Probability Theory

a rigorous treatment to probability theory and its applications to stochastic processes

Khanh Nguyen

August 2024

Chapter 0

Preliminaries

0.1 Linear Algebra

Proposition 1. *Let* $A \in \mathbb{R}^{n \times n}$*, then* A *and* A^T *have the same eigenvalues*

Chapter 1

Probability Theory

1.1 Minimal measure theory to fulfil this

1.1.1 Measurable Space

Definition 1 (σ-algebra, measurable space). *Let X be a set. A σ-algebra Σ on X is a collection of subsets of X such that:*

1. $\varnothing, X \in \Sigma$
2. $A \in \Sigma \implies X - A \in \Sigma$
3. $E_1, E_2, ... \in \Sigma \implies \bigcup_{i=1}^{\infty} E_i \in \Sigma$

The pair (X, Σ) is called measurable space, and elements of Σ are called measurable set.

Definition 2 (σ-algebra generated by basis). *Let X be a set and $\mathcal{B}$ be a collection of subsets of X. Define $\sigma(B)$ by the smallest σ-algebra containing $\mathcal{B}$, that is, the intersection of all σ-algebras containing $\mathcal{B}$. (since intersection of arbitrary collection of σ-algebras is another σ-algbera, the defintion is well-defined)*

Definition 3 (product of measurable spaces, product σ-algebra). *Let $\{(X_i, \Sigma_i)\}_{i \in I}$ be a collection of measurable spaces. Define the product of measurable spaces (X, Σ) by*

$$X := \prod_{i \in I} X_i$$
$$\Sigma := \sigma\left(\prod_{i \in I} \Sigma_i\right)$$

where products are cartesian products. Σ is called product σ-algebra.

Definition 4 (measurable function, the category of measurable spaces). *Let $(X, \Sigma_X), (Y, \Sigma_Y)$ be measurable spaces. A function $f : X \to Y$ is called measurable if for every measurable set E_Y in (Y, Σ_Y), the preimage $f^{-1}E_Y$ is measurable in (X, Σ_X). The pair measurable space, measurable function form a category called the category of measurable spaces denoted by* Meas*, the product in this category is precisely the product of measurable spaces.*

Definition 5 (subspace). *Let (X, Σ) be a measurable space and $A \in \Sigma$ be a measurable set. Then A induces a measurable space (A, Σ_A) defined by*

$$\Sigma_A = \{A \cap E : E \in \Sigma\}$$

1.1.2 Measure Space

Definition 6 (measure, measure space). *Let (X, Σ) be a measurable space. A measure μ on (X, Σ) is a function $\mu : X \to [0, +\infty]$ such that*

1. $\mu(\varnothing) = 0$
2. $\mu(\coprod_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$

where $\coprod$ denotes the disjoint union. The triplet (X, Σ, μ) is called measure space.

Definition 7 (subspace). *Let (X, Σ, μ) be a measure space and $A \in \Sigma$ be a measurable set. Then A induces a measure space (A, Σ_A, μ_A) defined by*

$$\mu_A(E_A) = \mu(E \cap A)$$

where $E_A = E \cap A$

Definition 8 (measure-preserving map, the category of measure spaces). *Let $f : (X, \Sigma_X, \mu_X) \to (Y, \Sigma_Y, \mu_Y)$ be a measurable function, f is called measure preserving-map if*

$$\mu_X(f^{-1}E) = \mu_Y(E)$$

for all $E \in \Sigma_Y$. The pair measure space and measure-preserving map form a category called the category of measure space.

Definition 9 (pushforward measure). *Let $f : (X, \Sigma_X) \to (Y, \Sigma_Y)$ be a measurable function, and $\mu_X : \Sigma_X \to [0, +\infty]$ be a measure on X. The pushforward measure of μ_X by f is the unique measure μ_Y such that f is a measure-preserving map*

Theorem 1 (change of variables). *Let $\phi : (X, \Sigma_X, \mu_X) \to (Y, \Sigma_Y, \mu_Y)$ be a measure-preserving map. A measurable function $f : Y \to \mathbb{R}$ is integrable with respect to μ_Y if and only if the composition $f\phi$ is integrable with respect to μ_X, in that case, the integrals coincide*

$$\int_Y \phi d\mu_Y = \int_X f\phi d\mu_X$$

$$\begin{tikzcd}
(X, \Sigma_X, \mu_X) \arrow[r, "\phi"] \arrow[dr, dashed, "f\phi"'] & (Y, \Sigma_Y, \mu_Y) \arrow[d, "f"] \\
& \mathbb{R}
\end{tikzcd}$$

Equivalently, ϕ induces an isomorphism ϕ_* in L^1 spaces that preserves integral, i.e. $\int_Y f d\mu_Y = \int_X \phi_*(f) d\mu_X$ defined by

$$\begin{aligned}
\phi_* : L^1(Y, \Sigma_Y, \mu_Y) &\to L^1(X, \Sigma_X, \mu_X) \\
f &\mapsto \phi_*(f) = f\phi
\end{aligned}$$

1.2 Probability Spaces and Random Variables

1.2.1 Probability Space

Definition 10 (probability space). *A probability space $(\Omega, \mathcal{F}, P)$ is a measure space such that $P(\Omega) = 1$. Ω is called sample space, measurable sets in $\mathcal{F}$ are called events, P is called probability measure.*

Definition 11 (independence of events). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{E_i\}_{i \in I}$ be a collection of events. The collection is called independent if for any finite subcollection $J \subseteq I$*

$$P\left(\bigcap_{j \in J} E_j\right) = \prod_{j \in J} P(E_j)$$

1.2.2 Random Variables

Definition 12 (pushforward probability space, random variable). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(\mathcal{X}, \mathcal{F}_X)$ be a measurable space, and $X : \Omega \to \mathcal{X}$ be a measurable function. Let $P_X : \mathcal{F}_X \to \mathbb{R}$ be the pushforward measure of X, then $(\mathcal{X}, \mathcal{F}_X, P_X)$ is another probability space. $(\mathcal{X}, \mathcal{F}_X, P_X)$ is called pushforward probability space, the measurable function X is called random variable, and the pushforward measure P_X is called probability distribution.*

$$\begin{aligned}
P_X : \mathcal{F}_X &\to \mathbb{R} \\
E_X &\mapsto P(X^{-1}E_X)
\end{aligned}$$

If $\mathcal{X}$ is the codomain of the random variable $X : \Omega \to \mathcal{X}$, we call X a random variable on $\mathcal{X}$.

Remark 1. *:*

- *In probability theory, we usually start with the unique probability space $(\Omega, \mathcal{F}, P)$, namely, abstract probability space, and all random variables are measurable functions from Ω. Denote the collection of all random variables $\Omega \to X$ by*

$$\mathrm{RV}[\mathcal{X}] := \mathrm{Hom}((\Omega, \mathcal{F}), (\mathcal{X}, \mathcal{F}_X))$$

- *Without confusion, we identify the two events $E_X \in \mathcal{F}_X$ with $E = X^{-1}E_X \in \mathcal{F}$ and write*

$$P(E_X) := P_X(E_X)$$

Definition 13 (joint distribution)**.** *Let $\{X_i : \Omega \to \mathcal{X}_i\}_{i\in I}$ be a collection of random variables. Then, $X : \Omega \to \mathcal{X}$ is a random variable on the product of measurable spaces $\mathcal{X} = \prod_{i\in I} \mathcal{X}_i$ defined by*

$$X(\omega) := \prod_{i\in I} X_i(\omega)$$

X is called the joint random variable, the probability distribution on X is called joint distribution.

Remark 2. *Let $\{X_i : \Omega \to \mathcal{X}_i\}_{i\in I}$ be a collection of random variables, and X be the joint random variable. An $\tilde{E}_j = \prod_{i\in I} E_i$ be an event in X such that $E_i = \mathcal{X}_i$ for all but index j, that is, projections of $\tilde{E}_j$ on all coordinates are the whole space except coordinate j. Then, we identify $\tilde{E}_j$ by E_j.*

Definition 14 (independence of random variables)**.** *Let $\{X_i : \Omega \to \mathcal{X}_i\}_{i\in I}$ be a collection of random variables, and X be the joint random variable. The collection is called (mutually) independent if every collection of events $\{E_i \in X_i\}_{i\in I}$ is independent.*

Definition 15 (function on random variables)**.** *Let $\mathcal{X}, \mathcal{Y}$ be measurable spaces, $X : \Omega \to \mathcal{X}$ be a random variable on X. Let $f : \mathcal{X} \to \mathcal{Y}$ be a measurable function. Then, f induces* [1] *a random variable on Y defined by*

$$\begin{aligned} f_* : \mathrm{RV}[\mathcal{X}] &\to \mathrm{RV}[\mathcal{Y}] \\ X &\mapsto f_*X = fX \end{aligned}$$

Conditioning should be introduced here, however, it is a difficult topic and I did not have enough maturity to write an abstract introduction to conditioning, so I will put conditioning after real-valued random variables. In short, conditioning on an event is to induce new probability measure

1.3 Real-Valued Random Variables

Assume $\mathbb{R}$ is equipped with the Borel-algebra: the σ-algebra generated by open sets of the usual topology.

Definition 16 (real-valued random variable)**.** *A real-valued random variable $X : \Omega \to \mathbb{R}$ is a random variable on the measurable space $\mathbb{R}$. The collection of real-valued random variables is denoted by* $\mathrm{RV}[\mathbb{R}]$

Proposition 2 (algebra over field)**.** *Since $\mathbb{R}$ is a field, $\mathrm{RV}[\mathbb{R}]$ is an algebra over $\mathbb{R}$ with vector addition, and scalar multiplication, vector multiplication are defined by*

- ***vector addition**: $(X + Y)(\omega) = X(\omega) + Y(\omega)$*
- ***scalar multiplication**: $(cX)(\omega) = cX(\omega)$*
- ***vector multiplication**: $(XY)(\omega) = X(\omega)Y(\omega)$*

Definition 17 (distribution function, absolutely continuous random variable)**.** *Let $X : \Omega \to \mathbb{R}$ be a real-valued random variable, define $F_X : \mathbb{R} \to [0, 1]$ by*

$$F_X(a) = P(X \le a)$$

F_X is called distribtion function of random variable X. X is called (absolutely) continuous if F_X is an absolutely continuous function. When X is continuous, there exists a L^1 function $f_X : \mathbb{R} \to \mathbb{R}$ so that

$$P(X \le a) = F_X(a) = \int_{-\infty}^{a} f_X(x)dx$$

f_X is called density function.

