Notes in Advanced Stochastic Models

Liu Chuan

May 4, 2007

Contents

Chapter 1

Preliminaries

1.1 Probability

Probability space $(\Omega, \mathcal{F}, P)$. Ω is sample space, set of all possible outcomes of a random experiment. Event E is a subset of Ω, said to occur if the outcome of the experiment is an element of E. $\mathcal{F}$ is a collection of events, and $\mathcal{F}$ is σ-field.

Remark. $\mathcal{F}$ is called a σ-field if the following holds.

(i) $\Omega \in \mathcal{F}$.

(ii) if $A, B \in \mathcal{F}$, then $A \backslash B \in \mathcal{F}$.

(iii) if $A_i \in \mathcal{F},\ i = 1, 2, \cdots$, then $\bigcup_{i=1}^{\infty} \in \mathcal{F}$.

Note. $\boxed{\mathcal{F} = \{\Omega, \phi\} \text{ is smallest } \sigma\text{-field.}}$

For each $E \in \mathcal{F}$, a number $P(E)$ is defined ($P : \mathcal{F} \to \mathbb{R}$ mapping), satisfying the following.

(i) $0 \leq P(E) \leq 1$.

(ii) $P(\Omega) = 1$.

(iii) For any sequences $\{E_i\}_{i=1}^{\infty} \subset \mathcal{F}$, which are mutually exclusive,

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Note. $P(A) = 1 \Leftrightarrow A$ holds almost surely (a.s.).
$P(A \cap B) = P(A)P(B) \Leftrightarrow A$ and B are independent.

Simple facts

1. If $E \subseteq F$, then $P(E) \leq P(F)$.
2. $P(E^c) = P(\Omega \backslash E) = 1 - P(E)$.
3. If $\{E_i\}_{i=1}^n$ are mutually exclusive,
$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$
4. $P(\bigcup_{i=1}^\infty) \leq \sum_{i=1}^\infty P(E_i)$ (Boole's Inequality)

$\{E_n\}_{n=1}^\infty$ is increasing if $E_n \subseteq E_{n+1}$ $\forall n$.
decreasing if $E_n \supseteq E_{n+1}$ $\forall n$.

Let $\{E_n\}_{n=1}^\infty$ be increasing. Define $\lim_{n\to\infty} E_n \triangleq \bigcup_{n=1}^\infty E_n$.
decreasing. Define $\lim_{n\to\infty} E_n \triangleq \bigcap_{n=1}^\infty E_n$.

Proposition 1.1.1. *If $\{E_n\}_{n=1}^\infty$ is either increasing or decreasing, then*
$$P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n)$$

Proof. Suppose $\{E_n\}_{n=1}^\infty$ is increasing.

Define $F_1 = E_1$
$F_n = E_n \backslash E_{n-1} = E_n \cap E_{n-1}^c$, $n = 2, 3, \cdots$

$\{F_n\}_{n=1}^\infty$ are mutually exclusive, and
$$\bigcup_{i=1}^n F_i = \bigcup_{i=1}^n E_i\ ,\ n = 1, 2, \cdots$$

$$\begin{aligned} P\left(\lim_{n\to\infty} E_n\right) &= P\left(\bigcup_{i=1}^\infty E_i\right) = P\left(\bigcup_{i=1}^\infty F_i\right) = \sum_{i=1}^\infty P(F_i) \\ &= \lim_{n\to\infty} \sum_{r=1}^n P(F_i) = \lim_{n\to\infty} P(\bigcup_{i=1}^n F_i) \\ &= \lim_{n\to\infty} P(E_n) \end{aligned}$$

$\square$

Example 1.1.1. Consider a population consisting of individuals able to produce offspring of the same kind. The number of individuals initially present, denoted by X_0, is called the size of the zeroth generation. All offspring of the zeroth generation constitute the first generation and their number is denoted by X_1. In general, let X_n denote the size of the nth generation.

Let $E_n = \{X_n = 0\}$, $n = 0, 1, 2, \cdots$. $E_n \uparrow$ or $E_n \subseteq E_{n+1}$.

If $\lim_{n\to\infty} P(E_n)$ exists,

$$\begin{aligned}
\lim_{n\to\infty} P(E_n) &= P\left(\lim_{n\to\infty} E_n\right) \\
&= P\left(\bigcup_{n=1}^{\infty} \{X_n = 0\}\right) \\
&= P(\text{the population ever dies out})
\end{aligned}$$

Theorem 1.1.1. *(Borel-Cantelli lemma) Let $\{E_n\}_{n=1}^{\infty}$ denote a sequence of events. If $\sum_{i=1}^{\infty} P(E_i) < \infty$, then*

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} E_i\right) = 0$$

Proof. Let $F_n \triangleq \bigcup_{i=n}^{\infty} E_i$. $F_n \downarrow$.

$$\begin{aligned}
P\left(\bigcap_{n=1}^{\infty} F_n\right) &= P\left(\lim_{n\to\infty} F_n\right) = \lim_{n\to\infty} P(F_n) \\
&= \lim_{n\to\infty} P\left(\bigcup_{i=n}^{\infty} E_i\right) \leq \lim_{n\to\infty} \sum_{i=n}^{\infty} P(E_i) = 0
\end{aligned}$$

