1}^{n} X_i \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$.

6.4 X and Y are discrete and independent

Theorem 6.9 (Sum of 2 Independent Poisson Random Variables)**.**
If $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are two independent random variables, $X + Y \sim \text{Poisson}(\lambda + \mu)$.

Theorem 6.10 (Sum of 2 Indepedent Binomial Random Variables)**.**
If $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$ are two independent random variables, $X + Y \sim \text{Bin}(n + m, p)$.

Theorem 6.11 (Sum of 2 Independent Geometric Random Variables)**.**
If $X \sim \text{Geom}(p)$ and $Y \sim \text{Geom}(p)$ are two independent random variables, $X+Y \sim \text{NB}(2, p)$.

6.5 Conditional distribution: Discrete Case

The **conditional probability mass function** of X given that $Y = y$ is defined by

$$\begin{aligned} p_{X|Y}(x \mid y) :&= P(X = x \mid Y = y) \\ &= \frac{p_{X,Y}(x, y)}{p_Y(y)} \end{aligned}$$

for all values of y such that $P_Y(y) > 0$.
Similarly, the **conditional distribution function** of X given that $Y = y$ is defined by

$$\begin{aligned} F_{X|Y}(x \mid y) :&= \frac{P(X \leq x, Y = y)}{P(Y = y)} \\ &= \sum_{a \leq x} p_{X|Y}(x \mid y) \end{aligned}$$

Theorem 6.12. If X is independent of Y, then the conditional probability mass function of X given $Y = y$ is the same as the marginal probability mass function of X for every y such that $p_Y(y) > 0$, i.e. $p_{X|Y}(x \mid y) = p_X(x)$.

6.6 Conditional distributions: Continuous Case

Suppose X and Y are jointly continuous random variables. Define the **conditional probability density function** of X given that $Y = y$ as

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

for all y such that $f_Y(y) > 0$. We define conditional probabilities of event associated with one random variable when we are given the value of a second random variable. That is, for $A \subset \mathbb{R}$ and y such that $f_Y(y) > 0$,

$$P(X \in A \mid Y = y) \int_A f_{X|Y}(x \mid y) \mathrm{d}x$$

In particular, the **conditional distribution function** of X given that $Y = y$ is defined by

$$F_{X|Y}(x, y) = P(X \leq x \mid Y = y) = \int_{-\infty}^{x} f_{X|Y}(t \mid y)\mathrm{d}t$$

Theorem 6.13. If X is independent of Y, then the conditional probability density function of X given $Y = y$ is the same as the marginal probability density function of X for every y such that $f_Y(y) > 0$, i.e.,

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

6.7 Joint Probability Distribution Function of Functions of Random Variables

Let X and Y be jointly distributed random variables with joint probability density function $f_{X,Y}$.
Suppose that

$$U = g(X, Y) \quad \text{and} \quad V = h(X, Y)$$

for some functions g and h.
The jointly probability density function of U and V is given by

$$f_{U,V}(u, v) = f_{X,Y}(x, y)|J(x, y)|^{-1}$$

where $x = a(u, v)$ and $y = b(u, v)$.
Here, g and h have continuous partial derivatives and

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

6.8 Jointly Distributed Random Variables: $n \geq 3$

Assume X, Y, Z are jointly continuous random variables, with

$$F_{X,Y,Z}(x, y, z) := P(X \leq x, Y \leq y, Z \leq z)$$

The marginal distribution functions are given as

$$\begin{aligned}
F_{X,Y}(x, y) &= \lim_{z \to \infty} F_{X,Y,Z}(x, y, z) \\
F_{X,Z}(x, z) &= \lim_{y \to \infty} F_{X,Y,Z}(x, y, z) \\
F_{Y,Z}(y, z) &= \lim_{x \to \infty} F_{X,Y,Z}(x, y, z) \\
F_X(x) &= \lim_{y,z \to \infty} F_{X,Y,Z}(x, y, z) \\
F_Y(y) &= \lim_{x,z \to \infty} F_{X,Y,Z}(x, y, z) \\
F_Z(z) &= \lim_{x,y \to \infty} F_{X,Y,Z}(x, y, z)
\end{aligned}$$

