

Probability Theory

Labix

May 8, 2025

Abstract

Notes for the basics of Probability Theory.

Contents

1	Foundations of Probability Theory	3
1.1	Definition of Probability	3
1.2	Multiplication Principle	4
1.3	Conditional Probability	5
1.4	Independence of Events	6
2	Probability Distributions	7
2.1	Random Variables and its Distribution	7
2.2	Cumulative Density Functions	10
2.3	Multivariate Random Variables	10
2.4	Algebra of Random Variables	11
3	Expectation and Variance	15
3.1	Expectations	15
3.2	Variance and Covariance	15
3.3	Moments	17
3.4	Conditional Expectations	18
4	Convergence of Random Variables	20
4.1	Convergence	20
4.2	Standardized Random Variables	21
5	Introduction to Stochastic Processes	22
5.1	Basic Terminology	22
5.2	Hitting Time and Stopping Time	22
6	Markov Chains	24
6.1	Markov Chains	24
6.2	Communicating Classes	26
6.3	Recurrence and Transience	26
6.4	Branching Process	28
6.5	Invariant Distributions on a Markov Chain	28
7	Brownian Motion	31
8	Martingale Processes	32
8.1	Basic Definitions	32
8.2	Martingale Transformations	32
9	Stochastic Calculus	33
9.1	Ito's Integral	33
9.2	The Integral as a Martingale	34

9.3	Ito's Lemma	34
9.4	Localization	34
10	Stochastic Differential Equations	35
10.1	Basic Definitions	35
10.2	Geometric Brownian Motions	35
10.3	Systems of SDEs	35
11	Mathematical Finance	36
12	Mathematical Models for Option Pricing	37
12.1	Background Terminology	37
12.2	One-Period Binomial Models	37
12.3	Multi-Period Binomial Models	38
12.4	Black-Scholes-Merton Model	39

1 Foundations of Probability Theory

1.1 Definition of Probability

Definition 1.1.1: Probability Space

A probability space is a measure space (Ω, \mathcal{F}, P) where the measure P lands in $[0, 1]$.

Explicitly, a probability space is a triple (Ω, \mathcal{F}, P) consisting of the following data:

- $\Omega \neq \emptyset$ is a set called the sample space.
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a σ -algebra called events.
- $P : \mathcal{F} \rightarrow [0, 1]$ is a set function.

such that the following are true:

- $P(\Omega) = 1$.
- If $\{A_n \mid n \in \mathbb{N}\} \subseteq \mathcal{F}$ are pairwise disjoint, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

Proposition 1.1.2

Let (Ω, \mathcal{F}, P) be a probability space. Let $A, B \in \mathcal{F}$ be events. Then the following are true.

- $P(\Omega \setminus A) = 1 - P(A)$
- $A \subset B \implies P(A) \leq P(B)$

Proof. Let $A \subset B \subset \Omega$ be events in Ω .

- A and $\Omega \setminus A$ are disjoint and $P(\Omega) = P(A) + P(\Omega \setminus A)$ and $P(\Omega \setminus A) = 1 - P(A)$
- We have that A and $B \setminus A$ are disjoint. Thus $P(B) = P(A) + P(B \setminus A)$. Since $P(B \setminus A) \geq 0$, we have $P(A) \leq P(B)$.

□

Definition 1.1.3: Uniform Probability Measure

Let Ω be a sample space. A probability measure P is uniform if for all $a, b \in \Omega$,

$$P(\{a\}) = P(\{b\})$$

Theorem 1.1.4

Let Ω be a sample space and P a uniform probability measure of Ω . Then for all $A \subset \Omega$,

$$P(A) = \frac{|A|}{|\Omega|}$$

Proof. Suppose that A consists of $|A|$ distinct elements and the event space $|\Omega|$ contains $|\Omega|$ distinct elements. Since every singleton set is pairwise disjoint, we have $P(A) = |A|P(\{a\})$ for any $a \in A$. Similarly, we have $P(\Omega) = |\Omega|P(\{a\})$. Thus we have that $P(A) = \frac{|A|P(\Omega)}{|\Omega|}$ and $P(A) = \frac{|A|}{|\Omega|}$ □

Theorem 1.1.5: Principle of Inclusion Exclusion

Let $A, B \subset \Omega$ be a sample space and P the probability measure.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof. Note that

$$\begin{aligned} A \cup (B \setminus A) &= A \cup (B \cap A^c) \\ &= (A \cup B) \cap (A \cup A^c) \\ &= A \cup B \end{aligned}$$

Note also that $A \cap (B \setminus A) = \emptyset$. Thus $P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B)$ \square

Theorem 1.1.6: Extended Principle of Inclusion Exclusion

Let $A_k \subset \Omega$ be a sample space and P the probability measure for all $k \leq n \in \mathbb{N}$. Then

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 \leq \dots \leq k} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k})$$

1.2 Multiplication Principle**Theorem 1.2.1: The Multiplication Principle**

Suppose that Experiment A has a outcomes and Experiment B has b outcomes. Then the performing both A and B results in ab possible outcomes.

Theorem 1.2.2: Sampling with replacement - Ordered

In the case of sampling k balls with replacement from an urn containing n balls, there are $|\Omega| = n^k$ possible outcomes when the order of the objects matters, where $\Omega = \{(s_1, \dots, s_k) : s_i \in \{1, \dots, n\} \forall i \in \{1, \dots, k\}\}$.

Theorem 1.2.3: Sampling without replacement - Ordered

In the case of sampling k balls without replacement from an urn containing n balls, there are $|\Omega| = \frac{n!}{(n-k)!}$ possible outcomes when the order of the objects matters, where $\Omega = \{(s_1, \dots, s_k) : s_i \in \{1, \dots, n\} \forall i \in \{1, \dots, k\}, i \neq j \implies s_i \neq s_j\}$.

Theorem 1.2.4: Sampling without replacement - Unordered

In the case of sampling k balls without replacement from an urn containing n balls, there are $|\Omega| = \binom{n}{k}$ possible outcomes when the order of the objects does not matter, where $\Omega = \{\omega \subset \{1, \dots, n\} : |\omega| = k\}$.

Theorem 1.2.5: Sampling with replacement - Unordered

In the case of sampling k balls with replacement from an urn containing n balls, there are $|\Omega| = \binom{n+k-1}{k}$ possible outcomes when the order of the objects does not matter, where $\Omega = \{\omega \subset \{1, \dots, n\} : \omega \text{ is a } k \text{ element multiset with elements from } \{1, \dots, n\}\}$.

1.3 Conditional Probability

Definition 1.3.1: Conditional Probability

Consider a probability space (Ω, P) . Let $A, B \subset \Omega$ with $P(B) > 0$. Then the conditional probability of A given B , denoted by $P(A|B)$ is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Theorem 1.3.2: Multiplication Rule

Let $n \in \mathbb{N}$. Then for any events A_1, \dots, A_n such that $P(A_2 \cap \dots \cap A_n) > 0$, we have

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

Proof. From the right hand side, we have

$$\begin{aligned} & P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= P(A_1) \frac{P(A_2 \cap A_1)}{P(A_1)} \frac{P(A_3 \cap A_2 \cap A_1)}{P(A_2 \cap A_1)} \dots \frac{P(A_n \cap \dots \cap A_1)}{P(A_1 \cap \dots \cap A_{n-1})} \\ &= P(A_1 \cap \dots \cap A_n) \end{aligned}$$

□

Theorem 1.3.3: Bayes' Rule

Let (Ω, P) be a probability measure. Let $A, B \subset \Omega$ with $P(A), P(B) > 0$. Then

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

Proof. We have that $P(A \cap B) = P(A|B)P(B)$ and $P(A \cap B) = P(B|A)P(A)$.

□

Theorem 1.3.4: Law of Total Probability

Let (Ω, P) be a probability measure. Let A_1, \dots, A_n be a partition of Ω with $P(A_i) > 0$ for all $i = 1, \dots, n$. Then for any $B \subset \Omega$,

$$P(B) = \sum_{k=1}^n P(A_k)P(B|A_k)$$

Proof. Note that since A_1, \dots, A_n is a partition, $B \cap A_1, \dots, B \cap A_n$ is also a partition.

$$\begin{aligned} \sum_{k=1}^n P(A_k)P(B|A_k) &= \sum_{k=1}^n P(B \cap A_k) \\ &= P\left(\bigcup_{k=1}^n B \cap A_k\right) \\ &= P(B \cap \Omega) \\ &= P(B) \end{aligned}$$

□

Theorem 1.3.5: General Bayes' Rule

Let (Ω, P) be a probability measure. Let A_1, \dots, A_n be a partition of Ω with $P(A_i) > 0$ for all $i = 1, \dots, n$. Then for any $B \subset \Omega$ with $P(B) > 0$,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^n P(B|A_k)P(A_k)}$$

Proof. Apply Bayes' rule and apply the multiplication rule. □

1.4 Independence of Events**Definition 1.4.1: Independent Events**

Two events A, B are said to be independent if

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

Proposition 1.4.2

If A, B are independent, then A^c, B, A, B^c and A^c, B^c are independent.