From now, whenever we write f_X, we assume that X is continuous.

[1] composition of measurable functions is measurable

1.3.1 Expectation and Variance of Real-Valued Random Variables

Expectation of Real-Valued Random Variables

Definition 18 (expectation). *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : \Omega \to \mathcal{X} = \mathbb{R}$ be a real-valued random variable. Define the expectation $\mathbb{E}[-] : \mathrm{RV}[\mathbb{R}] \to \mathbb{R}$ by*

$$\begin{aligned}
\mathbb{E}[X] &:= \int_\Omega X dP \\
&:= \int_\Omega (\mathrm{id}\, X) dP && (\mathrm{id} : \mathcal{X} = \mathbb{R} \to \mathbb{R}) \\
&= \int_\mathbb{R} \mathrm{id}\, dP_X && \text{(by change of variables w.r.t pushforward } X\text{)} \\
&= \int_\mathbb{R} x dP_X(x)
\end{aligned}$$

Proposition 3 (linearity of expectation). *Expectation is a linear map* $\mathrm{RV}[\mathbb{R}] \to \mathbb{R}$. *That is,*

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

Proposition 4 (expectation of function on real-valued random variables). *Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued measurable function, then*

$$\begin{aligned}
\mathbb{E}[fX] &= \int_\Omega (fX) dP \\
&= \int_\mathbb{R} f(x) dP_X(x) && \text{(by change of variables w.r.t pushforward } X\text{)}
\end{aligned}$$

Proposition 5 (expectation of product of two independent random variables). *Let X, Y be independent real-valued random variables, then*

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

Proposition 6 (inner product space). $\mathrm{RV}[\mathbb{R}]$ *is an inner product space over $\mathbb{R}$ where the inner product is defined by*

$$\langle X, Y \rangle = \mathbb{E}[XY]$$

Theorem 2 (Cauchy-Schwarz inequality). *Since* $\mathrm{RV}[\mathbb{R}]$ *is an inner product space over $\mathbb{R}$, then if X, Y are real-valued random variables, then*

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

Theorem 3 (Markov inequality). *Let $X : \Omega \to [0, +\infty)$, then for any $a > 0$, we have*

$$P(\{X > a\}) \leq \frac{\mathbb{E}[X]}{a}$$

Proposition 7 (expectation as a sum of tail probabilities). *If $X : \Omega \to \mathbb{N}$, then*

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} nP(X = n) = \sum_{n=0}^{\infty} P(X > n)$$

If $X : \Omega \to \mathbb{R}$, then

$$\mathbb{E}[X] = \int_\mathbb{R} x = \int_0^\infty P(X > a) da$$

Variance of Real-Valued Random Variables

Definition 19 (variance, p-th moment). *Let $X : \Omega \to \mathbb{R}$. Define* $\mathrm{Var} : \mathrm{RV}[\mathbb{R}] \to \mathbb{R}$ *by*

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

$\mathrm{Var}(X)$ *is called variance of X, $\mathbb{E}[X^p]$ is called p-th moment of X, and $\mathbb{E}[|X|^p]$ is called p-th absolute moment of X*

Definition 20 (Chebyshev inequality). *Let $X : \Omega \to \mathbb{R}$, then for any $a > 0$*

$$P(\{|X - \mathbb{E}[X]| > a\}) \leq \frac{\mathrm{Var}(X)}{a^2}$$

Definition 21 (covariance, correlation)**.** *Let X, Y be real-valued random variables. Define the covariance* $\text{Cov} : \text{RV}[\mathbb{R}] \times \text{RV}[\mathbb{R}] \to \mathbb{R}$ *by*

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

Define the correlation $\text{Corr} : \text{RV}[\mathbb{R}] \times \text{RV}[\mathbb{R}] \to [-1, +1]$ *by*

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

Proposition 8. *Given a collection $\{X_1, X_2, ..., X_n\}$ of real-valued random variables with finite second moments, i.e $\mathbb{E}[X_i^2] < \infty$, then*

$$\text{Var}(X_1 + X_2 + ..., +X_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, X_j) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{i,j \in [n]\times[n]: i<j} \text{Cov}(X_i, X_j)$$

START FROM HERE

1.3.2 Limit Theorems

Definition 22 (convergence)**.** *Let $(X_n)_{n\in\mathbb{N}}$ and X be real-valued random variables defined on the same probability space (Ω, F, P) with probability distribution $(\nu_n)_{n\in\mathbb{N}}$ and ν respectively.*

1. *$X_n \to X$ almost surely if there exists a subset $\Omega_1 \subseteq \Omega$ with $P(\Omega_1) = 1$ such that for all $\omega \in \Omega_1$, as $n \to \infty$*

$$X_n(\omega) \to X(\omega)$$

2. *$X_n \to X$ in probability if for all $\epsilon > 0$, as $n \to \infty$*

$$P(|X_n - X| \geq \epsilon) \to 0$$

3. *$X_n \to X$ in distribution (or $\nu_v \to \nu$ weakly) if for all $a < b$ with $\nu(\{a\}) = \nu(\{b\}) = 0$*

$$\nu_n(a, b) \to \nu(a, b)$$

Remark 3. *Some remarks on convergence*

1. *almost surely convergence $\implies$ convergence in probability $\implies$ convergence in distribution*
2. *for any constant $c \in \mathbb{R}$, $X_n \to c$ in probability $\iff$ $X_n \to c$ in distribution.*
3. *weak convergence of $\nu_n \to \nu$ is equivalent to*

$$\int_{\mathbb{R}} f(x) d\nu_n \to \int f(x) d\nu$$

as $n \to \infty$ for all bounded (absolutely) continuous function $f : \mathbb{R} \to \mathbb{R}$. f is called test function.

4. *$X_n \to X$ in distribution means exactly $\nu_n \to \nu$ weakly, and hence $(X_n)_{n\in\mathbb{N}}$ and X need not to be defined on the same probability space.*

Theorem 4 (weak law of large numbers)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d (independent and identically distributed) real-valued random variables. Assume that the mean $\mu = \mathbb{E}[X_1]$ finite. Let $\sigma = \sqrt{\text{Var}(X_1)}$ and $S_n = \sum_{i=1}^{n} X_i$. Then the empirical average $\frac{S_n}{n} \to \mu$ in probability, i.e. for all $\epsilon > 0$, as $n \to \infty$*

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \to 0$$

Proof. Assume $\sigma < \infty$ [2], by Chebyshev inequality for real-valued random variable $\frac{S_n}{n}$,

$$\begin{aligned}
P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) &\leq \frac{1}{\epsilon^2}\operatorname{Var}\left(\frac{S_n}{n}\right) \\
&= \frac{1}{\epsilon^2}\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] \\
&= \frac{1}{\epsilon^2}\mathbb{E}\left[\left(\frac{X_1 + \ldots + X_n}{n} - \mu\right)^2\right] \\
&= \frac{1}{\epsilon^2}\mathbb{E}\left[\left(\frac{(X_1 - \mu) + \ldots + (X_n - \mu)}{n}\right)^2\right] \\
&= \frac{1}{n^2\epsilon^2}\mathbb{E}\left[((X_1 - \mu) + \ldots + (X_n - \mu))^2\right] && \text{(linearity of expectation)} \\
&= \frac{1}{n^2\epsilon^2}\left(\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)^2] + \sum_{(i,j)\in[n]\times[n]:i\neq j}\mathbb{E}[(X_i - \mu)(X_j - \mu)]\right) && \text{(linearity of expectation)} \\
&= \frac{1}{n^2\epsilon^2}\left(\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)^2] + \sum_{(i,j)\in[n]\times[n]:i\neq j}\mathbb{E}[X_i - \mu]\mathbb{E}[X_j - \mu]\right) && (X_i - \mu \text{ and } X_j - \mu \text{ are independent}) \\
&= \frac{1}{n^2\epsilon^2}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mu)^2] \\
&= \frac{\operatorname{Var}(X_1)}{n\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0
\end{aligned}$$

□

Theorem 5 (strong law of large numbers). *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d real-valued random variables with finite mean $\mu : \mathbb{E}[X_1] \in \mathbb{R}$. Then, almost surely*

$$\left(\frac{S_n}{n}\right)_{n\in\mathbb{N}} \to \mu$$

Lemma 1 (Borel-Cantelli). *Let (Ω, F, P) be a probability space and $A_n \in F$ is a sequence of events. Then*

1. *if $\sum_{i=1}^{\infty} P(A_n) < \infty$ then almost surely A_n eventually stops occurring, i.e. there is $\Omega_0 \subseteq \Omega$ with $P(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$, $\omega \notin A_n$ for all but finitely many n $P(A_n) \to 0$ as $n \to \infty$*
2. *if $(A_n)_{n\in\mathbb{N}}$ are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$ then almost surely A_n occur infinitely often, i.e. there is $\Omega_0 \subseteq \Omega$ with $P(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$, $\omega \notin A_n$ for finitely many n*

Sketch proof of strong law of large numbers. We will show that for each $\epsilon > 0$, let A_n be the event $\left\{\left|\frac{S_n}{n} - \mu\right| > \epsilon\right\}$, then $\sum_{n=1}^{\infty} P(A_n) < \infty$. By Borel-Cantelli, for almost every $\omega \in \Omega$, the event A_n eventually stops occurring, hence

$$\limsup_{n\to\infty}\left|\frac{S_n(\omega)}{n} - \mu\right| \leq \epsilon$$

for almost every $\omega \in \Omega$. Let $\Omega_\epsilon \subseteq \Omega$ be the set where it holds, $P(\Omega_\epsilon) = 1$. Choose a sequence $(\epsilon_i)_{i\in\mathbb{N}} \to 0$, take $\Omega_0 = \bigcap_{i=1}^{\infty}\Omega_{\epsilon_i}$ □