6.8.1 Joint probability density function of X, Y and Z:$f_{X,Y,Z}(x, y, z)$

For any $D \subset \mathbb{R}^3$, we have

$$P((X, Y, Z) \in D) = \iiint_{(x,y,z)\in D} f_{X,Y,Z}(x, y, z)\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

6.8.2 Marginal probability density function of X, Y and Z

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}y\mathrm{d}z \\
f_Y(y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}x\mathrm{d}z \\
f_Z(z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}x\mathrm{d}y \\
f_{X,Y}(x, y) &= \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}z \\
f_{Y,Z}(y, z) &= \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}x \\
f_{X,X}(x, z) &= \int_{-\infty}^{\infty} f_{X,Y,Z}(x, y, z)\mathrm{d}y
\end{aligned}$$

6.8.3 Independent random variables

Theorem 6.14. For jointly continuous random variables, the following three statements are equivalent:

1. Random variables X, Y and Z are independent
2. For all $x, y, z \in \mathbb{R}$, we have

$$f_{X,Y,Z}(x, y, z) = f_X(x)f_Y(y)f_Z(z)$$

3. For all $x, y, z \in \mathbb{R}$, we have

$$F_{X,Y,Z}(x, y, z) = F_X(x)F_Y(y)F_Z(z)$$

7 Properties of Expectation

Theorem 7.1. If $a \le X \le b$, then $a \le \mathrm{E}(X) \le b$.

7.1 Expectation of Sums of Random Variables

Theorem 7.2.

1. If X and Y are jointly discrete with joint probability mass function $p_{X,Y}$, then

$$\mathrm{E}[g(X,Y)] = \sum_y \sum_x g(x,y) p_{X,Y}(x,y)$$

2. If X and Y are joint continuous with joint probability density function $f_{X,Y}$, then

$$\mathrm{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) \mathrm{d}x \mathrm{d}y$$

Some important consequences of theorem above are:

1. If $g(x,y) \ge 0$ whenever $p_{X,Y}(x,y) > 0$, then $\mathrm{E}[g(X,Y)] \ge 0$.
2. $\mathrm{E}[g(X,Y) + h(X,Y)] = \mathrm{E}[g(X,Y)] + \mathrm{E}[h(X,Y)]$
3. $\mathrm{E}[g(X) + h(Y)] = \mathrm{E}[g(X)] + \mathrm{E}[h(Y)]$.
4. **Monotone Property**
   If jointly distributed random variables X and Y satisfy $X \le Y$, then

$$E(X) \le E(Y)$$

Theorem 7.3 (Boole's Inequality)**.**

$$P(\cup_{k=1}^{n} A_k) \le \sum_{k=1}^{n} P(A_k)$$

7.2 Covariance, Variance of Sums, Correlations

Definition 7.1. The **covariance** of jointly distributed random variables X and Y, denoted by $\mathrm{cov}(X,Y)$, is defined by

$$\mathrm{cov}(X,Y) = \mathrm{E}(X - \mu_X)(Y - \mu_Y)$$

where μ_X, μ_Y denote the means of X and Y respectively.
If $\mathrm{cov}(X,Y) \ne 0$, we say that X and Y are correlated.

Theorem 7.4 (Alternative formulae for covariance)**.**

$$\begin{aligned}\text{cov}(X,Y) &= \text{E}(XY) - \text{E}(X)\text{E}(Y)\\ &= \text{E}[X(Y-\mu_Y)]\\ &= \text{E}[Y(X-\mu_X)]\end{aligned}$$

Theorem 7.5. If X and Y are independent, then for any functions $g, h : \mathbb{R} \to mathbbR$, we have

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

Theorem 7.6. If X and Y are independent, then $\text{cov}(X,Y) = 0$.

Theorem 7.7 (Some properties of covariance)**.**

1. $\text{Var}(X) = \text{cov}(X,X)$
2. $\text{cov}(X,Y) = \text{cov}(Y,X)$
3. $\text{cov}\left(\sum_{i=1}^n a_iX_i, \sum_{j=1}^m b_jY_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_ib_j\text{cov}(X_i,Y_j)$

Theorem 7.8.