Proof. We only proof the first and third item.

- Without loss of generality we prove the first and reader mirrors the second.

$$\begin{aligned} P(A^c \cap B) &= P(B) - P(A \cap B) \\ &= P(B)(1 - P(A)) \\ &= P(B)P(A^c) \end{aligned}$$

- Note that $P(A \cap B) = P(A)P(B)$

$$\begin{aligned} P(A^c \cap B^c) &= 1 - P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c) \end{aligned}$$

□

2 Probability Distributions

2.1 Random Variables and its Distribution

Definition 2.1.1: Random Variable

Let (Ω, \mathcal{F}, P) be a probability space. Let (E, \mathcal{E}) be a measurable space. An (E, \mathcal{E}) valued random variable is an \mathcal{F} -measurable function $X : \Omega \rightarrow E$.

Definition 2.1.2: Independent Random Variables

Let (Ω, \mathcal{F}, P) be a probability space. Let (E, \mathcal{E}) be a measurable space. Let $X, Y : \Omega \rightarrow E$ be random variables. We say that X and Y are independent if for any $A, B \in \mathcal{E}$, we have that $X^{-1}(A)$ and $Y^{-1}(B)$ are independent events in \mathcal{F} .

Definition 2.1.3: Discrete and Continuous Random Variables

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable.

- We say that X is discrete if $\text{im}(X)$ is a countable subset of \mathbb{R} .
- We say that X is continuous otherwise.

Recall that X is an \mathcal{F} -measurable function if $X^{-1}(B) \in \mathcal{F}$ for $B \in \mathcal{E}$.

Definition 2.1.4: Probability Distribution

Let (Ω, E, \mathbb{P}) be a probability space. Let (E, \mathcal{E}) be a measurable space. Let $X : \Omega \rightarrow E$ be a measurable function. Define the probability distribution of X to be the pushforward measure $P \circ X^{-1} = P_X : \mathcal{E} \rightarrow [0, 1]$ defined by

$$P_X(A) = P(X^{-1}(A))$$

for $A \in \mathcal{E}$.

Definition 2.1.5: Probability Density Function

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Define the probability density function of X to be the Radon–Nikodym derivative

$$f_X = \frac{dX_*P}{d\mu}$$

where μ is the Lebesgue measure.

Recall that this means that f_X satisfies the property that

$$P_X(A) = \int_A f_X d\mu$$

for any measurable set $A \subseteq \mathbb{R}$. In particular, if $A = \{a\} \subseteq \mathcal{F}$, then we have

$$P_X(a) = f_X(a)$$

The probability distribution function has its input as every measurable subset of \mathbb{R} , while the probability density function takes input as individual points of \mathbb{R} . They are really the same thing because having its probability be determined on singletons is sufficient to determine the probability of every measurable subset.

Proposition 2.1.6

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a discrete random variable. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a function. Then the probability density function of $Y = g \circ X$ is given by

$$f_Y(y) = \sum_{x \in g^{-1}(y)} f_X(x)$$

Proposition 2.1.7

Suppose that X is a continuous random variable with density f_X and $g : \mathbb{R} \rightarrow \mathbb{R}$ is strictly monotone and differentiable with inverse function denoted g^{-1} , then $Y = g(X)$ has density

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy}(g^{-1}(y)) \right|$$

for all $y \in \mathbb{R}$

Example 2.1.8: Bernoulli Distribution

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. We say that X has a Bernoulli distribution if the probability density function of X is given by

$$f_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

for some $p \in [0, 1]$.

Example 2.1.9: Binomial Distribution

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. We say that X has a binomial distribution if the probability density function of X is given by

$$f_X(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for some $p \in [0, 1]$.

Definition 2.1.10: Poisson Distribution

A discrete random variable X is said to have Poisson Distribution with parameter $\lambda > 0$ if $\text{im}(X) = \mathbb{N}_0$ and

$$p_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Definition 2.1.11: Geometric Distribution

A discrete random variable X is said to have Geometric Distribution with parameter $p \in (0, 1)$ if $\text{im}(X) = \mathbb{N}_0$ and

$$p_X(x) = p(1 - p)^{x-1}$$

Let $I \subseteq \mathbb{R}$ be an interval. Recall that $\mathcal{B}(I)$ refers to the borel measurable subsets of I . Denote λ the Lebesgue measure on \mathbb{R}^n .

Example 2.1.12: Uniform Distribution

Let $[a, b] \subseteq \mathbb{R}$ be an interval. Let X be a random variable on the probability space $([a, b], \mathcal{B}([a, b]), P)$. We say that X has a uniform distribution if its probability density function is given by

$$f_X(A) = \frac{\lambda(A)}{b-a}$$

for $A \subseteq [a, b]$.

In particular, when $A = \{c\} \subseteq [a, b]$ is the one-point set, we have $P_X(c) = \frac{1}{b-a}$ so that the probability of any one point set is uniform.

Example 2.1.13

Let X be a uniform distribution on $[a, b]$. Then the probability density function of X is given by

$$f_X(x) = \frac{1}{b-a}$$

Example 2.1.14: Normal Distribution

Let X be a random variable on the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$. We say that X has a normal distribution if its probability density function is given by

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

for some $\mu \in \mathbb{R}$ and $\sigma > 0$.

Definition 2.1.15: Exponential Distribution

A continuous random variable X is said to have Exponential Distribution with parameter $\lambda > 0$ if its density function is given by

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

and its cumulative function given by

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

Definition 2.1.16: Gamma Distribution

A continuous random variable X is said to have Gamma Distribution with shape parameter $\alpha > 0$ and rate parameter $\beta > 0$ if its density function is given by

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

2.2 Cumulative Density Functions

Definition 2.2.1: Cumulative Distribution Function

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Define the cumulative distribution function $F_X : \mathbb{R} \rightarrow \mathbb{R}$ of X to be

$$F_X(x) = P_X(X \leq x)$$

Proposition 2.2.2

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Then the following are true.

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

Proposition 2.2.3

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Then the following are true regarding the cumulative distribution function F_X .

- F_X is monotonically increasing: $x \leq y \implies F_X(x) \leq F_X(y)$
- F_X is right continuous: If (x_n) is a sequence such that $x_1 \geq \dots \geq x_n \geq x_{n+1} \geq \dots \geq x$ and $(x_n) \rightarrow x$, then $F_X(x_n) \rightarrow F_X(x)$
- $F_X(-\infty) = 0$ and $F_X(\infty) = 1$

Proposition 2.2.4

Suppose that X is a random variable on a probability space $(\Omega, \mathcal{E}, \mathbb{P})$ with cumulative distribution function F_X . If $a < b$, then $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$

2.3 Multivariate Random Variables

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space. The definition of random variables and probability distribution is well-adapted to the case when the random variable X lands in \mathbb{R}^n . In this case, we may find the relationship between the probability density function of X and the probability density function of its individual components.

Definition 2.3.1: Joint Probability Mass Function

Let X, Y be discrete random variables. The joint probability mass function of X and Y is the function

$$p_{X,Y}(x, y) = P(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}) = P((X, Y) = (x, y))$$

for all $(x, y) \in \mathbb{R}^2$

Theorem 2.3.2

Let $p_{X,Y}$ be the joint probability mass function of two random variables X, Y .

- $p_X(x) = \sum_y p_{X,Y}(x, y)$
- $p_Y(y) = \sum_x p_{X,Y}(x, y)$

Definition 2.3.3: Joint Cumulative Distribution Function

Let X, Y be random variables. The joint cumulative distribution function of X and Y is the function

$$F_{X,Y}(x, y) = P(\{\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y\}) = P(X \leq x, Y \leq y)$$

for all $(x, y) \in \mathbb{R}^2$

Theorem 2.3.4

Let $F_{X,Y}$ be the joint cumulative distribution function of two random variables X, Y .

- $\lim_{x,y \rightarrow -\infty} F_{X,Y}(x, y) = 0$
- $\lim_{x,y \rightarrow \infty} F_{X,Y}(x, y) = 1$
- $x \leq x'$ and $y \leq y'$ implies $F_{X,Y}(x, y) \leq F_{X,Y}(x', y')$
- $F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$
- $F_Y(y) = \lim_{x \rightarrow \infty} F_{X,Y}(x, y)$

Definition 2.3.5: Jointly Continuous

Let X, Y be random variables. X and Y are jointly continuous if

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du$$

for a function $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$ satisfying

- $f_{X,Y}(u, v) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v) dv du = 1$

We call $f_{X,Y}$ the joint density function of (X, Y) .

Theorem 2.3.6

Let $F_{X,Y}$ be the joint cumulative distribution function of two random variables X, Y .