Theorem 6 (central limit theorem). *Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d random variables with finite mean μ and finite variance σ^2 and let $S_n = \sum_{i=1}^{n} X_i$ and $W_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then W_n converges in distribution to a standard Gaussian random variable Z, that is, for all $a < b$, as $n \to \infty$*

$$P(W_n \in [a, b]) \to P(Z \in [a, b]) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Theorem 7 (Poisson limit theorem - law of small numbers). *For $n \in \mathbb{N}$, let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with $P(X_i = 1) = \frac{\lambda}{n}$ for some $\lambda > 0$ modelling the occurrence of n independent rare events Then, $S_n = \sum_{i=1}^{n} X_i$ is a random variable modelling the number of occurrences*

As $n \to \infty$, S_n converges in distribution to $Pois(\lambda)$, i.e. for each $k = 0, 1, \ldots,$ as $n \to \infty$

$$P(S_n = k) \to \left(e^{-\lambda}\frac{\lambda^k}{k!}\right)$$

[2] this simple proof is only for the case of finite second moment

Proof.

$$\begin{aligned}
P(S_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k e^{(n-k)\log\left(1-\frac{\lambda}{n}\right)} \\
&= \left(\frac{n(n-1)...(n-k-1)}{n^k}\right)\left(\frac{\lambda^k}{k!}e^{(n-k)\left(-\frac{\lambda}{n}+o\left(\frac{\lambda}{n}\right)\right)}\right) \\
&\to \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}$$

□

Definition 23 (Fourier transform)**.** *Let X be a real-valued random variable with probability distribution μ. The Fourier transform of X is also called its characteristic function, is defined by*

$$\phi(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} d\mu(x)$$

Theorem 8. *The Fourier transform satisfies the following properties*

1. *$\phi(t) \in \mathbb{C}$ with $|\phi(t)| \in [0,1]$ for all $t \in \mathbb{R}$ and $\phi(0) = 1$*
2. *ϕ determines the distribution of X with $\phi^{(k)}(0) = i^k\mathbb{E}[X^k]$*
3. *if $\phi_n(t) = \mathbb{E}[e^{itX_n}]$ and $\phi(t) = \mathbb{E}[e^{itX}]$, then $\phi_n \to \phi$ pointwise on $[-a,+a]$ for some $a > 0$ implies $X_n \to X$ in distribution*

Proof of central limit theorem. Suppose $\mu = 0, \sigma = 1$, let $\psi(t) = \mathbb{E}[e^{itX_1}]$. Then $\psi(0) = 1, \psi'(0) = 0$ and $\psi''(0) = -1$, Taylor theorem,

$$\psi(t) = 1 - \frac{t^2}{2} + o(t^2)$$

where $\frac{o(f(\epsilon))}{f(\epsilon)} \to 0$ as $\epsilon \to 0$ for any function f For any $t \in \mathbb{R}$,

$$\begin{aligned}
\phi_n(t) &= \mathbb{E}[e^{itW_n}] \\
&= \mathbb{E}\left[e^{i\frac{t}{\sqrt{n}}\sum_{i=1}^n X_i}\right] \\
&= \mathbb{E}\left[e^{i\frac{t}{\sqrt{n}}X_1}...e^{i\frac{t}{\sqrt{n}}X_n}\right] \\
&= \mathbb{E}\left[e^{i\frac{t}{\sqrt{n}}X_1}\right]...\mathbb{E}\left[e^{i\frac{t}{\sqrt{n}}X_n}\right] && (X_i \text{ are independent}) \\
&= \psi\left(\frac{t}{\sqrt{n}}\right)^n \\
&= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \\
&= e^{n\log\left(1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)} \\
&= e^{n\left[\left(-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)+o\left(-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)\right]} \\
&\to e^{-\frac{t^2}{2}}
\end{aligned}$$

Since $\mathbb{E}[e^{itZ}] = e^{-\frac{t^2}{2}}$ is the characteristic function of standard Gaussian Z. So, $W_n \to Z$ in distribution. □

Definition 24 (Laplace transform)**.** *Let X be non-negative real-valued random variable with probability distribution μ. Then the Laplace transform of X (or μ) is defined to be*

$$\Lambda(\lambda) = \mathbb{E}[e^{-\lambda X}] = \int_0^\infty e^{-\lambda x} d\mu(x)$$

Definition 25 (generating function)**.** *Let X be a non-negative $\mathbb{N}$-valued random variable with probability mass function $p_n = P(X = n)$ for $n \in \{0, 1, ...\}$. Then the generating function of X (or p_n) is defined to be*

$$G(s) = \mathbb{E}[s^X] = \sum_{n=0}^\infty s^n p_n$$

for $s \geq 0$ so that the sum converges

CONTINUE FROM HERE

1.4 Conditioning

1.4.1 Conditioning on Event

Definition 26 (conditioning on event, conditional probability, conditional probability space, conditional probability distribution)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space and E be an event such that $P(E) > 0$, then E induces another probability space $(\Omega, \mathcal{F}, P(\cdot|E))$ where the probability measure $P(\cdot|E)$ is defined by*

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

$P(A|E)$ is call conditional probability, $P(\cdot|E)$ is called conditional probability measure and $(\Omega, \mathcal{F}, P(\cdot|E))$ is called conditional probability space. If $X : \Omega \to \mathcal{X}$ is a random variable, conditioning on event E yields a new probability distribution, namely, conditional probability distribution $P_{X|E} = P(\{X \in \cdot\}|E) : \mathcal{F}_X \to \mathbb{R}$

$$P_{X|E}(E_X) = P(\{X \in E_X\}|E) = P(X^{-1}E_X|E) = \frac{P(X^{-1}E_X \cap E)}{P(E)}$$

Definition 27 (conditional expectation on event, conditional variance)**.** *Let X be a real-valued random variable on (Ω, F, P) and E is a event with $P(E) > 0$. Define the conditional expectation of X conditioned by event E by*

$$\mathbb{E}[X|E] = \int_\Omega X dP(\cdot|E) = \frac{\mathbb{E}[X 1_E]}{P(E)}$$

Define the conditional variance of X conditioned by event E by

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

1.4.2 Conditioning on Discrete Random Variable

Definition 28 (conditioning on discrete random variable)**.** *Let $X : \Omega \to \mathcal{X}$ be a random variable, $Y : \Omega \to \mathcal{Y}$ be a discrete random variable. Define $P(X|Y) : \mathcal{Y} \to \mathrm{Hom}(\mathcal{F}_X, \mathbb{R})$ as a function from value in $\mathcal{Y}$ to a distribution on $\mathcal{X}$ by*

$$\begin{aligned} P(X|Y) : \mathcal{Y} &\to \mathrm{Hom}(\mathcal{F}_X, \mathbb{R}) \\ y &\mapsto P(\{X \in \cdot\}|\{Y = y\}) \end{aligned}$$

Definition 29 (conditional expectation on discrete random variable)**.** *Let $f : \mathrm{Hom}(\mathcal{F}_X \to \mathbb{R}) \to \mathbb{R}$ be a real-valued function on distribution on $\mathcal{X}$ (e.g expectation, variance). Then, the composition of $f \circ P(X|Y)$ is a function $\mathcal{Y} \to \mathbb{R}$. When f is expectation, we have*

$$\begin{aligned} \mathbb{E}[X|Y] : \mathcal{Y} &\to \mathbb{R} \\ y &\mapsto \mathbb{E}[X|\{Y = y\}] \end{aligned}$$

Proposition 9 (tower property of conditional expectation)**.** *When X is a real-valued random variables and Y is a discrete random variable, let $A \subseteq \mathcal{X}$ be an event, then $P(X \in A)$ can be recovered from $P(X|Y)$ by*

$$P(X \in A) = \sum_{y \in \mathcal{Y}} P(X \in A|Y = y)P(Y = y) = \mathbb{E}_Y[P(X \in A|Y)]$$

$\mathbb{E}[X]$ can be recovered from $\mathbb{E}[X|Y] : \mathcal{Y} \to \mathbb{R}$ by

$$\mathbb{E}[X] = \sum_{y \in \mathcal{Y}} \mathbb{E}[X 1_{\{Y=y\}}] = \sum_{y \in \mathcal{Y}} \mathbb{E}[X|Y]P(Y = y) = \mathbb{E}[\mathbb{E}[X|Y]]$$

More generally, let $f : \mathcal{X} = \mathbb{R} \to \mathbb{R}$, then

$$\mathbb{E}[f(X)] = \sum_{y \in \mathcal{Y}} \mathbb{E}[f(X) 1_{\{Y=y\}}] = \sum_{y \in \mathcal{Y}} \mathbb{E}[f(X)|Y]P(Y = y) = \mathbb{E}[\mathbb{E}[f(X)|Y]]$$

$P(X \in -)$ is a mixture of conditional distributions $P(X \in -|Y = y)$ with mixture coefficients $P(Y = y)$. $\mathbb{E}[X]$ is a mixture of conditional expectation $\mathbb{E}[X|Y = y]$ with mixture coefficients $P(Y = y)$

Remark 4 (marginal distribution on many variables)**.** *Let $\phi(X, Y)$ be a function of random variables which is itself another random variable. We can write*

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

Now, $\psi = \mathbb{E}[\phi|X]$ is a random variable, then

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

1.4.3 Conditioning on Continuous Random Variable

Definition 30 (conditioning on continuous random variable)**.** *When both X, Y are continuous random variables on $\mathbb{R}$, define $P(X \in A|Y = y_0)$ and $E[X|Y = y_0]$ by the conditional density*

$$f_X(x|Y = y_0) = \frac{f(x, y_0)}{\int_{\mathbb{R}} f(x, y_0)dx}$$

1.4.4 Conditioning on σ-Algebra

Definition 31 (conditioning on σ-algebra)**.** *Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra* [3] *in $(\Omega, \mathcal{F}, P)$ and $X : \Omega \to \mathcal{X}$ be a random variable. Define $P(X|\mathcal{G}) : \mathcal{G} \to \mathrm{Hom}(\mathcal{F}_X, \mathbb{R})$ as a function from events in $\mathcal{G}$ to a distribution on $\mathcal{X}$ by*

$$\begin{aligned} P(X|\mathcal{G}) : \mathcal{G} &\to \mathrm{Hom}(\mathcal{F}_X, \mathbb{R}) \\ E_{\mathcal{G}} &\mapsto P(\{X \in \cdot\}|E_{\mathcal{G}}) \end{aligned}$$

When $\mathcal{G}$ is countable (discrete random variable), the set function $P(X|\mathcal{G}) : \mathcal{G} \to \mathrm{Hom}(\mathcal{F}_X, \mathbb{R})$ can be characterized by a function

$$\begin{aligned} \Omega &\to \mathrm{Hom}(\mathcal{F}_X, \mathbb{R}) \\ g &\mapsto P(\{X \in \cdot\}|G) \end{aligned}$$

where $G \in \mathcal{G}$ is the smallest set in $\mathcal{G}$ containing g, the smallest set in $\mathcal{G}$ containing g is the intersection of all events in $\mathcal{G}$ containing g. This is coincide with the discrete random variable case.