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k) + 2\sum_{1\le i<j\le n} \text{cov}(X_i,X_j)$$

If $X_1,\dots,X_n$ are independent random variables, then

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k)$$

In other words, ***under independence, variance of sum = sum of variances***

Definition 7.2 (Correlation Coefiicient)**.**
The correlation coefficient of random variables X and Y, denoted by $\rho(X,Y)$, is defined by

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Theorem 7.9.

$$-1 \le \rho(X,Y) \le 1$$

1. The correlation coefiicient is a measure of the degree of linearity between X and Y. If $\rho(X,Y) = 0$, then X and Y are said to be uncorrelated.
2. $\rho(X,Y) = 1$ if and only if $Y = aX + b$ where $a = \frac{\rho_Y}{\rho_X} > 0$.
3. $\rho(X,Y) = -1$ if and only if $Y = aX + b$ where $a = -\frac{\rho_Y}{\rho_X} < 0$.
4. $\rho(X,Y)$ is dimensionless.
5. If X and Y are independent, then $\rho(X,Y) = 0$.

7.3 Conditional expectation

Definition 7.3.

1. If X and Y are jointly distributed discrete random variables, then

$$\mathrm{E}[X \mid Y=y] = \sum_x x p_{X|Y}(x \mid y) \quad \text{if } p_Y(y) > 0$$

2. If X and Y are jointly distributed continuous random variables, then

$$\mathrm{E}[X \mid Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) \mathrm{d}x \quad \text{if } f_Y(y) > 0$$

Theorem 7.10 (Some important formulas)**.**

$$\mathrm{E}[g(X) \mid Y=y] = \begin{cases} \sum_x g(x) p_{X|Y}(x \mid y) & \text{for discrete case} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y) \mathrm{d}x & \text{for continuous case} \end{cases}$$

and hence

$$\mathrm{E}\left[\sum_{k=1}^{n} X_k \mid Y=y\right] = \sum_{k=1}^{n} \mathrm{E}[X_k \mid Y=y]$$

7.3.1 Computing expectation by conditioning

Theorem 7.11.

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X \mid Y]] = \begin{cases} \sum_y \mathrm{E}(X \mid Y=y) P(Y=y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathrm{E}(X \mid Y=y) f_Y(y) \mathrm{d}y & \text{if } Y \text{ is continuous} \end{cases}$$

7.3.2 Computing probabilities by conditioning

Theorem 7.12.
Let $X = I_A$ where A is an event. Then we have

$$\mathrm{E}[I_A] = P(A) \qquad \mathrm{E}[I_A \mid Y=y] = P(A \mid Y=y)$$

and we have

$$\begin{aligned} P(A) &= \mathrm{E}(I_A) = \mathrm{E}[\mathrm{E}(I_A \mid Y)] \\ &= \begin{cases} \sum_y \mathrm{E}(I_A \mid Y=y) P(Y=y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathrm{E}(I_A \mid Y=y) f_Y(y) \mathrm{d}y & \text{if } Y \text{ is continuous} \end{cases} \\ &= \begin{cases} \sum_y P(A \mid Y=y) P(Y=y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} P(A \mid Y=y) f_Y(y) \mathrm{d}y & \text{if } Y \text{ is continuous} \end{cases} \end{aligned}$$

7.4 Conditional Variance

Definition 7.4. The **conditional variance** of X given that $Y = y$ is defined as

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

Theorem 7.13.

$$\mathrm{Var}(X) = \mathrm{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathrm{E}[X \mid Y])$$

7.5 Moment Generating Functions

Definition 7.5. The **moment generating function** of random variable X, denoted by M_X, is defined as

$$\begin{aligned} M_X(t) &= \mathrm{E}[e^{tX}] \\ &= \begin{cases} \sum_x e^{tx} p_X(x), & \text{if } X \text{ is discrete with probability mass function } p_X \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)\mathrm{d}x & \text{if } X \text{ is continuous with probability density function } f_X \end{cases} \end{aligned}$$

Theorem 7.14 (Properties of Moment Generating Function)**.**

1. $M_X^{(n)}(0) = \mathrm{E}[X^n]$.
2. Multiplicative Property: If X and Y are independent, then
$$M_{X+Y}(t) = M_X(t)M_Y(t)$$
3. Uniqueness Property: Let X and Y be random variables with their moment generating functions M_X and M_Y respectively. Suppose that there exists an $h > 0$ such that
$$M_X(t) = M_Y(t), \quad \forall t \in (-h, h)$$
then X and Y have the same distribution.