- $f_{X,Y}(x, y) = \begin{cases} \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y) & \text{if the derivative exists at } (x, y) \\ 0 & \text{otherwise} \end{cases}$
- $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$
- $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$

Proposition 2.3.7: L

t (Ω, E, \mathbb{P}) be a probability space. Let (E, \mathcal{E}) be a measurable space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be a random variables. Then the following are equivalent.

- X and Y are independent.
- $f_{(X,Y)} = f_X f_Y$.
- $F_{(X,Y)} = F_X F_Y$

2.4 Algebra of Random Variables

Proposition 2.4.1

Let (Ω, E, \mathbb{P}) be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be a random variables. Then we have

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{(X,Y)}(t, z-t) dt$$

Proposition 2.4.2

Let $X \approx \text{Poi}(\lambda)$ and $Y \approx \text{Poi}(\mu)$ be independent. $X + Y \approx \text{Poi}(\lambda + \mu)$.

Proof.

$$\begin{aligned}
 p_{X+Y}(m) &= \sum_{k \in \mathbb{Z}} \frac{\lambda^k}{k!} e^{-k} \frac{\mu^{m-k}}{(m-k)!} e^{k-m} \\
 &= \frac{1}{m!} e^{-m} \sum_{k=0}^m m! \frac{\lambda^k}{k!} \frac{\mu^{m-k}}{(m-k)!} \\
 &= \frac{1}{m!} e^{-m} \sum_{k=0}^m \binom{m}{k} \lambda^k \mu^{m-k} \\
 &= \frac{(\lambda + \mu)^m}{m!} e^{-m}
 \end{aligned}$$

□

Proposition 2.4.3

Let $X_1, \dots, X_n \approx \text{Bern}(p)$ be independent. $X_1 + \dots + X_n \approx \text{Bin}(n, p)$.

Proof. We prove by induction. When $n = 2$,

$$\begin{aligned}
 p_{X_1+X_2}(0) &= p_{X_1}(0)p_{X_2}(0) \\
 &= 1 - 2p + p^2 \\
 p_{X_1+X_2}(1) &= p_{X_1}(0)p_{X_2}(1) + p_{X_1}(1)p_{X_2}(0) \\
 &= (1-p)(p) + p(1-p) \\
 &= 2p(1-p) \\
 p_{X_1+X_2}(2) &= p_{X_1}(1)p_{X_2}(1) + p_{X_1}(2)p_{X_2}(0) \\
 &= p^2 \\
 p_{\text{Bin}(2,p)}(x) &= \binom{2}{x} p^x (1-p)^{2-x}
 \end{aligned}$$

For $x \in \{0, 1, 2\}$, the two probability density functions match thus for the case $n = 2$, it is true. Now suppose that $X_1 + \dots + X_{n-1} \approx \text{Bin}(n-1, p)$. Let $Y = \text{Bin}(n-1, p) + X_n$. For $m \in \{0, \dots, n\}$,

$$\begin{aligned}
 p_Y(m) &= \sum_{k \in \mathbb{Z}} p_{\text{Bin}(n-1,p)}(k) p_{X_n}(m-k) \\
 &= \sum_{k=0}^m p_{\text{Bin}(n-1,p)}(k) p_{X_n}(m-k) \\
 &= \sum_{k=0}^m \binom{n-1}{k} p^k (1-p)^{n-1-k} p_{X_n}(m-k) \\
 &= \sum_{k=m-1}^m \binom{n-1}{k} p^k (1-p)^{n-1-k} p_{X_n}(m-k) \\
 &= \binom{n-1}{m-1} p^{m-1} (1-p)^{n-m} p_{X_n}(1) + \binom{n-1}{m} p^m (1-p)^{n-1-m} p_{X_n}(0) \\
 &= \binom{n-1}{m-1} p^{m-1} (1-p)^{n-m} + \binom{n-1}{m} p^m (1-p)^{n-m} \\
 &= \binom{n}{m} p^m (1-p)^{n-m}
 \end{aligned}$$

Thus for the case $X_1 + \dots + X_n$ it is true.

□

Proposition 2.4.4

Let $X \approx \text{Bin}(m, p)$ and $Y \approx \text{Bin}(n, p)$ be independent. $X + Y \approx \text{Bin}(m + n, p)$.

Proof.

$$\begin{aligned}
 p_{X+Y}(t) &= \sum_{k \in \mathbb{Z}} p_X(k) p_Y(t-k) \\
 &= \sum_{k=0}^t \binom{m}{k} p^k (1-p)^{m-k} \binom{n}{t-k} p^{t-k} (1-p)^{n-t+k} \\
 &= \sum_{k=0}^t \binom{m}{k} \binom{n}{t-k} p^t (1-p)^{m+n-t} \\
 &= p^t (1-p)^{m+n-t} \sum_{k=0}^t \frac{m!}{k!(m-k)!} \frac{n!}{(t-k)!(n-t+k)!}
 \end{aligned}$$

□

Proposition 2.4.5

Let $\lambda > 0$. Let $n \in \mathbb{N}$. Let T_1, \dots, T_n be independent random variables with exponential distribution parameter λ . Then

$$Z = \sum_{k=1}^n T_k \approx \text{Gamma}(n, \lambda)$$

Proof. We prove by induction. When $n = 2$,

$$\begin{aligned}
 f_Z(z) &= \int_{-\infty}^{\infty} f_{T_1}(x) f_{T_2}(z-x) dx \\
 &= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx \\
 &= \lambda^2 e^{-\lambda z} \int_0^z dx \\
 &= \lambda^2 z e^{-\lambda z}
 \end{aligned}$$

Thus the case $n = 2$ is true. Suppose that it is true for $n = k - 1$. Let $X \approx \text{Gamma}(n - 1, \lambda)$.

$$\begin{aligned}
 f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_{T_n}(z-x) dx \\
 &= \int_0^z \frac{\lambda^{n-1}}{\Gamma(n-1)} x^{n-2} e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx \\
 &= \frac{\lambda^n}{\Gamma(n-1)} e^{-\lambda z} \int_0^z x^{n-2} dx \\
 &= \frac{\lambda^n}{\Gamma(n-1)} e^{-\lambda z} \frac{1}{n-1} z^{n-1} \\
 &= \frac{\lambda^n}{\Gamma(n)} z^{n-1} e^{-\lambda z}
 \end{aligned}$$

Thus we are done

□

Proposition 2.4.6

Let $m, n \in \mathbb{N}$ and $\lambda > 0$. Let $X \approx \text{Gamma}(m, \lambda)$ and $Y \approx \text{Gamma}(n, \lambda)$ be independent. $X + Y \approx \text{Gamma}(m + n, \lambda)$.

Proof.

$$\begin{aligned}
 f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \\
 &= \int_0^z \frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x} \frac{\lambda^n}{\Gamma(n)} (z-x)^{n-1} e^{-\lambda(z-x)} dx \\
 &= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} e^{-\lambda z} \int_0^z x^{m-1} (z-x)^{n-1} dx \\
 &= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} e^{-\lambda z} \int_0^z x^{m-1} \sum_{k=0}^{n-1} \binom{n-1}{k} z^{n-1-k} (-x)^k dx \\
 &= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} e^{-\lambda z} \sum_{k=0}^{n-1} \binom{n-1}{k} z^{n-1-k} (-1)^k \int_0^z x^{m-1+k} dx \\
 &= \frac{\lambda^{m+n}}{\Gamma(m)\Gamma(n)} z^{m+n-1} e^{-\lambda z} \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^k \frac{1}{m+k}
 \end{aligned}$$

□

Theorem 2.4.7

Suppose that T_1, T_2, \dots are independent random variables with exponential distribution parameter λ . Define for $t \geq 0$,

$$N_t = \begin{cases} 0 & \text{if } T_1 > t \\ 1 & \text{if } T_1 \leq t < T_1 + T_2 \\ 2 & \text{if } T_1 + T_2 \leq t < T_1 + T_2 + T_3 \\ \dots & \end{cases}$$

Then, for any $t \geq 0$, we have that $N_t \approx \text{Poi}(\lambda t)$.

Definition 2.4.8: Poisson Process

The family of random variables $\{N_t : t \geq 0\}$ is said to be Poisson process of intensity λ if

- $N_0 = 0$
- for any t_0, \dots, t_n with $0 = t_0 < t_1 < t_2 < \dots < t_n$, the random variables $N_{t_1}, N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \dots, N_{t_n} - N_{t_{n-1}}$ are independent, and $N_{t_i} - N_{t_{i-1}} \approx \text{Poi}(\lambda(t_i - t_{i-1}))$

3 Expectation and Variance

3.1 Expectations

Definition 3.1.1: Expectations

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Define the expectation of X to be

$$E[X] = \int_{\Omega} X dP$$

Lemma 3.1.2

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Then we have

$$E[X] = \int_{\mathbb{R}} x f_X(x) dx$$

Proposition 3.1.3: Law of the Unconscious Staticians

Let (Ω, \mathcal{F}, P) be a probability space. Let $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ be random variables. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a function. Then we have

$$E[g \circ (X_1, \dots, X_n)] = \int_{\mathbb{R}^n} g(x_1, \dots, x_n) f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \cdots dx_n$$

Proposition 3.1.4

Let (Ω, \mathcal{F}, P) be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be random variables. Then the following are true.