Remark 5 (conditional expectation on σ-algebra)**.** *When X is a real-valued random variable, we have the conditional expectation*

$$\begin{aligned} \mathbb{E}[X|\mathcal{G}] : \mathcal{G} &\to \mathbb{R} \\ G &\mapsto \mathbb{E}[X|G] \end{aligned}$$

If $\mathcal{G}$ is countable, then the set function $\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \to \mathbb{R}$ can be characterized by a function

$$\begin{aligned} \Omega &\to \mathbb{R} \\ g &\mapsto \mathbb{E}[X|G] \end{aligned}$$

where $G \in \mathcal{G}$ is the smallest set in $\mathcal{G}$ containing g

Proposition 10 (tower property of conditional expectation)**.** *Let $\mathcal{H} \subseteq \mathcal{G}$ be sub-σ-algebras in $(\Omega, \mathcal{F}, P)$ and $X : \Omega \to \mathbb{R}$ be a real-valued random variable. Then,*

$$\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$$

[3] a σ-algebra that is a subset

Chapter 2

Time-Homogeneous Markov Chain

In this chapter, all Markov chains are time-homogeneous

2.1 Markov Chain Basics

Definition 32 (time-homogeneous Markov chain). *A stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ with countable state space S is called a Markov chain if*

$$P(X_{n+1} = y|X_0 = x_0, ..., X_n = x_n) = P(X_{n+1} = y|X_n = x_n)$$

for all $n \in \mathbb{N}_0$. If $P(X_{n+1} = y|X_n = x_n) = P(X_1 = y|X_0 = x)$ for all $n \in \mathbb{N}_0$, then X is called time-homogeneous, in that case we define $\Pi : S \times S \to \mathbb{R}$

$$\Pi(x, y) = P(X_1 = y|X_0 = x)$$

Π is called transition probability matrix. When S is finite, Π is a right stochastic matrix, that is, all elements are non-negative and sum of every row is 1

Remark 6 (n-step). *Let $\Pi^n : S \times S \to \mathbb{R}$ be defined by*

$$\Pi^n(x, y) = P(X_n = y|X_0 = x) = \sum_{(x=z_0,z_1,...,z_n=y)} \Pi(z_0, z_1)\Pi(z_1, z_2)...\Pi(z_{n-1}, z_n)$$

where the sum is over all possible path $(x = z_0, z_1, ..., z_n = y)$ of length n. Π^n is also a right stochastic matrix.

Proposition 11 (Π acting on row vector). *Let X be a time-homogeneous Markov chain with countable state space S and $\mu : S \to \mathbb{R}$ be a distribution for X_0, then the distribution of X_n, denoted by $\mu_n : S \to \mathbb{R}$ is*

$$\mu_n(y) = P(X_n = y) = \sum_{x\in S} P(X_n = y, X_0 = x) = \sum_{x\in S} \mu_0(x)\Pi^n(x, y)$$

If S is finite, we can write it in matrix form $\mu_n = \mu_0\Pi^n$

$$\begin{aligned} \mathrm{Hom}(S, \mathbb{R}) \times \mathrm{Hom}(S \times S, \mathbb{R}) &\to \mathrm{Hom}(S, \mathbb{R}) \\ (\mu_0, \Pi^n) &\mapsto \mu_n = \mu_0\Pi^n \end{aligned}$$

Proposition 12 (Π acting on column vector). *Let $f : S \to \mathbb{R}$ be a function on random variable X which is also a linear function on S. Let $\Pi^n f : S \to \mathbb{R}$ be defined by*

$$(\Pi^n f)(x) = \mathbb{E}[f(X_n)|X_0 = x] = \sum_{y\in S} \Pi^n(x, y)f(y)$$

If S is finite, we can write it in matrix form $\Pi^n f$

$$\begin{aligned} \mathrm{Hom}(S \times S, \mathbb{R}) \times \mathrm{Hom}(S, \mathbb{R}) &\to \mathrm{Hom}(S, \mathbb{R}) \\ (\Pi^n, f) &\mapsto \Pi^n f \end{aligned}$$

Proposition 13 (stationary measure). *Since $1 : S \to \mathbb{R}$ is a right eigenvector of Π with eigenvalue 1, Π also has a left eigenvector with eigenvalue 1. In other words, there exists a measure $\nu : S \to \mathbb{R}$ such that $\nu\Pi = \nu$, such ν is called stationary measure*

Proposition 14 (spectrum of Π)**.** *All complex eigenvalues of* Π *have norm less than or equal* 1

Proof. Let (λ, g) be a right eigenvalue eigenvector of Π where $\lambda \in \mathbb{C}$, $g : S \to \mathbb{C}$. Suppose S is finite and $|g(x)|$ achieves maximum at x_0, then

$$|\lambda||g(x_0)| = |\lambda g(x_0)| = |(\Pi g)(x_0)| = \left|\sum_{y \in S} \Pi(x_0, y) g(y)\right| \leq \left|\sum_{y \in S} \Pi(x_0, y)\right| |g(x_0)| = |g(x_0)|$$

That is, $|\lambda| \leq 1$ □

Proposition 15 (intercommunicating states)**.** *Let S be the state space, two states $x, y \in S$ are called intercommunicate, denoted by $x \sim y$ if there exist $m, n \in \mathbb{N}_0$ such that*

$$P(X_m = y | X_0 = x) > 0$$
$$P(X_n = x | X_0 = y) > 0$$

The relation is an equivalence relation that partitions the state space S into equivalence classes of intercommunicating.

Definition 33 (irreducible Markov chain)**.** *A time-homogeneous Markov chain is irreducible if S is a single equivalence class with respect to intercommunicating intercommunicating equivalence relation $\sim$*

2.2 Recurrence Transience

Remark 7 (notation)**.** *Given a stochastic process $X = (X_0, X_1, ...)$, let E be any event and Y be a random variable, then we write*

$$P_x(E) = P(E | X_0 = x)$$
$$\mathbb{E}_x[Y] = \mathbb{E}[Y | X_0 = x]$$

Definition 34 (T_x^k, f_{xy}, N_x, $G(x, y)$)**.** *Let X be a stochastic process with countable state space S, define the following:*

1. *Let $T_x^0 = 0$, for any $k > 0$, let T_x^k be the random variable modelling the time when X visits x at the k-th time, that is,*

$$T_x^k = \min\{n \in \mathbb{N}_0 : n > T_y^{k-1}, X_n = x\}$$

2. *Let f_{xy} be the probability that X visits y in finite time given $X_0 = x$, that is,*

$$f_{xy} = P_x(T_y^1 < \infty)$$

3. *Let N_x be the random variable modelling the number of visiting state x, that is*

$$N_x = \sum_{n=0}^{\infty} 1_{\{X_n = x\}}$$

4. *Let $G(x, y)$ be the expected number of visiting y given $X_0 = x$, that is*

$$G(x, y) = \mathbb{E}_x[N_y]$$

Definition 35 (recurrence transience)**.** *Let X be a stochastic process with countable state space S. A state $x \in X$ is called recurrent if $f_{xx} = 1$ and called transient if $f_{xx} < 1$*

Proposition 16 (expected number of visits for irreducble Markov chain)**.** *Let X be a time-homogeneous irreducible Markov chain with countable state space S. Let $X_0 = x$, the the expected number of visiting x is*

$$G(x, x) = \frac{1}{1 - f_{xx}}$$

Proof. We will show that N_x is a geometric random variable with parameter $p = 1 - f_{xx}$. For each $k > 0$, we have

$$P_x(N_x = k) = P_x(T_x^{k-1} < \infty, T_x^k = \infty)$$
$$P_x(N_x \geq k) = P_x(T_x^{k-1} < \infty)$$

We have

$$
\begin{aligned}
&P_x(N_x \geq k) - P_x(N_x \geq k+1) \\
&= P_x(N_x = k) \\
&= P_x(T_x^{k-1} < \infty, T_x^k = \infty) \\
&= \sum_{n=0}^{\infty} P_x(T_x^{k-1} = n, T_x^k = \infty) \\
&= \sum_{n=0}^{\infty} P_x(T_x^{k-1} = n) P_x(T_x^k = \infty | T_x^{k-1} = n) \\
&= \sum_{n=0}^{\infty} P_x(T_x^{k-1} = n) P_x(T_x^1 = \infty | T_x^0 = n) && \text{(time-homogeneous)} \\
&= \sum_{n=0}^{\infty} P_x(T_x^{k-1} = n) P(T_x^1 = \infty | X_0 = x) \\
&= \sum_{n=0}^{\infty} P_x(T_x^{k-1} = n)(1 - f_{xx}) \\
&= (1 - f_{xx}) \sum_{n=0}^{\infty} P_x(T_x^{k-1} = n) \\
&= (1 - f_{xx}) P_x(T_x^{k-1} < \infty) \\
&= (1 - f_{xx}) P_x(N_x \geq k)
\end{aligned}
$$