□

Note. $\boxed{\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} E_i = \limsup_{i\to\infty} E_i = \overline{\lim_{i\to\infty}}\, E_i}$

Remark. $\boxed{\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} E_i = \{\text{an infinite number of } E_i \text{ occurs}\}}$

1.2 Random Variable

A random variable (r.v.) X is a mapping from Ω to $\mathbb{R}$, satisfying $\{X \leq a\} \in \mathcal{F}$, $\forall a \in \mathbb{R}$. A distribution function F of the r.v. X, $F(x) \triangleq P(X \leq x)$, $\forall x \in \mathbb{R}$, and $\overline{F}(x) \triangleq 1 - F(X) = P(X > x)$.

A r.v. X is said to be discrete if its set of possible value is countable. In this case

$$F(x) = \sum_{y \leq x} P(X = y) \ \forall x \in \mathbb{R}$$

A r.v. X is said to be continuous if there is a function $f(x)$ called the probability density function, such that

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx \ \forall a \in \mathbb{R}$$

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

The joint distribution function F of two r.v.s X and Y is

$$F(x, y) = P(X \leq x, Y \leq y) \ \forall (x, y) \in \mathbb{R}^2$$

Let $Y_n \uparrow \infty$, $E_n \triangleq \{X \leq x, Y \leq y_n\} \uparrow$.

$$\{X \leq x\} = \bigcup_{n=1}^{\infty} (X \leq x, Y \leq y_n)$$

$$F_X(x) = P(X \leq x) = P(\bigcup_{n=1}^{\infty} E_n) = \lim_{n \to \infty} P(X \leq x, Y \leq y_n) = \lim_{n \to \infty} F(x, y_n)$$

Two r.v.s X and Y are called independent if

$$F(x, y) = F_X(x)F_Y(y)$$

X and Y are called jointly continuous if there exits a function $f(x, y)$ called the joint probability density function, such that,

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)du\, dv \ \forall (x, y) \in \mathbb{R}^2$$

1.3 Mathematic Expectation

The mathematical expectation or mean of a r.v. X, $E[X]$, is defined by

$$E[X] \triangleq \int_{-\infty}^{\infty} xdF(x) = \begin{cases} \sum_x xP(X = x) & \text{if } X \text{ is discrete.} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{if } X \text{ is continuous.} \end{cases}$$

Let $h : \mathbb{R} \to \mathbb{R}$. The (measurable) function $h(X)$ is r.v.

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)dF(x) = \begin{cases} \sum_x h(x)P(X = x) & \text{if } X \text{ is discrete.} \\ \int_{-\infty}^{\infty} h(x)f(x)dx & \text{if } X \text{ is continuous.} \end{cases}$$

$h : \mathbb{R}^n \to \mathbb{R}$

$$E[h(X_1, X_2, \cdots, X_n)] = \int_{\mathbb{R}^n} h(x_1, x_2, \cdots, x_n) dF(x_1, x_2, \cdots, x_n)$$

The variance of r.v. X

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

Standard deviation $\sigma_X = \sqrt{\mathrm{Var}(X)}$.
The covariance of two random variable X and Y

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

X and Y are called uncorrelated if $\mathrm{cov}(X, Y) = 0$ or $E[XY] = E[X]E[Y]$.

Properties of expectation and variance

1. $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]\ \forall \alpha, \beta \in \mathbb{R}$.

2. $$\begin{aligned} \mathrm{Var}(\sum_{i=1}^{n} X_i) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{cov}(X_i, X_j) \end{aligned}$$

3. $E[XY] = E[X]E[Y]$, if X and Y are independent.

1.4 Moment Generating Function

The moment generating function of a r.v. X

$$\psi_X(t) \equiv \psi(t) \triangleq E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} dF(x),\ t \in \mathbb{R}$$

$$\begin{aligned} \psi'(t) &= \int_{-\infty}^{\infty} x e^{tx} dF(x) = E[X e^{tX}] && \psi'(0) = E[X] \\ \psi''(t) &= \int_{-\infty}^{\infty} x^2 e^{tx} dF(x) = E[X^2 e^{tX}] && \psi''(0) = E[X^2] \end{aligned}$$

Remark. $\boxed{\psi^{(n)}(0) = E[X^n],\ n \geq 1}$

The moment generating function of r.v.s $X_1, X_2, \cdots, X_n$

$$\psi(t_1, t_2, \cdots, t_n) \triangleq E\left[e^{\sum_{t=1}^{n} t_i X_i}\right]$$

Example 1.4.1. $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$.
X_1 and X_2 are independent.

$$\begin{aligned}\psi_{X_1+X_2}(t) &= E[e^{t(X_1+X_2)}] = E[e^{tX_1}e^{tX_2}] = E[e^{tX_1}]E[e^{tX_2}]\\ &= \psi_{X_1}(t)\psi_{X_2}(t) = e^{(\mu_1+\mu_2)t} + \frac{\sigma_1^2+\sigma_1^2}{2}t^2\end{aligned}$$

$\Rightarrow X_1 + X_2 \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Define the characteristic function of X,

$$\phi_X(t) = \phi(t) = E[e^{itx}] = \int_{-\infty}^{\infty} e^{itx} dF(x)$$

where $i = \sqrt{-1}$.