- If X, Y are random variables and $a, b \in \mathbb{R}$, then

$$E[aX + bY] = aE[X] + bE[Y]$$

- If $P(X \geq Y) = 1$, then

$$E[X] \geq E[Y]$$

Proposition 3.1.5

Let (Ω, \mathcal{F}, P) be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be random variables. Then X, Y are independent if and only if

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

for any two functions $g, h : \mathbb{R} \rightarrow \mathbb{R}$.

3.2 Variance and Covariance

Definition 3.2.1: Variance

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Define the variance of X to be

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

Lemma 3.2.2

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Then the following are true.

- $\text{Var}(X) \geq 0$.
- $\text{Var}(X) = 0$ if and only if $P_X(E[X]) = 1$.
- $\text{Var}(X) = E[X^2] - E[X]^2$
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$ for any $a, b \in \mathbb{R}$.

Proposition 3.2.3

Suppose that X_1, \dots, X_n are independent variables with finite variance. Then

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

Definition 3.2.4: Covariance

Let X, Y be two random variables. The covariance of X, Y is defined as

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

Proposition 3.2.5

Suppose that X, Y are random variables.

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$
- If X, Y are independent, $\text{Cov}(X, Y) = 0$
- $\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$

Proposition 3.2.6: Variance of Sums

For random variables X_1, \dots, X_n , we have

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j)$$

Theorem 3.2.7

Given two random variables X and Y , we have

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)}$$

Theorem 3.2.8: Correlation Coefficient

The correlation coefficient between two random variables X and Y is given by

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

Proposition 3.2.9

Let X and Y be random variables. We have

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

Moreover, for any $a, b, c, d \in \mathbb{R}$ with $a, c > 0$, we have

$$\rho(aX + b, cY + d) = \rho(X, Y)$$

Proposition 3.2.10

Let X, Y be random variables.

- $\rho(X, X) = 1$
- $\rho(X, -X) = -1$
- X, Y are uncorrelated if $\rho(X, Y) = 0$

3.3 Moments

Definition 3.3.1: k th Moment

Let X be a random variable. For $k \in \mathbb{N}$ we define the k th moment of X as $E[X^k]$ whenever the expectation exists.

Definition 3.3.2: Moment Generating Function

The moment-generating function of a random variable X is the function M_X defined as

$$M_X(t) = E[e^{tX}]$$

for all $t \in \mathbb{R}$ for which the expectation is well defined.

Theorem 3.3.3

Assume that M_X exists in a neighbourhood of 0, that is, there exists $\epsilon > 0$ such that for all $t \in (-\epsilon, \epsilon)$ we have $M_X(t) < \infty$. Then for all $k \in \mathbb{N}$ the k th moment of X exists, and

$$E[X^k] = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0}$$

Proof. We have that $E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x) dx$ for any continuous cumulative probability. On the other hand,

$$\begin{aligned} \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} &= \left. \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \right|_{t=0} \\ &= \left. \int_{-\infty}^{\infty} \frac{\partial^k}{\partial t^k} e^{tx} f_X(x) dx \right|_{t=0} \\ &= \left. \int_{-\infty}^{\infty} x^k e^{tx} f_X(x) dx \right|_{t=0} \\ &= \int_{-\infty}^{\infty} x^k f_X(x) dx \end{aligned}$$

□

Proposition 3.3.4

Assume that all expectations in the statement are well defined.

- For any $a, b \in \mathbb{R}$, $M_{aX+b}(t) = e^{tb} M_X(at)$
- If X, Y are independent, then $M_{X+Y}(t) = M_X(t) M_Y(t)$

Theorem 3.3.5

Let X, Y be two random variables. Assume that the moment generating functions of X, Y exists and are finite on an interval of the form $(-\epsilon, \epsilon)$. Assume further that $M_X(t) = M_Y(t)$ for all $t \in (-\epsilon, \epsilon)$. Then X, Y have the same distribution.

Theorem 3.3.6

Let X be a non-negative random variable whose expectation is well defined. We then have

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

Theorem 3.3.7

Let X be a random variable whose variance is well defined. Then

$$P(|X - E(X)| \geq x) \leq \frac{\text{Var}(X)}{x^2}$$

for all $x > 0$

3.4 Conditional Expectations**Definition 3.4.1: Conditional Expectations on Subalgebras**

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Let \mathcal{H} be a σ -subalgebra of \mathcal{F} . Define $E[X | \mathcal{H}] : \Omega \rightarrow \mathbb{R}$ to be a random variable such that the following are true.

- $E[X | \mathcal{H}]$ is \mathcal{H} -measurable.
- For any $A \in \mathcal{H}$, we have $E[X \cdot 1_A] = E[E[X | \mathcal{H}] \cdot 1_A]$

Lemma 3.4.2

Let (Ω, \mathcal{F}, P) be a probability space. Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable. Let \mathcal{H} be a σ -subalgebra of \mathcal{F} . Then the random variable $E[X | \mathcal{H}]$ exists and is unique up to almost surely equality.

Lemma 3.4.3

Let (Ω, \mathcal{F}, P) be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be random variables. Let \mathcal{H} be a σ -subalgebra of \mathcal{F} . Then the following are true.

- Stability: If X is \mathcal{H} -measurable, then $E[XY | \mathcal{H}] = XE[Y | \mathcal{H}]$.
- Independence: If $\sigma(X)$ and \mathcal{H} are independent, then $E[X | \mathcal{H}] = E[X]$.

Definition 3.4.4: Conditional Expectation on Random Variables

Let (Ω, \mathcal{F}, P) be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be random variables. Define the conditional expectation of X on Y to be

$$E[X | Y] = E[X | \sigma(Y)]$$

Definition 3.4.5: Conditional Density

Let (Ω, \mathcal{F}, P) be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be random variables. Define the conditional density of X on the event $\{\omega \in \Omega \mid Y(\omega) = y\}$ by

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

Lemma 3.4.6

Let (Ω, \mathcal{F}, P) be a probability space. Let $X, Y : \Omega \rightarrow \mathbb{R}$ be random variables. Then we have

$$E[X \mid Y](\omega) = E[X \mid Y = Y(\omega)] = \int_{-\infty}^{\infty} x f_{X \mid Y}(x, Y(\omega)) \, dx$$

4 Convergence of Random Variables

4.1 Convergence

Definition 4.1.1: Convergence in Mean Square

We say that a sequence of random variables X_1, X_2, \dots converges in mean square to a random variable X if

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0$$

Definition 4.1.2: Convergence in Probability

We say that a sequence of random variables X_1, X_2, \dots converges in probability to a random variable X if for every $\epsilon > 0$, we have

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

Theorem 4.1.3

Let X_1, X_2, \dots be a sequence of random variables, and X another random variable. If $X_n \rightarrow X$ in mean square as $n \rightarrow \infty$ then $X_n \rightarrow X$ in probability as $n \rightarrow \infty$.

Definition 4.1.4: Convergence in Distribution

We say that a sequence of random variables X_1, X_2, \dots converges in distribution to a random variable X if

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

for every x in the set $C = \{x \in \mathbb{R} : F_X \text{ is continuous at } x\}$.

Theorem 4.1.5

For any random variable X , the set of discontinuity points of F_X is countable.

Theorem 4.1.6

Let X_1, X_2, \dots be a sequence of random variables, and X another random variable. If $X_n \rightarrow X$ in probability, then $X_n \rightarrow X$ in distribution.

Theorem 4.1.7

Let X_1, X_2, \dots be a sequence of random variables such that $X_n \rightarrow c$ in distribution, where $c \in \mathbb{R}$, then X_n converges in probability to c .

Theorem 4.1.8: Law of large numbers in mean square

Let X_1, X_2, \dots be a sequence of independent random variable, each with mean μ and variance σ^2 . Then

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} \rightarrow \mu$$

in mean square.

Theorem 4.1.9: Weak law of large numbers

Let X_1, X_2, \dots be a sequence of independent random variable, each with mean μ and variance $\sigma^2 \neq 0$. Then

$$\lim_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{n} \rightarrow \mu$$

in probability.