Therefore,

$$P_x(N_x \geq k) = f_{xx}^k \text{ and } P_x(N_x = k) = f_{xx}^k(1 - f_{xx})$$

□

Proposition 17 (recurrence transience as a class property)**.** *Let $x, y \in S$ be two inter-communicating states. Then, x, y are either both transient or both recurrent*

Proof. It suffices to show that $G(x,x) = \infty$ if and only if $G(y,y) = \infty$. By assuming $x \sim y$, there exists $k, l \in \mathbb{N}$ such that

$$\Pi^k(x,y) > 0 \text{ and } \Pi^l(y,x) > 0$$

Note that, for all $n \in \mathbb{N}_0$, we have the probability of starting from x then coming back to x after $k+n+l$ steps is greater than probability of starting from x, going to y in k steps, staying in y in n steps, then coming back to x in l steps, that is

$$\Pi^{k+l+n}(x,x) \geq \Pi^k(x,y)\Pi^n(y,y)\Pi^l y, x$$

Summing over $n \in \mathbb{N}_0$, we have

$$
\begin{aligned}
G(x,x) &= \mathbb{E}_x\left[\sum_{m=0}^{\infty} 1_{\{X_m = x\}}\right] \\
&= \sum_{m=0}^{\infty} \Pi^m(x,x) \\
&\geq \sum_{m=k+l}^{\infty} \Pi^m(x,x) \\
&\geq \sum_{n=0}^{\infty} \Pi^k(x,y)\Pi^n(y,y)\Pi^l(y,z) \\
&= \Pi^k(x,y)\Pi^l(y,z) \sum_{n=0}^{\infty} \Pi^n(y,y) \\
&= \Pi^k(x,y)\Pi^l(y,z) G(y,y)
\end{aligned}
$$

Therefore, $G(y,y) = \infty$ implies $G(x,x) = \infty$ □

Definition 36 (recurrent transient Markov chain). *A time-homogeneous irreducible Markov chain with countable state space is called recurrent/transient if its states are recurrent/transient*

Corollary 1 (irreducible finite state Markov chain). *A time-homogeneous irreducible Markov chains with finite state space are recurrent*

Proof. Since Markov chain is finite, there is at least one state with expected number of visits being infinity, that state is recurrent. Moreover, irreducibility implies every other state is inter-communicating with the recurrent state, therefore, all states are recurrent. □

Proposition 18. *Let X be a time-homogeneous irreducible Markov chain with countable state space S. Then,*

1. *If X is recurrent, then $P(N_x = \infty) = 1$ for all $x \in S$ and $G(x, y) = \infty$ for all $x, y \in S$.*
2. *If X is transient, then $P(N_x < \infty) = 1$ for all $x \in S$ and $G(x, y) < \infty$ for all $x, y \in S$*

Proof. Since the distribution of $X = (X_n)_{n \in \mathbb{N}_0}$ is a mixture with different starting position, i.e. for any event E

$$P(E) = \sum_{x \in S} \mu(x) P_x(E)$$

It suffices to prove for the case when X starts from any state $s \in S$.

1. X is recurrent

   For any $y \in S$, since X returns to x infinitely many times, let $X_n = x$, by irreducibility, there exists $m \in \mathbb{N}$ such that

   $$P(X_{n+m} = y | X_n = x) = \Pi^m(x, y) \geq 0$$

   Therefore, everytime X visits x, there is a positive probability X visits $y \in S$, that is, visiting y is a sequence of i.i.d Bernoulli random variables of positive parameter. Hence, number of visits y is infinite. Then, $G(x, y) = \infty$

2. X is transient

   Probability of starting from x, going to y in m steps, then going back to x in p steps is less than probability of starting from x and going back to x in $n = m + p$ steps.

   $$\begin{aligned} G(x, x) &= \sum_{n=0}^{\infty} \Pi^n(x, x) \\ &\geq \sum_{n=m}^{\infty} \Pi^n(x, x) \\ &\geq \sum_{p=0}^{\infty} \Pi(x, y)^m \Pi^p(y, x) \\ &= \Pi(x, y)^m \sum_{p=0}^{\infty} \Pi^p(y, x) \\ &= \Pi(x, y)^m G(y, x) \end{aligned}$$

   Hence, $G(x, x)$ is finite implies $G(y, x)$ is finite. $G(y, x)$ finite implies $P_y(N_x = \infty) = 0$

□

Remark 8 (transient - escape to infinity). *In the transient case, X escape to infinity with probability 1 in the following sense: For any finite set of states F, with probability 1*

$$\max\{n \in \mathbb{N} : X_n \in F\} < \infty$$

Proposition 19 (escaping from a finite set). *Let X be an irreducible Markov chain with countable state space S. Let $F \subseteq S$ be a finite, and $T_{F^c} = T_{F^c}(X) = \min\{n \geq 0 : X_n \notin F\}$ be the first time X exits from F. Then there exists $C > 0$ and $\rho \in (0, 1)$ such that for all $n \in \mathbb{N}_0$ and all initial distributions*

$$P(T_{F^c}(X) > n) \leq C\rho^n$$

Proof. Let $\rho \in [0,1]$ be defined by

$$\rho = \max\{P(X_1 = y|X_0 = x) : y \in F, x \in F\}$$

We can assume $\rho < 1$ since if $P(X_1 = y|X_0 = x) = 1$, as F is finite, we can merge two states x, y into a new state and the merging process terminates with $\rho < 1$ and a finite set of states F. Then

$$\begin{aligned} P(T_{F^c}(X) > n) &= P(X_0 \in F, X_1 \in F, ..., X_n \in F) \\ &= P(X_0 \in F) \prod_{i=0}^{n} P(X_{i+1} \in F|X_i \in F) \\ &\leq P(X_0 \in F)\rho^{n-1} \end{aligned}$$

□

Theorem 9 (Pólya 1921). *The symmetric random walk on $\mathbb{Z}^d$ is recurrent in dimension $d = 1, 2$ and transient in $d \geq 3$*

Proof. *TODO* □

CONTINUE FROM HERE

2.3 Stationary Measure

Proposition 20 (limiting distribution of transient Markov chain). *Let X be an irreducible transient Markov chain. Then for all $x, y \in S$,*

$$\Pi^n(x, y) \to 0$$

as $n \to \infty$. Consequently, for any initial distribution, $P(X_n = y) \to 0$ for all $y \in S$

Proof. Since X is transient, for any $x, y \in S$

$$G(x, y) = \sum_{n=0}^{\infty} \Pi^n(x, y) < \infty$$

Then, $\Pi^n(x, y) \to 0$ as $n \to \infty$. □

Definition 37 (stationary distribution, stationary measure). *Let X be a Markov chain with countable state space S and transition matrix Π. A probability distribution μ on S is called stationary distribution for X if*

$$\sum_{x \in S} \mu(x)\Pi(x, y) = \mu(y)$$

for all $y \in S$. That is, $\mu\Pi = \mu$. In other words, if X_0 has distribution μ, then X_1 also has distribution μ, hence so do $X_2, X_3, ...$. In general, any $\nu : S \to [0, +\infty)$ with $\nu\Pi = \nu$ and $\sum_{x \in S} \nu(x) \in (0, +\infty]$ is called stationary measure for X

2.4 Positive Recurrence, Null Recurrence, Existence of Stationary Measure

Definition 38 (positive recurrent Markov chain, null recurrent Markov chain). *Let X be an irreducible Markov chain with countable state space S. Let $T_x = T_x^1$, we call X positive recurrent if $\mathbb{E}_x[T_x] < \infty$ for all $x \in S$ and null recurrent if $\mathbb{E}_x[T_x] = \infty$ for all $x \in S$.*

Proposition 21 (positive recurrence and positive recurrence as a class property). *If x and y are two intercommunicating recurrent states, then they are either both positive recurrent or both null recurrent.*

Proof. *TODO* □

Theorem 10 (existence of stationary measure for recurrence Markov chain). *Let X be an irreducible recurrent Markov chain with state space S. For each $x \in S$,*

$$\nu(y) = \mathbb{E}_x\left[\sum_{n=0}^{T_x - 1} 1_{\{X_n = y\}}\right] = \mathbb{E}_x\left[\sum_{n=0}^{\infty} 1_{\{X_n = y\}} 1_{\{n < T_x\}}\right] = \sum_{n=0}^{\infty} \mathbb{E}_x[1_{\{X_n = y, n < T_x\}}] = \sum_{n=0}^{\infty} P_x(X_n = y, n < T_x)$$

is a stationary measure. If X is positive recurrent, then we can normalize ν to a stationary distribution.