Theorem 7.15 (Typical Moment Generating Functions)**.**

1. When $X \sim \mathrm{Be}(p)$, $M(t) = 1 - p + pe^t$.
2. When $X \sim \mathrm{Bin}(n, p)$, $M(t) = (1 - p + pe^t)^n$.
3. When $X \sim \mathrm{Geom}(p)$, $M(t) = \frac{pe^t}{1-(1-p)e^t}$.
4. When $X \sim \mathrm{Poisson}(\lambda)$, $M(t) = e^{\lambda e^t - 1}$.
5. When $X \sim U(\alpha, \beta)$, $M(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta - \alpha)t}$.
6. When $X \sim \mathrm{Exp}(\lambda)$, $M(t) = \frac{\lambda}{\lambda - t}$ for $t < \lambda$.
7. When $X \sim N(\mu, \sigma^2)$, $M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.

7.6 Joint Moment Generating Functions

Definition 7.6. For any n random variables $X_1, \ldots, X_n$, the joint moment generating function, $M(t_1, \ldots, t_n)$, is defined for all real values $t_1, \ldots, t_n$ by

$$M(t_1, \ldots, t_n) = \mathrm{E}[e^{t_1 X_1 + \cdots + t_n X_n}]$$

The individual moment generating functions can be obtained from $M(t_1, \ldots, t_n)$ by letting all but one of the t_j be 0. That is,

$$M_{X_i}(t) = \mathrm{E}[e^{tX_i}] = M(0, \ldots, 0, t, 0, \ldots, 0)$$

where the t is in the ith place.
It can be proved that $M(t_1, \ldots, t_n$ uniquely determines the joint distribution of $X_1, \ldots, X_n$.
n random variable $X_1, \ldots, X_n$ are independent if and only if

$$M(t_1, \ldots, t_n) = M_{X_1}(t_1) \cdots M_{X_n}(t_n)$$

8 Limit Theorems

8.1 Chebyshev's Inequality and the Weak Law of Large Numbers

Theorem 8.1 (Markov's Inequality)**.**
Let X be a nonnegative random variable. For $a > 0$, we have

$$P(X \geq a) \leq \frac{\mathrm{E}(X)}{a}$$

Theorem 8.2 (Chebyshev's Inequality)**.**
Let X be a random variable with finite mean μ and variance σ^2, then for $a > 0$, we have

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

Theorem 8.3 (Consequences of Chebyshev's Inequality)**.**
If $\mathrm{Var}(X) = 0$, then the random variable X is a constant. Or in other words,

$$P(X = \mathrm{E}(X)) = 1$$

Theorem 8.4 (The Weak Law of Large Numbers)**.**
Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, with common mean μ. Then, for any $\epsilon > 0$,

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \text{ as } n \to \infty$$

8.2 Central Limit Theorem

Theorem 8.5 (Central Limit Theorem)**.**
Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having mean μ and variance σ^2. Then the distribution of

$$\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

tends to the standard normal distribution as $n \to \infty$. That is,

$$\lim_{n\to\infty} P\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{2\pi}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\mathrm{d}t$$

8.3 The Strong Law of Large Numbers

Theorem 8.6 (The Strong Law of Large Numbers)**.**
Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having a finite mean $\mu = \mathrm{E}(X_i)$. Then with probability 1,

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \text{ as } n \to \infty$$

In other words,

$$P\left(\left\{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right\}\right) = 1$$

9 Problems

1 (AY1314Sem1) Let X_1 and X_2 have a bivariate normal distribution with parameters $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\rho = \frac{1}{2}$. Find the probability that all of the roots of the following equation are real:

$$X_1 x^2 + 2X_2 x + X_1 = 0$$

2 (AY1617Sem1) Let (X_1, X_2) have a bivariate normal distribution with means 0, variances μ_1^2 and μ_2^2, respectively, and with the correlation coefficient $-1 < \rho < 1$.

(a) Determine the distribution of $aX_1 + bX_2$, where a and b are two real numbers such that $a^2 + b^2 > 0$.

(b) Find a constants b such that $X_1 + bX_2$ is independent of X_1. Justify your answer.

(c) Find the probability that the following equation has real roots:

$$X_1 x^2 - 2X_1 x - bX - 2 = 0$$

where b is the constant found in part (ii).