4.2 Standardized Random Variables**Definition 4.2.1: Standardized Random Variables**

Let X be a random variable with finite variance. We define the standardized version of X to be the random variable Z given by

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}}$$

Theorem 4.2.2: Central Limit Theorem

Let X_1, X_2, \dots be a sequence of independent and identically distributed random variables, each with mean μ and variance $\sigma^2 \neq 0$. Let $S_n = X_1 + \dots + X_n$. Then the standardized version of S_n ,

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

converges in distribution as $n \rightarrow \infty$ to a Gaussian random variable with mean 0 and variance 1. That is,

$$\lim_{n \rightarrow \infty} P(Z_n \leq x) = \lim_{n \rightarrow \infty} F_{Z_n}(x) = F_Y(y) = \int_{-\infty}^x -\frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

5 Introduction to Stochastic Processes

5.1 Basic Terminology

Definition 5.1.1: Stochastic Processes

Let (Ω, \mathcal{F}, P) be a probability space. Let (S, Σ) be a measurable space. A stochastic process is a collection of random variables

$$\{X(t) : \Omega \rightarrow S \mid t \in T\}$$

indexed by some set T . In this case we write the stochastic process as $X : \Omega \times T \rightarrow S$.

Definition 5.1.2: Terminology

Let (Ω, \mathcal{F}, P) be a probability space. Let (S, Σ) be a measurable space. Let $X : \Omega \times T \rightarrow S$ be a stochastic process.

- Define the state space of the stochastic process to be S .
- Write $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$.

Definition 5.1.3: Discrete Time Stochastic Processes

Let (Ω, \mathcal{F}, P) be a probability space. Let (S, Σ) be a measurable space. Let $X : \Omega \times T \rightarrow S$ be a stochastic process. We say that the stochastic process has discrete time if $T \subseteq \mathbb{N}$.

Definition 5.1.4: Filtrations

Let $\mathcal{F} = \{\mathcal{F}_t \mid t \in T\}$ be a family of σ -algebras. We say that \mathcal{F} is a filtration if $\mathcal{F}_t \leq \mathcal{F}_s$ for $t \leq s$.

Definition 5.1.5: Adapted Stochastic Process

Let (Ω, \mathcal{F}, P) be a probability space. Let (S, Σ) be a measurable space. Let $X : \Omega \times T \rightarrow S$ be a stochastic process. Let $\mathcal{F} = \{\mathcal{F}_t \mid t \in T\}$ be a filtration. We say that X is adapted to \mathcal{F} if X_t is \mathcal{F}_t -measurable for all $t \in T$.

5.2 Hitting Time and Stopping Time

Definition 5.2.1: Hitting Time

Let (Ω, \mathcal{F}, P) be a probability space. Let (S, Σ) be a measurable space. Let $X : \Omega \times T \rightarrow S$ be a stochastic process. Let $A \in \Sigma$ be a measurable set. Define the hitting time of X to be the random variable $H_A : \Omega \rightarrow T$ given by

$$H_A(\omega) = \inf\{t \in T \mid X(\omega, t) \in A\}$$

Recall that for a σ -algebra \mathcal{F} and a subset $Y \subseteq \mathcal{F}$, $\sigma(Y)$ is the smallest σ -algebra that contains Y .

Definition 5.2.2: Stopping Time

Let (Ω, \mathcal{F}, P) be a probability space. Let (S, Σ) be a measurable space. Let $X : \Omega \times T \rightarrow S$ be a stochastic process. Let $S : \Omega \rightarrow T$ be a random variable. We say that S is a stopping time if S has the property that

$$\{\omega \in \Omega \mid S(\omega) \leq t\} \subseteq \sigma(X_t)$$

for all $t \in T$.

In other words, events determined by S up to some time $t \in T$ is dictated entirely using the variables X_0, \dots, X_t .

Lemma 5.2.3

Let (Ω, \mathcal{F}, P) be a probability space. Let (S, Σ) be a measurable space. Let $X : \Omega \times T \rightarrow S$ be a stochastic process. Let $S : \Omega \rightarrow T$ be a random variable. Then S is a stopping time if and only if S has the property that

$$\{\omega \in \Omega \mid S(\omega) = t\} \subseteq \sigma(X_t)$$

for all $t \in T$.

Theorem 5.2.4: Début Theorem

6 Markov Chains

6.1 Markov Chains

Definition 6.1.1: Stochastic Matrix

Let $P = (p_{i,j}) \in M_n(\mathbb{R})$ be a matrix. We say that P is a stochastic matrix if the following are true.

- $0 \leq p_{i,j} \leq 1$ for $1 \leq i, j \leq n$.
- For any fixed $1 \leq k \leq n$, we have

$$\sum_{j=1}^n p_{k,j} = 1$$

Definition 6.1.2: Markov Chain

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $I = \{1, \dots, k\} \subseteq \mathbb{N}$. Let $\{X_n : \Omega \rightarrow I \mid n \in \mathbb{N}\}$ be a discrete stochastic process. Let $\lambda : I \rightarrow [0, 1]$ be a probability distribution. Let $P = (p_{i,j})$ be a stochastic matrix. We say that $(X_n)_{n \geq 0}$ is a Markov chain with initial distribution λ and transition matrix P if the following are true.

- $P_{X_0} = \lambda$
- For any $i_0, \dots, i_{n+1} \in I$, we have

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_k = i_k \text{ for } 0 \leq k \leq n) = \lambda(i_0) \cdot \prod_{j=0}^n p_{i_j, i_{j+1}}$$

In this case we say that $(X_n)_{n \geq 0}$ is Markov(λ, P).

In other words, Markov chains are a sequence of random variables where the next step does not depend on what states you visited previously, but only on the state now. There is an important property that makes studying Markov chains worth while.

Proposition 6.1.3

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $I = \{1, \dots, k\}$. Let $\{X_n : \Omega \rightarrow I \mid n \in \mathbb{N}\}$ be a sequence of random variables. Let $\lambda : I \rightarrow [0, 1]$ be a probability distribution. Then $(X_n)_{n \geq 0}$ is a Markov chain if and only if

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_k = i_k \text{ for } 0 \leq k \leq n) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

for any $i_0, \dots, i_{n+1} \in I$. In this case, the transition matrix of the Markov chain is given by

$$P = \begin{pmatrix} \mathbb{P}(X_1 = 1 \mid X_0 = 1) & \cdots & \mathbb{P}(X_1 = k \mid X_0 = 1) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(X_1 = 1 \mid X_0 = k) & \cdots & \mathbb{P}(X_1 = k \mid X_0 = k) \end{pmatrix}$$

Theorem 6.1.4: Markov Property

Let $(X_n)_{n \geq 0}$ be a Markov chain. Suppose that $X_m = E_m$ is given. Then $(X_{m+n})_{n \geq 0}$ is a Markov chain.

Notice that the transition probability $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ still depends on i, j, n . We further restrict our study of Markov chains to the following type.

Definition 6.1.5: Homogenous Markov Chains

We say that a Markov chain $(X_n)_{n \geq 0}$ is homogenous if

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i)$$

This means that homogenous Markov chains are time independent.

Lemma 6.1.6

The transition matrix P of a homogenous Markov chain is a Stochastic matrix.

Theorem 6.1.7

Let $(X_n)_{n \geq 0}$ be Markov. Then for all $m, n \geq 0$, we have

- $\mathbb{P}(X_n = j | X_0 = i) = \mathbb{P}(X_1 = j | X_0 = i)^n$
- $\mathbb{P}(X_n = j) = \sum_{i \in I} \mathbb{P}(X_0 = i) \mathbb{P}(X_n = j | X_0 = i)$

This theorem is saying that $\mathbb{P}(X_n = j | X_0 = i)$ is equal to the i, j th entry of the matrix P^n if P is the transition matrix. To find out the total probability of reaching the j th state at the n th step regardless of the starting point, you sum up the $\mathbb{P}(X_n = j | X_0 = i)$ multiplying the probability that you start at state i .

We often like to create new Markov chains from old (at least from exams). If we want to show that a sequence $(X_n)_{n \geq 0}$ of random variables is Markov, try and write $X_{n+1} = X_n + \text{some term only related to } n + 1$.

We use

$$h_i^A = \mathbb{P}(H^A < \infty | X_0 = i)$$

to denote the corresponding probability. We use

$$k_i^A = E_i[H_A] = \sum_{k=0}^{\infty} k \mathbb{P}(H_A = k | X_0 = i) + \infty \mathbb{P}(H_A = \infty | X_0 = i)$$

to denote the expectation.

Proposition 6.1.8

We can find the probabilities of all hitting times by solving the system of linear equations

$$\begin{cases} \mathbb{P}(H_A < \infty | X_0 = i) = 1 & i \in A \\ \mathbb{P}(H_A < \infty | X_0 = i) = \sum_{j \in I} \mathbb{P}(X_1 = j | X_0 = i) \mathbb{P}(H_A < \infty | X_0 = j) & i \notin A \end{cases}$$

Proposition 6.1.9

We can find the expected number of hitting times by solving the system of linear equations

$$\begin{cases} E_i[H_A] = 0 & i \in A \\ E_i[H_A] = 1 + \sum_{j \notin A} \mathbb{P}(X_1 = j | X_0 = i) E_j[H_A] & i \notin A \end{cases}$$

Theorem 6.1.10: Strong Markov Property

Let $(X_n)_{n \geq 0}$ be Markov(λ, P) and T a stopping time of $(X_n)_{n \geq 0}$. Then conditional on both $\{X_T = i\}$ and $\{T < \infty\}$, we have $(X_{T+n})_{n \geq 0}$ is Markov(δ_i, P) and independent of X_0, \dots, X_T .