Proof. The technique is called cycle trick (*need to redo*). It suffices to show that $\sum_{z\in S} \nu(z)\Pi(z,y) = \nu(y)$ for all $y \in S$. We have

Case 1: $y \neq x$

$$
\begin{aligned}
&\nu(y) \\
&= \sum_{n=0}^{\infty} P_x(X_n = y, n < T_x) \\
&= \sum_{n=1}^{\infty} P_x(X_n = y, n < T_x) && (P_x(X_0 = y, 0 < T_x) = 0) \\
&= \sum_{n=1}^{\infty} \sum_{z\in S} P_x(X_{n-1} = z, X_n = y, n < T_x) \\
&= \sum_{z\in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, X_n = y, n < T_x) && (\text{Tonelli theorem}) \\
&= \sum_{z\in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, X_n = y, n-1 < T_x) && (\{X_n = y, n < T_x\} = \{X_n = y, n-1 < T_x\}) \\
&= \sum_{z\in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, n-1 < T_x) P_x(X_n = y | X_{n-1} = z, n-1 < T_x) \\
&= \sum_{z\in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, n-1 < T_x) P_x(X_n = y | X_{n-1} = z) && (\{n-1 < T_x\} = \{X_0 \neq x, .., X_{n-1} \neq x\}) \\
&= \sum_{z\in S} \Pi(z,y) \sum_{n=1}^{\infty} P_x(X_{n-1} = z, n-1 < T_x) \\
&= \sum_{z\in S} \Pi(z,y)\nu(z)
\end{aligned}
$$

Case 2: $y = x$

$$
\begin{aligned}
&\nu(x) \\
&= \sum_{n=0}^{\infty} P_x(X_n = x, n < T_x) \\
&= 1 \\
&= \sum_{n=1}^{\infty} P_x(n = T_x) \\
&= \sum_{n=1}^{\infty} \sum_{z \in S} P_x(X_{n-1} = z, n = T_x) \\
&= \sum_{z \in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, n = T_x) && \text{(Tonelli theorem)} \\
&= \sum_{z \in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, X_n = x, n-1 < T_x) && (\{X_n = x, n-1 < T_x\} = \{n < T_x\}) \\
&= \sum_{z \in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, n-1 < T_x) P_x(X_n = x | X_{n-1} = z, n-1 < T_x) \\
&= \sum_{z \in S} \sum_{n=1}^{\infty} P_x(X_{n-1} = z, n-1 < T_x) P_x(X_n = x | X_{n-1} = z) && (\{n-1 < T_x\} = \{X_0 \neq x, .., X_{n-1} \neq x\}) \\
&= \sum_{z \in S} \Pi(z, x) \sum_{n=1}^{\infty} P_x(X_{n-1} = z, n-1 < T_x) \\
&= \sum_{z \in S} \Pi(z, x) \nu(z)
\end{aligned}
$$

Hence, ν is stationary. □

Theorem 11 (uniqueness of stationary measure for recurrent Markov chain)**.** *Let X be a recurrent Markov chain and $\nu : S \to \mathbb{R}$ be defined by*

$$\nu(y) = \mathbb{E}_x \left[\sum_{n=0}^{T_x - 1} 1_{\{X_n = y\}} \right] = \sum_{n=0}^{\infty} P_x(X_n = y, n < T_x)$$

Then if $\tilde{\nu} : S \to \mathbb{R}$ is another stationary measure for X, then there exists $C \in \mathbb{R}$ such that $\tilde{\nu}(y) = C\nu(y)$ for all $y \in S$

Proof. Without loss of generality, assume $\tilde{\nu}(x) = 1$ we will prove that $\tilde{\nu}(y) = \nu(y)$. By stationary of $\tilde{\nu}$, we have

$$\tilde{\nu}(y) = \sum_{z_1 \in S} \tilde{\nu}(z_1) \Pi(z_1, y) = \Pi(x, y) + \sum_{z_1 \neq x} \tilde{\nu}(z_1) \Pi(z_1, y)$$

Apply the same decomposition for z_1

$$
\begin{aligned}
&\tilde{\nu}(y) \\
&= \Pi(x, y) + \sum_{z_1 \neq x} \tilde{\nu}(z_1) \Pi(z_1, y) \\
&= \Pi(x, y) + \sum_{z_1 \neq x} \left(\Pi(x, z_1) + \sum_{z_2 \neq x} \tilde{\nu}(z_2) \Pi(z_2, z_1) \right) \Pi(z_1, y) \\
&= \Pi(x, y) + \sum_{z_1 \neq x} \Pi(x, z_1) \Pi(z_1, y) + \sum_{z_1 \neq x} \sum_{z_2 \neq x} \tilde{\nu}(z_2) \Pi(z_2, z_1) \Pi(z_1, y) \\
&= \Pi(x, y) + P_x(X_2 = y, X_1 \neq x) + \sum_{z_1 \neq x} \sum_{z_2 \neq x} \tilde{\nu}(z_2) \Pi(z_2, z_1) \Pi(z_1, y)
\end{aligned}
$$

TODO - finish this later □

2.5 Long Time Limit of Markov Chain

Remark 9. *For transient Markov chain, we have shown that $\Pi^n(x,y) = P_x(X_n = y) \to 0$ as $n \to \infty$ for all $x, y \in S$*

Definition 39 (period of an irreducible Markov chain). *Let X be an irreducible Markov chain with state space S and transition matrix Π. The period $r_x \in \mathbb{N}$ of a state x is defined by*

$$\gcd\{n \in \mathbb{N} : \Pi^n(x,x) > 0\}$$

It can be shown that $r_x = r_y$ for all $x, y \in S$. Hence, the period $r \in \mathbb{N}$ of X is defined by the period of any $x \in S$. A Markov chain is called periodic if $r \geq 2$ and aperiodic if $r = 1$

Remark 10 (cyclic structure of periodic Markov chain). *For a periodic Markov chain with period r, we can divide the state space S into r equivalence classes $S_1, S_2, ..., S_r$. Let $i \in [r]$, for any state in S_i, the only transition possible is to another state in S_{i+1} (where $S_{r+1} = S_1$). If we define $Y_n = X_{nr}$, then Y is an aperiodic Markov chain with state space S_i where $Y_0 \in S_i$. Therefore, any periodic Markov chain can be broken down to aperiodic Markov chains.*

Theorem 12 (long time limit of aperiodic positive recurrent Markov chain). *Let X be a aperiodic positive recurrent Markov chain with state space S and transition matrix Π. Let μ denote the unique stationary distribution. Then for any initial distribution μ_0, X_n converges to μ in distribution, that is*

$$P_{\mu_0}(X_n = y) \to \mu(y)$$

as $n \to \infty$ for all $y \in S$

Proof. *TODO- coupling* □

Theorem 13 (long time limit of null recurrent Markov chain). *Let X be a null recurrent Markov chain with state space S, then*

$$\Pi^n(x,y) \to 0$$

as $n \to \infty$ for any $x, y \in S$. Hence, for any initial distribution μ_0,

$$P_{\mu_0}(X_n = y) \to 0$$

as $n \to \infty$ for any $y \in S$

2.6 Renewal Process

Definition 40 (discrete renewal process). *A discrete renewal process τ is a sequence of $\mathbb{N}_0$-valued random variable $(\tau_n)_{n\in\mathbb{N}_0}$ where $\tau_0 = 0$ and $(\tau_n - \tau_{n-1})_{n\in\mathbb{N}}$ are i.i.d $\mathbb{N} \cup \{\infty\}$-valued random variables with probability mass function $f(k) = P(\tau_1 = k)$ for $k \in \mathbb{N} \cup \{\infty\}$. That is, the distribution of increments is fixed.*

Remark 11. *The natural interpretation of $(\tau_n)_{n\in\mathbb{N}_0}$ is the collection of times when we change the light bulb such that light bulbs have i.i.d random lifetimes with probability mass function f*

Remark 12. *Given a Markov chain $(X_n)_{n\in\mathbb{N}_0}$ with $X_0 = x$, the sequence T_x^m for $m \in \mathbb{N}_0$ is a renewal process where $f(\infty) > 0$ if and only if x is transient.*

Remark 13 (discussion on the Markov chain of renewal process). $\Pi(n, n-1) = 1$, $\Pi(0, k-1) = f(k)$

Theorem 14 (renewal theorem). *Let τ be a discrete renewal process, if τ is transcient, that is, $f(\infty) > 0$ or null recurrent that is $f(\infty) = 0$ and $\sum_{k\in\mathbb{N}} kf(k) = \infty$, then*

$$P(n \in \tau) = P(n \in \{\tau_1, \tau_2, ...\}) \to 0$$

as $n \to \infty$. If τ is positive recurrent that is $f(\infty) = 0$ and $\sum_{k\in\mathbb{N}} kf(k) < \infty$, and τ is aperiodic, that is, $r = \gcd(n : f(n) > 0) = 1$, then

$$P(n \in \tau) = P(n \in \{\tau_1, \tau_2, ...\}) \to \frac{1}{\sum_{k\in\mathbb{N}} kf(k)}$$

as $n \to \infty$

2.7 Reversible Measure, Reversible Markov Chain

Definition 41 (reversible measure, reversible Markov chain). *Let X be a Markov chain with state space S and transition matrix Π. A measure $\nu : S \to \mathbb{R}$ is a reversible measure of X if*

$$\nu(x)\Pi(x,y) = \nu(y)\Pi(y,x)$$

for all $x, y \in S$. The condition is called detailed balance. A Markov chain is called reversible if it has a reversible measure.

Remark 14. *A reversible measure ν must be stationary, since*

$$\nu(x) = \sum_{y \in S} \nu(x)\Pi(x,y) = \sum_{y \in S} \nu(y)\Pi(y,x)$$

If we interpret a distribution as the distribution of masses over all states, then each time step, masses are transferred. Stationary means for each state, the in-mass equals the out-mass. Reversibility means for each pair of state x, y, the mass $x \to y$ equals the mass $y \to x$

Proposition 22 (time reversibility). *Let ν be a reversible distribution of a Markov chain X. If X_0 has distribution ν, then $(X_0, ..., X_n)$ has the same distribution as its time reversal $(X_n, ..., X_0)$, that is*

$$P_\nu(X_0 = x_0, ..., X_n = x_n) = P_\nu(X_0 = x_n, ..., X_n = x_0)$$

Moreover, if given a stationary measure ν, $(X_0, ..., X_n)$ has the same distribution as its time reversal $(X_n, ..., X_0)$, then ν is reversible.