6.2 Communicating Classes

Definition 6.2.1: Talks and Communicates

Let i, j be states. We say that i talks to j which is $i \rightarrow j$ if there exists $n \in \mathbb{N}$ such that

$$\mathbb{P}(X_n = j | X_0 = i) > 0$$

We say that two states i and j communicate if $i \leftrightarrow j$.

Definition 6.2.2: Communicating Class

Let $C \subseteq I$. C is a communicating class if $\forall i, j \in C, i \leftrightarrow j$ and $\forall i \in C$ and $\forall k \in I \setminus C, i \not\rightarrow k$

Definition 6.2.3: Closed and Absorbing

We say that a class is closed if no states in the class talks to any states outside of that class. We say that a class is absorbing if it forms a closed class by itself.

Definition 6.2.4: Irreducible Markov Chains

A Markov chain is said to be irreducible if for all $i, j \in I, i \leftrightarrow j$.

6.3 Recurrence and Transience

Definition 6.3.1: Recurrence and Transience

A state $i \in I$ is recurrent if $\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1$. A state $i \in I$ is transient if $\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 0$.

Definition 6.3.2: k -th Passage Time

The first passage time of state i is a stopping time such that

$$T_i(\omega) = \inf\{n \geq 1 | X_n(\omega) = i\}$$

The k -th passage time is a stopping time such that

$$T_i^{(k)}(\omega) = \inf\{n \geq T_i^{(k-1)} | X_n(\omega) = i\}$$

The k -th excursion time is defined by

$$S_i^{(k)} = \begin{cases} T_i^{(k)} - T_i^{(k-1)} & \text{if } T_i^{(k-1)} < \infty \\ \infty & \text{otherwise} \end{cases}$$

Intuitively, $T_i^{(k)}$ outputs the time it takes for the Markov chain to reach state i for the k th time and $S_i^{(k)}$ outputs the number of steps taken between consecutive visits.

Lemma 6.3.3

For $k = 2, 3, \dots$ and if $T_i^{(k-1)} < \infty$, then $S_i^{(k)}$ is independent of $\{X_m : m \leq T_i^{(k-1)}\}$ and

$$\mathbb{P}(S_i^{(k)} = n | T_i^{(k-1)} < \infty) = \mathbb{P}_i(T_i = n)$$

Definition 6.3.4: Visit Counting Function

Let $i \in I$ be a state. Define

$$V_i = \sum_{k=0}^{\infty} 1_{\{X_k=i\}}$$

the number of visits ever to state i .

Lemma 6.3.5

If V_i counts the number of visits to state i , then

$$E_i(V_i) = \sum_{k=0}^{\infty} P_{ii}^{(k)}$$

Lemma 6.3.6

For $k = 0, 1, 2, \dots$, we have

$$\mathbb{P}_i(V_i > k) = (\mathbb{P}_i(T_i < \infty))^k$$

This lemma means that V_i asserts a geometric distribution.

Theorem 6.3.7: Characteristic of Recurrent and Transient States

Let $(X_n)_{n \geq 0}$ be a Markov chain. Let T_i be the first passage time of state i . If $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent and $\sum_{k=1}^{\infty} \mathbb{P}(X_1 = k | X_0 = k)^n = \infty$. If $\mathbb{P}_i(T_i < \infty) < 1$, then i is transient and $\sum_{k=1}^{\infty} \mathbb{P}(X_1 = k | X_0 = k)^n < \infty$.

Theorem 6.3.8

Let C be a communicating class. Then either all states in C are transient or all are recurrent.

Theorem 6.3.9

Let $C \subset I$ be a class of a Markov chain. Then

- Every recurrent class is closed
- Every finite closed class is recurrent

Theorem 6.3.10

If P is irreducible and recurrent then for all $j \in I$,

$$\mathbb{P}(T_j < \infty) = 1$$

Definition 6.3.11: Component Independent Simple Random Walk on \mathbb{Z}^d

A component independent simple random walk on \mathbb{Z}^d for $d \in \mathbb{N}$ is defined as

$$X_{n+1} = X_n + Z_{n+1}$$

where $Z_{n+1} = (Z_{n+1}^1, \dots, Z_{n+1}^d)$ with

$$Z_m^j = \begin{cases} 1 & \text{with probability } p_j \\ -1 & \text{with probability } q_j \end{cases}$$

where Z_m^j are independent for all j, m .

Proposition 6.3.12

Let $(X_n)_{n \geq 0}$ be a CISRW on \mathbb{Z}^d . If there exists $j \in \{1, \dots, d\}$ such that $p_j \neq \frac{1}{2}$ then (X_n) is transient.

Proposition 6.3.13

Let $(X_n)_{n \geq 0}$ be a CISRW on \mathbb{Z}^d and $p_j = q_j = \frac{1}{2}$ for all j . Then

- If $d \leq 2$ then all states are recurrent
- If $d \geq 3$ then all states are transient

6.4 Branching Process**Definition 6.4.1: Branching Process**

Let $(X_n)_{n \geq 0}$ be the number of individuals in a population at time n where $X_n \in \{0\} \cup \mathbb{N}$. At every time step each individual in the population gives birth to a random number of off-spring. Thus X_{n+1} is defined to be

$$X_{n+1} = \sum_{k=1}^{X_n} Z_k^n$$

where $Z_k^n \sim Z$ and $\mathbb{P}(Z \geq 0) = 1$. Then $(X_n)_{n \geq 0}$ is said to be a branching process.

Proposition 6.4.2

Define $G(s) = E[s^Z]$ and $F_{n+1}(s) = E[S^{X_{n+1}} | X_0 = 1]$ for $s \in (0, 1)$. Then

$$F_n(s) = G(F_{n-1}(s))$$

Proposition 6.4.3

For a braching process $(X_n)_{n \geq 0}$ with off-spring distribution Z where $E[Z] = \mu$, we have

$$E[X_n] = \mu^n$$

Theorem 6.4.4: Extinction Probability

The extinction probability is the smallest non-negative root of $G(\alpha) = \alpha$.

Theorem 6.4.5

Suppose that $G(0) > 0$, $\mu = E[Z] = G'(1)$. Then $\mu \leq 1$ implies certain extinction. $\mu > 1$ implies uncertain extinction.

6.5 Invariant Distributions on a Markov Chain**Definition 6.5.1: Invariant Distribution**

A measure $\lambda = (\lambda_i : i \in I)$ with non-negative entries is said to be an invariant distribution if

- $\lambda P = \lambda$
- $\pi_i \in [0, 1]$ for all i
- $\sum_{i \in I} \pi_i = 1$

The following theorem shows that after applying a Markov process to an invariant distribution, the new distribution will be the same as the old one. Notice here that right multiplication of a distribution gives the next step on the Markov chain instead of the usual left multiplication.

Theorem 6.5.2

Let $(X_n)_{n \geq 0}$ be Markov(π, P), where π is an invariant distribution for P . Then $\forall m \geq 0$, $(X_{n+m})_{n \geq 0}$ is Markov(π, P).

Proof. By the Markov property, the new sequence will be Markov. But we do not yet know its distribution. We have that

$$\begin{aligned} \mathbb{P}(X_m = i) &= (\pi P^m)_i && \text{(this denotes the } i\text{th entry of } P) \\ &= (\pi P P^{m-1})_i \\ &= (\pi P^{m-1})_i \end{aligned}$$

Thus by induction, we see that $\mathbb{P}(X_m = i) = \pi_i$ thus we are done. \square

Theorem 6.5.3

Let I be finite. Suppose that for some $i \in I$,

$$P_{ij}^n \rightarrow \pi_j$$

for all $j \in I$. Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

Let us look at an example.

Example 6.5.4: Gene Mutation

Recall the gene mutation example that has transition matrix $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$. Recall that

$$P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n \\ ? & ? \end{pmatrix}$$

As $n \rightarrow \infty$, observe that the first entry tends to $\frac{\beta}{\alpha+\beta}$ the second entry tends to $\frac{\alpha}{\alpha+\beta}$. This means that $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ is an invariant distribution for the gene mutation model.

We can also find the invariant distribution directly: Let $\pi = (\pi_1, \pi_2)$ be an invariant distribution for P . We have that

$$\begin{aligned} \pi P &= \pi \\ \left(\pi_1(1-\alpha) + \pi_2\beta \quad \pi_1\alpha + \pi_2(1-\beta) \right) &= \left(\pi_1 \quad \pi_2 \right) \\ \left(\pi_2\beta - \pi_1\alpha \quad \pi_1\alpha - \pi_2\beta \right) &= 0 \end{aligned}$$

Notice that these two equations mean the same thing. (In general if you have an $n \times n$ transition matrix, only the first $n-1$ equations will be useful and the last one will be redundant.) This is why we need to use the fact that $\pi_1 + \pi_2 = 1$ as our final equation.

Now solving it, we have that $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ which is the exact same answer as the one given in the first method.

A third way to think about it is that notice that $P^T \pi^T = \pi^T$ is the equation that we are attempting to solve. Recall that 1 is necessarily an eigenvector of a transition matrix which

means that this equation really is just a question of finding eigenvectors from the eigenvalue 1.

The above examples gives a lot of practical information when calculating the invariant distribution of a transition matrix.

Definition 6.5.5: Expected Time Spent

Define the expected time spent in state i in between visits to state k by

$$\gamma_i^k = E_k \left(\sum_{n=0}^{T_k-1} \mathcal{X}_{X_n=i} \right)$$

This uses k as a reference, and as we keep going back to k , this records how much time we have spent in i in the process of leaving and returning to k .

Theorem 6.5.6

Let P be irreducible and recurrent. Then

- $\gamma_k^k = 1$
- $\gamma^k = (\gamma_i^k : i \in I)$ is an invariant measure satisfying $\gamma^k P = \gamma^k$
- $0 < \gamma_i^k < \infty \forall i \in I$

Theorem 6.5.7

Let P be irreducible and λ an invariant measure and $\lambda_k = 1$ for some k . Then $\lambda \geq \gamma^k$. If we also have that P is recurrent, then $\lambda = \gamma^k$

This theorem asserts that as long as P is irreducible and recurrent, then any invariant measure is exactly equal to the γ^k we defined, which means that there is really only one invariant measure up to rescaling.

Definition 6.5.8: Positive Recurrent

A state $i \in I$ is positive recurrent if it is recurrent and

$$E_i[T_i] < \infty$$

A state which is not positive recurrent but is recurrent is called null recurrent.

Theorem 6.5.9

Let P be irreducible. Then the following are equivalent.

- Every state is positive recurrent
- Some state i is positive recurrent
- P has an invariant distribution π . More over, $\pi_i = \frac{1}{E_i[T_i]}$

7 Brownian Motion

Definition 7.0.1: Brownian Motion

Let (Ω, E, P) be a probability space. Let (S, Σ) be a measurable space. Let $B : \Omega \times [0, T] \rightarrow S$ be continuous stochastic processes. We say that B is a Brownian motion if the following are true.

- $B_0 = 0$.
- For any $0 \leq t_1 < t_2 < t_3 \leq T$,

$$B_{t_2} - B_{t_1} \quad \text{and} \quad B_{t_3} - B_{t_2}$$

are independent variables.

- For any $0 \leq t_1 < t_2 \leq T$,

$$B_{t_2} - B_{t_1} \sim N(0, t_2 - t_1)$$

- For any $\omega \in \tilde{\Omega}$ such that $P(\tilde{\Omega}) = 1$, we have that $B(\omega, -) : T \rightarrow S$ is a continuous function.

8 Martingale Processes

8.1 Basic Definitions

Definition 8.1.1: Martingale Process

Let (Ω, \mathcal{E}, P) be a probability space. Let (S, Σ) be a measurable space. Let $M : \Omega \times [0, T] \rightarrow S$ be a continuous stochastic process adapted to a filtration $\mathcal{F} = \{\mathcal{F}_t \mid t \in [0, T]\}$. We say that M is Martingale with respect to \mathcal{F} if the following are true.

- $E[|M_k|] < \infty$ for all k .
- $E[M_{t_2} \mid \mathcal{F}_{t_1}] = M_{t_1}$ for all $0 \leq t_1 < t_2 \leq T$.

Definition 8.1.2: Continuous Martingale Process

Let (Ω, \mathcal{E}, P) be a probability space. Let (S, Σ) be a measurable space. Let $M : \Omega \times [0, T] \rightarrow S$ be a continuous stochastic process. We say that M is continuous if for all $\omega \in \tilde{\Omega}$ such that $P(\tilde{\Omega}) = 1$, the function

$$M(\omega, -) : [0, T] \rightarrow S$$

is continuous.

Definition 8.1.3: Standard Brownian Filtration

8.2 Martingale Transformations

9 Stochastic Calculus

9.1 Ito's Integral

Given a Martingale process $M : \Omega \times [0, T] \rightarrow \mathbb{R}$ on some filtered probability space, we may ask to integrate M over the product measure of the measurable space $\Omega \times [0, T]$ because M is in particular a measurable function. Ito's integral gives a construction of such an integral under certain assumption on M .

Definition 9.1.1: Simple Functions

Let (Ω, \mathcal{F}, P) be a probability space. Define $\mathcal{H}_0^2 \subseteq L^2(\Omega \times [0, T])$ to be the vector subspace whose measurable functions are of the form

$$f(\omega, t) = \sum_{i=0}^{n-1} a_i(\omega) 1_{(t_i, t_{i+1}]}(t)$$

Definition 9.1.2: Ito Integral of Simple Functions

Let (Ω, \mathcal{F}, P) be a probability space. Let $f \in \mathcal{H}_0^2$. Define the Ito integral of $f(\omega, t) = \sum_{i=0}^{n-1} a_i(\omega) 1_{(t_i, t_{i+1}]}(t)$ by the formula

$$I(f)(\omega) = \sum_{i=0}^{n-1} a_i(\omega) (B_{t_{i+1}} - B_{t_i})$$

In particular, I is a function $I : \mathcal{H}_0^2 \rightarrow L^2(\Omega)$.

Lemma 9.1.3: Ito's Isometry I

Let (Ω, \mathcal{F}, P) be a probability space. The map $I : \mathcal{H}_0^2 \rightarrow L^2(\Omega)$ is a linear isometry.

Definition 9.1.4: Ito Integrable Functions

Let (Ω, \mathcal{F}, P) be a probability space. Define the space of Ito integrable functions $\mathcal{H}^2 \subseteq L^2(\Omega \times [0, T])$ to be the vector subspace whose measurable functions $f \in L^2(\Omega \times [0, T])$ satisfies the fact that

$$E \left[\int_{[0, T]} f^2(\omega, t) dt \right] = \int_{\Omega} \int_{[0, T]} f^2(\omega, t) dt dP < \infty$$

Lemma 9.1.5

Let (Ω, \mathcal{F}, P) be a probability space. Then \mathcal{H}_0^2 is dense in \mathcal{H}^2 .

Definition 9.1.6: Ito Integral of Integrable Functions

Let (Ω, \mathcal{F}, P) be a probability space. Let $f \in \mathcal{H}^2$. Let $(f_n)_{n \in \mathbb{N} \setminus \{0\}}$ be a sequence such that $f_n \rightarrow f$ in $L^2(\Omega \times [0, T])$. Then define the Ito integral of f by

$$I(f) = \lim_{n \rightarrow \infty} I(f_n)$$

In particular, I is a function $I : \mathcal{H}^2 \rightarrow L^2(\Omega)$.

Lemma 9.1.7: Ito's Isometry II

Let (Ω, \mathcal{F}, P) be a probability space. The map $I : \mathcal{H}^2 \rightarrow L^2(\Omega)$ is a linear isometry.

9.2 The Integral as a Martingale**9.3 Ito's Lemma****9.4 Localization**

10 Stochastic Differential Equations

10.1 Basic Definitions

Definition 10.1.1: Stochastic Differential Equations

Methods of solving differential equations can also be applied to stochastic differential equations.

10.2 Geometric Brownian Motions

Definition 10.2.1: Geometric Brownian Motion

Let (Ω, E, P) be a probability space. Let $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ be a continuous Martingale process with respect to the standard Brownian filtration $\mathcal{B} = \{B_t \mid t \in [0, T]\}$. We say that S follows a geometric Brownian motion if it satisfies the following stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

where μ is called the percentage drift and σ is called the percentage volatility.

Recall that dS_t really means the Radon–Nikodym derivative $\frac{dS_t}{d\mu}$ where μ here (not the same one as above) refers to the standard measure on \mathbb{R} , while dt means the genuine infinitesimally small increment in \mathbb{R} .

Lemma 10.2.2

Let (Ω, E, P) be a probability space. Let $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ be a continuous Martingale process with respect to the standard Brownian filtration $\mathcal{B} = \{B_t \mid t \in [0, T]\}$. Suppose that S follows a geometric Brownian motion. Then the solution of the stochastic differential equation is given by

$$S_t = S_0 e^{(\mu - \sigma^2/2)t + \sigma B_t}$$

Proposition 10.2.3

Let (Ω, E, P) be a probability space. Let $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ be a continuous Martingale process with respect to the standard Brownian filtration $\mathcal{B} = \{B_t \mid t \in [0, T]\}$. Suppose that S follows a geometric Brownian motion. Then the following are true.

- For each t , S_t has a log normal distribution.
- For each t , we have

$$E[S_t] = S_0 e^{\mu t}$$

- For each t , we have

$$\text{Var}(S_t) = S_0^2 e^{2\mu t} (e^{\sigma^2 t} - 1)$$

We also provide a discrete approach to solving the differential equation for easy simulation using the Monte Carlo method.