Proof. TODO □

Theorem 15 (loop condition for reversibility). *An irreducible Markov chain is reversible if and only if the transition matrix Π satisfies the loop condition, that is, given $x \in S$,*

$$\frac{\Pi(x_0, x_1)}{\Pi(x_1, x_0)} \cdots \frac{\Pi(x_{n-1}, x_n)}{\Pi(x_n, x_{n-1})} = 1$$

for all path $(x = x_0, x_1, ..., x_{n-1}, x_n = x)$. In that case, we can construct a stationary measure by

$$\nu(y) = \nu(x) \frac{\Pi(y_0, y_1)}{\Pi(y_1, y_0)} \cdots \frac{\Pi(y_{n-1}, y_n)}{\Pi(y_n, y_{n-1})}$$

for a path $(y = y_0, y_1, ..., y_{n-1}, y_n = x)$

Proof. TODO □

Remark 15 (reversible Markov chain as random walk on electric network). *Any reversible MC can be seen as a random walk on a graph $G = (V, E)$ with $V = S$ and $(x, y) \in E$ if $\Pi(x, y) > 0$ with conductance $C(x, y)$ where*

$$C(x,y) = \nu(x)\Pi(x,y)$$

with $\nu : S \to \mathbb{R}$ is a stationary measure

$$\nu(x) = \sum_{(x,z) \in E} C(x,z)$$

2.8 Hitting Probability, Expected Hitting Time

Given a Markov chain X with state space S, let $A, B \subseteq S$ be two disjoint subsets of S, let $T_A = \min(n \geq 0 : X_n \in A)$, $T_B = \min(n \geq 0 : X_n \in B)$ be the first times the Markov chain visiting A and B.

2.8.1 Hitting Probability

Let

$$f(x) = P_x(T_A < T_B)$$

be the probability of hitting A before B . The boundary conditions are for every $x \in A$, $f(x) = 1$, for every $x \in B$, $f(x) = 0$. If $x \notin A \cup B$, then

$$\begin{aligned}
f(x) &= P_x(T_A < T_B) \\
&= \sum_{y \in S} P_x(X_1 = y, T_A < T_B) && \text{(one step analysis)} \\
&= \sum_{y \in S} P_x(X_1 = y) P_x(T_A < T_B | X_1 = y) \\
&= \sum_{y \in S} P_x(X_1 = y) P_y(T_A < T_B) \\
&= \sum_{y \in S} \Pi(x, y) f(y) = (\Pi f)(x)
\end{aligned}$$

Hence, $(\Pi - I)f = 0$. f is called a harmonic function of the operator Π. (*related to Laplace equation*)

2.8.2 Expected Hitting Time

Let

$$g(x) = \mathbb{E}_x[T_A]$$

be the expected hitting time for A. The boundary condition is for every $x \in A$, $g(x) = 0$. If $x \notin A$, then

$$\begin{aligned}
g(x) &= \mathbb{E}_x[T_A] \\
&= \mathbb{E}_x\left[\sum_{y \in S} T_A 1_{\{X_1 = y\}}\right] \\
&= \mathbb{E}_x\left[1 + \sum_{y \in S} (T_A - 1) 1_{\{X_1 = y\}}\right] \\
&= 1 + \sum_{y \in S} \mathbb{E}_x[(T_A - 1) 1_{\{X_1 = y\}}] \\
&= 1 + \sum_{y \in S} P_x(X_1 = y) \mathbb{E}_x[T_A - 1 | X_1 = y] \\
&= 1 + \sum_{y \in S} P_x(X_1 = y) \mathbb{E}_y[T_A] \\
&= 1 + \sum_{y \in S} \Pi(x, y) g(y) = 1 + (\Pi g)(x)
\end{aligned}$$

Hence, $(\Pi - I)g = 1$ (*related to Poisson equation*)

2.9 Monte Carlo, Metropolis, Gibbs sampling

SKIP - NOT IN EXAM

Chapter 3

Martingale

3.1 Martingale basics

Definition 42 (σ-algebra filtration)**.** *A filtration on (Ω, F, P) is an increasing sequence of σ-algebras $(F_n)_{n\geq 0}$ with*

$$F_0 \subseteq F_1 \subseteq ...$$

Remark 16. *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a stochastic process, then the filtration defined by $F_n = \sigma(X_0, X_1, ..., X_n)$ is called the canonical filtration generated by X*

Definition 43 (martingale)**.** *Given a filtration $G_0 \subseteq G_1 \subseteq G_2 \subseteq ...$, a real-valued stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ is called a martingale adapted to the filtration G. if*

1. *For all $n \in \mathbb{N}_0$, $\mathbb{E}[|X_0|] < \infty$ and $\mathbb{E}[X_n|G_n] = X_n$. That is, G_n contains all information of X_n, $\sigma(X_n) \subseteq G_n$*
2. *For all $n \in \mathbb{N}_0$, $\mathbb{E}[X_{n+1}|G_n] = X_n$. This is the notion of fair game, that is, given the past information (G_n), the expectation ($\mathbb{E}[X_{n+1} - X_n|G_n]$) of $X_{n+1} - X_n$ is zero.*

we have, $\mathbb{E}[X_{n+2}|G_{n+1}] = X_{n+1}$, then $X_n = \mathbb{E}[X_{n+1}|G_n] = \mathbb{E}[\mathbb{E}[X_{n+2}|G_{n+1}]|G_n] = \mathbb{E}[X_{n+2}|G_n]$. hence, for any $n < m$, then $\mathbb{E}[X_m|G_n] = X_n$

Remark 17 (sub-martingale, super-martingale)**.** *If we replace the second condition for martingle by $\mathbb{E}[X_{n+1}|G_n] \geq X_n$, it is called sub-martingale and $\mathbb{E}[X_{n+1}|G_n] \leq X_n$, it is called super-martingale*

3.1.1 Doob Decomposition, Doob Martingale

Given a stochastic process $(X_n)_{n\in\mathbb{N}_0}$ and let $F_n = \sigma(X_0, X_1, ..., X_n)$ be the canonical filtration generated by X. Let $D_n = X_n - X_{n-1}$. Then, let $M_0 = 0$ and

$$M_n = M_{n-1} + D_n - \mathbb{E}[D_n|F_{n-1}] = \sum_{i=1}^{n}(D_i - \mathbb{E}[D_i|F_{i-1}])$$

Proposition 23. *$(M_n)_{n\in\mathbb{N}_0}$ is a martingale*

Let $A_n = \sum_{i=1}^{n} \mathbb{E}[D_i|F_{i-1}]$, note that, A_n is not a random variable but a sequence of real numbers

Theorem 16 (Doob decomposition)**.** *Every stochastic process $(X_n)_{n\in\mathbb{N}_0}$ can be decomposed into*

$$X_n = X_0 + M_n + A_n$$

where M_n is a martingale and A_n is a sequence of real numbers.

Proposition 24 (Doob martingale, martingale decomposition)**.** *If Y is a random variable and $G_0 \subseteq G_1 \subseteq G_2 \subseteq ...$ is a filtration in (Ω, F, P), then $Z_n = \mathbb{E}[Y|G_n]$ is a martingale. This is a direct application of tower property*

$$\mathbb{E}[Z_n|G_{n-1}] = \mathbb{E}[\mathbb{E}[Y|G_n]|G_{n-1}] = \mathbb{E}[Y|G_{n-1}] = Z_{n-1}$$

If $G_n = F$, then $Z_0 = \mathbb{E}[Y]$ and $Z_n = Y$, martingale decomposition

$$Y = \mathbb{E}[Y] + (Z_n - Z_0) = \mathbb{E}[Y] = \sum_{i=1}^{n}(Z_i - Z_{i-1})$$

TODO - generalize this

3.1.2 Martingale in Markov chain

Let X be a Markov chain with state space S and transition matrix Π. Let $f : S \to \mathbb{R}$ be a bounded function on S, then Doob decomposition gives

$$f(X_n) = f(X_0) + M_n + A_n$$

where M_n is a martingale adapted to the canonical filtration of X and

$$\begin{aligned}
A_n &= \sum_{i=1}^{n} \mathbb{E}[f(X_i) - f(X_{i-1})|F_{i-1}] \\
&= \sum_{i=1}^{n} \mathbb{E}[f(X_i) - f(X_{i-1})|X_0, X_1, ..., X_{i-1}] \\
&= \sum_{i=1}^{n} \mathbb{E}[f(X_i) - f(X_{i-1})|X_{i-1}] \\
&= \sum_{i=1}^{n} \mathbb{E}[f(X_i)|X_{i-1}] - f(X_{i-1}) \\
&= \sum_{i=1}^{n} (\Pi - I) f(X_{i-1})
\end{aligned}$$

Hence, if $(\Pi - I)f = 0$ (f is harmonic) then $f(X_n)$ is a martingale, if $(\Pi - I)f = -1$, then $f(X_n) + n$ is a martingale. Now let $A, B \subseteq S$ be disjoint and $T_A = \{n \geq 0 : X_n \in A\}$ be the first time hitting A, we want to compute

$$f(x) = P_x(T_A < T_B) \text{ and } g(x) = \mathbb{E}_x[T_A]$$

Through one step analysis, we have shown that

$$(\Pi - I)f = 0 \text{ and } (\Pi - I)g = 1$$

Then, $f(X_n)$ is a martingale before time $T_A \wedge T_B = \min\{T_A, T_B\}$, $g(X_n) + n$ is a martingale before time T_A

3.2 Azuma-Hoeffding Inequality

Theorem 17 (Azuma-Hoeffding). *Let $(X_n)_{0 \leq n \leq N}$ be a martingale with X_0 and its increments $D_i = X_i - X_{i-1}$ satisfy $|D_i| \leq K$ for all $1 \leq i \leq N$ almost surely (true for a set $\Omega_0 \subseteq \Omega$ of realizations with $P(\Omega_0) = 1$). Then, for all $a > 0$,*

$$P\left(\frac{X_N}{\sqrt{N}} \geq +a\right) \leq e^{-\frac{a^2}{2K}} \text{ and } P\left(\frac{X_N}{\sqrt{N}} \leq -a\right) \leq e^{-\frac{a^2}{2K}}$$

Proof. *TODO* □

TODO - generalize to the case where $|D_i| \leq K_i$

3.3 Stopped Martingale

Definition 44 (stopping time). *A random variable τ on $\mathbb{N}_0 \cup \{\infty\}$ is called stopping time with respect to the filtration $(F_n)_{n \in \mathbb{N}_0}$ if $\{\tau = n\} \in F_n$ for all $n \geq 0$ (I am kinda get it but not really get it. at least I don't do probability so just know enough to pass the exam). τ models the stopping time, that is, to decide when to stop a martingale, we only have the information available up to that time. (sub-σ-algebra is information)*

Proposition 25. *If τ_1 and τ_2 are stopping time with respect to the filtration $(F_n)_{n \in \mathbb{N}_0}$, then $\tau_1 \wedge \tau_2 = \min\{\tau_1, \tau_2\}$ and $\tau_1 \vee \tau_2 = \max\{\tau_1, \tau_2\}$ are stopping times.*

Definition 45 (stopped σ-field [1]). *Let τ be a stopping time with respect to the filtration $(F_n)_{n \geq 0}$, the stopped σ-field F_τ associated with the stopping time τ is defined by*

$$F_\tau = \{A \in F : A \cap \{\tau = n\} \in F_n \text{ for all } n \geq 0\}$$

that is, the collection of measurable events A in which we can determine whether it will occur or not based on the available information up time time τ

[1] σ-algebra is also called σ-field

Lemma 2 (stopped martingale is a martingale). *Let $(X_n)_{n\geq 0}$ be a martingale adapted to a filtration $(F_n)_{n\geq 0}$ and τ be a stopping time with respect to $(F_n)_{n\geq 0}$. Then $Y_n = X_{n\wedge\tau}$, the martingale X_n stopped at time τ, is also a martingale with respect to $(F_n)_{n\geq 0}$. More generally, if θ is another stopping time with $\theta \leq \tau$ almost surely, then $X_{n\wedge\tau} - X_{n\wedge\theta}$ is also a martingale.*

Proof. TODO □

3.3.1 Upcrossing Inequality, Martingale Convergence Theorem, Backward Martingale

Definition 46 (upcrossing). *Let $(X_n)_{n\geq 0}$ be a super-martingale adapted to the filtration $(F_n)_{n\geq 0}$. An upcrossing by X of the interval (a,b) with $a < b$ consists of a pair of times $k < l$ with $X_k \leq a$ and $X_l \geq b$. Let U_n be the number of complete upcrossings X makes before (before and at) time n and define*

$$\begin{aligned}
\tau_1 &= \min\{i \geq 0 : X_i \leq a\}\\
\tau_2 &= \min\{i \geq \tau_1 : X_i \geq b\}\\
&\dots\\
\tau_{2k+1} &= \min\{i \geq \tau_{2k} : X_i \leq a\}\\
\tau_{2k+2} &= \min\{i \geq \tau_{2k+1} : X_i \geq b\}
\end{aligned}$$

where the minimum of an empty set is taken to be ∞. Note that, τ_i is a stopping time and $U_n = \max\{k : \tau_{2k} \leq n\}$

Lemma 3 (upcrossing inequality). *Let $(X_n)_{n\geq 0}$ be a super-martingale and U_n be the number of complete upcrossings over (a,b) before time n, then*

$$\mathbb{E}[U_n] \leq \frac{\mathbb{E}[(a - X_n)^+]}{b-a} \leq \frac{|a| + \mathbb{E}[|X_n^-|]}{b-a}$$

where $x^+ = \max\{x, 0\}$, $x^- = \min\{x, 0\}$

Proof. TODO □

Theorem 18 (martingale convergence theorem). *If $(X_n)_{n\geq 0}$ is a super-martingale and $\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n^-|] < \infty$, then there exists a random variable X_∞ such that almost surely $X_n \to X_\infty$ as $n \to \infty$ and $\mathbb{E}[|X_\infty|] < \infty$. For sub-martingale, the condition is $\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n^+|] < \infty$.*

Proof. TODO □

Corollary 2. *If $(X_n)_{n\geq 0}$ is a non-negative super-martingale then $X_\infty = \lim_{n\to\infty} X_n$ exists almost surely and $\mathbb{E}[X_\infty] \leq \mathbb{E}[X_0]$*

Corollary 3. *Let $(X_n)_{n\geq 0}$ be a martingale with $|X_{n+1} - X_n| \leq M < \infty$ almost surely for all $n \geq 0$, then almost surely either $\lim_{n\to\infty} X_n$ exists and finite or $\limsup_{n\to\infty} X_n = +\infty$ and $\liminf_{n\to\infty} X_n = -\infty$. That is, either X_n converges or osscilates between $-\infty$ and $+\infty$*

Definition 47 (backward martingale). *$(X_n)_{n\geq 0}$ is called a backward martingale adapted to the decreasing filtration $F_0 \supseteq F_1 \supseteq \dots$ if*

$$\mathbb{E}[X_n | F_{n+1}] = X_{n+1}$$

Note that, $X_n = \mathbb{E}[X_0|F_n]$ for all $n \geq 0$, and $(\dots, X_2, X_1, X_0)$ is a martingale adapted to the filtration $\dots \subseteq F_2 \subseteq F_1 \subseteq F_0$

Theorem 19. *Let $(X_n)_{n\geq 0}$ be a backward martingale adapted to a decreasing filtration $(F_n)_{n\geq 0}$, then almost surely $X_n \to X_\infty$ and $\mathbb{E}[X_\infty] = \mathbb{E}[X_0]$*

Lemma 4 (when does martingale limit preserve mean). *If for some $K > 0$, the martingale $(X_n)_{n\geq 0}$ is bounded, that is, $P(|X_n| \leq K) = 1$ for all large n, then almost surely $X_n \to X_\infty$ and $\mathbb{E}[X_\infty] = \mathbb{E}[X_0]$*

Proof. TODO □

3.4 Uniform Integrable Martingale, Optional Stopping Theorem

Definition 48 (uniform integrability). *A sequence of random variables $(X_n)_{n\geq 0}$ is called uniformly integrable if for each $\epsilon > 0$, there exists $K > 0$ such that*

$$\sup_{n\geq 0} \mathbb{E}[|X_n| 1_{|X_n|>K}] \leq \epsilon$$

Remark 18 (L^p $(p > 1)$ implies uniformly integrable). *If $\sup_n \mathbb{E}[|X_n|^p] < \infty$ for some $p > 1$, then Markov inequality implies that $(X_n)_{n\geq 0}$ is uniformly integrable.*

Theorem 20. *Let $(X_n)_{n\geq 0}$ be a martingale that is uniformly integrable, then almost surely $X_n \to X_\infty$ and $\mathbb{E}[X_\infty] = \mathbb{E}[X_0]$*

Proof. TODO □

Theorem 21 (optional stopping theorem). *Let $(X_n)_{n\geq 0}$ be a martingale and τ a finite stopping time adapted to the same filtration $(F_n)_{n\geq 0}$. If the sequence $(X_{n\wedge\tau})_{n\geq 0}$ is uniformly integrable, then $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ $(X_\infty = X_\tau)$*

Remark 19. *Doob martingale is uniformly integrable. If X_n is a martingale with $\mathbb{E}[|X_n|] < \infty$, then for any convex function ϕ, $\phi(X_n)$ is a sub-martingale.*

3.5 Doob Maximal Inequality

Theorem 22 (Doob maximal inequality). *Let $(X_i)_{i\in\mathbb{N}}$ be a sub-martingale with respect to filtration $(F_i)_{i\in\mathbb{N}}$. Let $S_n = \max_{1\leq i\leq n} X_i$ be the running maximum of X_i, then for any $l > 0$,*

$$P(S_n \geq l) \leq \frac{1}{l}\mathbb{E}[X_n^+ 1_{\{S_n\geq l\}}] = \frac{1}{l}\mathbb{E}[X_n^+]$$

where $X_n^+ = X_n \vee 0 = \max\{X_n, 0\}$. In particular, if $(X_i)_{i\in\mathbb{N}}$ is a martingale and the absolute value function is convex, then $|X_i|$ is a sub-martingale, then let $M_n = \max_{1\leq i\leq n} |X_i|$,

$$P(M_n \leq l) \leq \frac{1}{l}\mathbb{E}[|X_n| 1_{\{M_n\leq l\}} \leq \frac{1}{l}\mathbb{E}[|X_n|]$$

Corollary 4. *For any $p > 1$, $x \mapsto (x^+)^p$ and $x \mapsto |x|^p$ are convex functions, then*

$$P(S_n \leq l) \leq \frac{1}{l^p}\mathbb{E}[(X_n^+)^p 1_{\{S_n\geq l\}}] \leq \frac{1}{l^p}\mathbb{E}[(X_n^+)^p]$$
$$P(M_n \leq l) \leq \frac{1}{l^p}\mathbb{E}[|X_n|^1_{\{S_n\geq l\}}] \leq \frac{1}{l^p}\mathbb{E}[|X_n|^p]$$

where X_n being sub-martingale and martingale correspondingly.

Theorem 23 (Doob L^p maximal inequality). *For any $p > 1$,*

$$\mathbb{E}[(S_n^+)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[(X_n^+)^p]$$
$$\mathbb{E}[M_n^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p]$$

3.6 Square-Integrable Martingale and Quadratic Variation

Definition 49 (square-integrable martingale, quadratic variation process). *A martingale $(X_n)_{n\geq 0}$ is called square-integrable if $\mathbb{E}[X_n^2] < \infty$. If $(X_n)_{n\geq 0}$ is a square-integrable martingale, then X_n^2 is a sub-martingale with Doob decomposition*

$$X_n^2 = M_n + \langle X\rangle_n$$

where M_n is a martingale and $\langle X\rangle_n$ is a monotone increasing sequence

$$\langle X\rangle_n = \sum_{n=2}^{n} \mathbb{E}[(X_i - X_{i-1})^2 | F_{i-1}]$$

$\langle X\rangle_n$ is called quadratic variation process of X

Theorem 24. *Let $(X_n)_{n\in\mathbb{N}}$ be a square-integrable martingale and $\langle X\rangle_n$ its quadratic variation process. Then*

1. *on the event $\{\langle X\rangle_\infty < \infty\}$, almost surely $\lim_{n\to\infty} X_n$ exists and finite.*
2. *on the event $\{\langle X\rangle_\infty = \infty\}$, almost surely $\lim_{n\to\infty} \frac{X_n}{\langle X\rangle_n} = 0$*

3.7 Martingale from Change of Measure

SKIP - I JUST DON'T WANT TO DO THIS