Proposition 10.2.4

The discretization of the defining stochastic differential equation of a geometric Brownian motion is given by

$$S_{t+dt} - S_t = \mu S_t dt + \sigma S_t N \sqrt{dt}$$

where dt denotes a small change in time, and $N \sim N(0, 1)$ refers a random variable with a normal distribution.

10.3 Systems of SDEs

11 Mathematical Finance

Assets: Items Stocks: Makes you own a small percentage of a company. Shares: The unit for stocks. (Owning share = Owning a certain number of stocks) Bonds: Buying debt from a loan issuer. Options: Contract that gives the ability to buy / sell assets at a certain time in the future. Futures: Contract to buy / sell assets at a certain time in the future. Value: Total gain / loss from the whole ordeal.

Derivatives: Just think of them as options / future for now. Strike price: the cost of exercising the derivative at the expiration date.

Idea: Derivatives themselves can also be traded.

Call: Looking to buy. Put: Looking to sell.

European options: can only exercise at the specified date. American options: can exercise any time before the specified date.

Derivative Traders: Hedger: Actually exercises the derivative. (To minimize loss) Speculator: Actually bets on the ups and downs of the derivatives. Arbitrageurs: Finding certain entry point to guarantee some winnings by calling / putting a combination of derivatives. (Too powerful and rare so we usually assume no arbitrage opportunities occur)

The price of an option is a function of

- S_0 : Current stock price.
- K : Strike price.
- T : Time to expiration.
- r : Risk free interest rate.
- Dividends that are expected to be paid.

More notation:

- S_T : Stock price at expiration day.
- C/P : American call / put.
- c/p : European call / put.

We assume no arbitrage opportunities (if there is then it disappears quickly).

We have

- $C \leq S_0$
- $S_0 - Ke^{rT} \leq c \leq S_0$.
- $P \leq K$
- $Ke^{-rT} - S_0 \leq p \leq Ke^{-rT}$

More:

- $c + Ke^{-rT} = p + S_0$ (idea: European call and put options worth the same price at time T)
- $S_0 - K \leq C - P \leq S_0 - Ke^{-rT}$

12 Mathematical Models for Option Pricing

12.1 Background Terminology

Risk-neutral: A situation where investors do not expect an increase in return when the risk increases. This has two consequences.

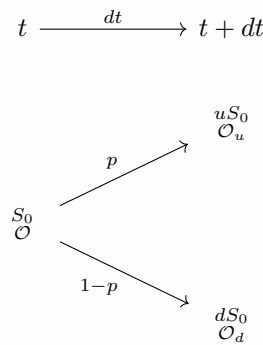
- The expected return of a stock is the risk free rate.
- The discount rate for the expected pay off of an option is the risk free.

Risk-less: A portfolio is risk-less when regardless of the outcome, the net total is constant.

12.2 One-Period Binomial Models

Definition 12.2.1: The Binomial Model

Suppose that t is the current time. An asset has a current stock price of S_0 with an unknown option price \mathcal{O} . After a time step of dt , the stock price either increase by a multiplicative factor of $u > 1$ or the stock price decreases by a multiplicative factor of $d < 1$. Denote the corresponding option value (the net gain from the option) by \mathcal{O}_u and \mathcal{O}_d respectively.



The number $0 < p < 1$ is an unknown quantity denoting the probability that the stock price increases.

Lemma 12.2.2

Assume there is no arbitrage opportunity and the world is risk-neutral. Assume the binomial model. Suppose a portfolio of buying $\Delta \in \mathbb{N}$ stocks and shorting a call option. Then the portfolio is risk-less when

$$\Delta = \frac{\mathcal{O}_u - \mathcal{O}_d}{uS_0 - dS_0}$$

Proposition 12.2.3

Assume there is no arbitrage opportunity and the world is risk-neutral. Assume the binomial model. Denote $V(t) = S_0\Delta - \mathcal{O}$ the value of the portfolio of buying Δ stocks and shorting a call option. Then the following are true.

- The probability p that the stock price increases is given by

$$p = \frac{e^{rdt} - d}{u - d}$$

- The option price is given by

$$\mathcal{O} = e^{-rdt} (p\mathcal{O}_u + (1-p)\mathcal{O}_d)$$

Proof. Accounting for the discount rate, we have

$$V(t + dt) = (uS_0\Delta - \mathcal{O}_u)e^{-r dt} = (dS_0\Delta - \mathcal{O}_d)e^{-r dt}$$

by definition of Δ . Equating $V(t)$ and $V(t + dt)$ gives the desired result. \square

Lemma 12.2.4

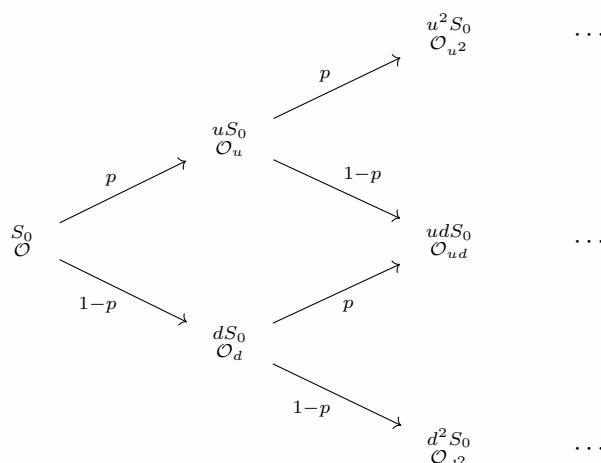
Assume there is no arbitrage opportunity and the world is risk-neutral. Assume the binomial model. Suppose that the implied volatility is σ . Then we have

$$u = e^{\sigma\sqrt{dt}} \quad \text{and} \quad d = e^{-\sigma\sqrt{dt}}$$

12.3 Multi-Period Binomial Models

Definition 12.3.1: The Multi-Period Binomial Model

Suppose that t is the current time. An asset has a current stock price of S_0 with an unknown option price \mathcal{O} . After a time step of dt , the stock price either increase by a multiplicative factor of $u > 1$ or the stock price decreases by a multiplicative factor of $d < 1$. Denote the corresponding option value (the net gain from the option) by \mathcal{O}_u and \mathcal{O}_d respectively.



The number $0 < p < 1$ is an unknown quantity denoting the probability that the stock price increases. The process iterates for n -times.

Notice at each time-step, the process is identical except with a different current stock and option price. Therefore we can iterate the one-period binomial model n -times to recover the option value.

Proposition 12.3.2

Assume there is no arbitrage opportunity and the world is risk-neutral. Assume the multi-period binomial model with period n . Then the European call option price is given by

$$\mathcal{O} = e^{-nr dt} \left(\sum_{i=0}^n \binom{n}{i} p^{n-i} (1-p)^i \mathcal{O}_{u^{n-i} d^i} \right)$$

12.4 Black-Scholes-Merton Model

Definition 12.4.1: Black-Scholes-Merton Model

Let T be a set time (in years). Denote $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ the stochastic processes that is the value of a certain stock. Let μ be the annual expected return. Let σ be the annual volatility of the stock price. The Black-Scholes-Merton Model makes the following assumptions

- The percentage in stock is normally distributed:

$$\frac{S_{t+dt} - S_t}{S_t} \sim N(\mu dt, \sigma^2 dt)$$

- μ and σ are constants over time.
- The risk free interest rate is constant over time.
- There are no dividends during the lifetime of the stock.
- There are no arbitrage opportunities.

Lemma 12.4.2

Let T be a set time (in years). Denote $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ the stochastic processes that is the value of a certain stock. Let μ be the annual expected return. Let σ be the volatility of the stock price. Assume the Black-Scholes-Merton model. Then we have

$$\ln(S_t) \sim N\left(\ln(S_0) + \left(\mu - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right)$$

Rate of return: The percentage return averaged over a certain amount of time.

Lemma 12.4.3

Let T be a set time (in years). Denote $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ the stochastic processes that is the value of a certain stock. Assume the Black-Scholes-Merton model. Let r be the continuously compounded rate of return of S . Then we have

$$r \sim N\left(\mu - \frac{\sigma^2}{2}, \frac{\sigma^2}{T}\right)$$

Volatility: a measure of uncertainty of return. So it is precisely the variance of the rate of return. Hence the annual volatility is σ . Checks out.

Proposition 12.4.4

Let T be a set time (in years). Denote $S : \Omega \times [0, T] \rightarrow \mathbb{R}$ the stochastic processes that is the value of a certain stock. Assume the Black-Scholes-Merton model. Then the European call option price according to the model is

$$c = S_0 F_N(d_1) - K e^{-rT} F_N(d_2)$$

where K is the strike price, r is the risk-free interest rate and F_N is the cumulative distribution function of $N(0, 1)$ and

$$d_1 = \frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \quad \text{and} \quad d_2 = \frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}$$