Note. $\boxed{e^{i\theta} = \cos\theta + i\sin\theta}$

Remark. $\boxed{\phi \text{ always exists and uniquely determines the distribution of } X.}$

The joint characteristic function of $X_1, X_2, \cdots, X_n$,

$$\phi(t_1, t_2, \cdots, t_n) = E[e^{i\sum_{i=1}^{n} t_n X_n}]$$

1.5 Conditional Expectation

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad \forall E_1, E_2 \in \mathcal{F} \text{ with } P(E_2) > 0$$

Remark. $\boxed{\text{If } P(E_1 \cap E_2) = P(E_1)P(E_2), \text{ then } P(E_1|E_2) = P(E_1).}$

Let X and Y be two discrete r.v.s, the conditional probability mass function of X given Y is defined to be,

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

provided that $P(Y = y) > 0$.

The continuous distribution function of X given $Y = y$ is defined to be,

$$F(x|y) = \sum_{z \le x} P(X = z|Y = y)$$

The conditional expectation of X given Y is

$$E(x|Y = y) = \int x dF(x|y) = \sum_{x} P(X = x|Y = y)$$

Let X and Y be jointly continuous r.v.s having joint probability density function $f(x, y)$. The conditional density function of X given $Y = y$ is defined for all y such that $f_Y(y) > 0$ by

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

$$P(E|Y = y) = \int_E f(x|y)dx$$

The conditional distribution function of x given $Y = y$ is

$$F(x|y) = \int_{-\infty}^{x} f(z|y)dz$$

The conditional expectation of X given $Y = y$ is,

$$E(X|Y = y) = \int_{-\infty}^{\infty} xf(x|y)dx$$

Define the following r.v denoted by $E[X|Y]$.

$$E[X|Y] = E[X|Y = y] \text{ if } Y = y$$

Theorem 1.5.1. *For all r.v.s X and Y,*

$$E[X] = E[E(X|Y)] = \int_{-\infty}^{\infty} E(x|Y = y)dF_Y(y)$$

Corollary 1.5.1. *1. If Y is discrete, then*

$$E[X] = \sum_y E(X|Y = y)P(Y = y)$$

2. If Y is continuous, then

$$E[X] = \int_{-\infty}^{\infty} E(X|Y = y)f_Y(y)dy$$

Proof of the case when both X and Y are discrete.

$$\begin{aligned}
\sum_y E(X|Y = y) &= \sum_y \sum_x xP(X = x|Y = y)P(Y = y) \\
&= \sum_y \sum_x xP(x = x, Y = y) \\
&= \sum_x x \left[\sum_y P(X = x, Y = y) \right] \\
&= \sum_x xP(X = x) \\
&= E[X]
\end{aligned}$$

□

Corollary 1.5.2. *1.*

$$E\left[\sum_{i=1}^{n} X_i \middle| Y = y\right] = \sum_{i=1}^{n} E[X_i|Y = y]$$

2.

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

Corollary 1.5.3. *Let $A \in \mathcal{F}$ and Y be a r.v., then*

$$P(A) = \int_{-\infty}^{\infty} P(A|Y = y)dF_Y(y)$$

Proof. Define a r.v. X by

$$X = \begin{cases} 1 & \text{if } A \text{ occur} \\ 0 & \text{otherwise} \end{cases}$$

Note. $\boxed{X = \mathbf{1}_A = \chi_A \quad \text{(indicator function)}}$

$$\begin{aligned} P(A) &= E[X] \\ P(A|Y = y) &= E[X|Y = y] \end{aligned}$$

□

Note. $\boxed{\begin{array}{l} P^y(A) = P(A|Y = y) \\ E^y(X) = E(X|Y = y) \end{array}}$

Theorem 1.5.2. *For all r.v.s X, Y, W, we have*

$$\begin{aligned} E[X|W] &= E[E(X|W, Y)|W] \\ &= E[E(X|W)|W, Y] \end{aligned}$$

Example 1.5.1. A miner is trapped in a mine containing three doors. The first door leads to a tunnel that takes him to safety after two hours of travel. The second door leads to a tunnel that returns him to the mine after three hours of travel. The third door leads to a tunnel that returns him to the mine after five hours. Assuming that the miner is at all times equally likely to choose any one of the doors. What are the expected time and its variance when the miner reaches safety?

X : time when the miner reaches safety.
Y : # of door be chosen.

$$
\begin{aligned}
E[e^{tx}] &= \sum_{i=1}^{3} E[e^{tx}|Y=i]P(Y=i) \\
E[e^{tx}|Y=1] &= e^{2t} \\
E[e^{tx}|Y=2] &= E[e^{t(3+x')}] = e^{3t}E[e^{tx}] \\
E[e^{tx}|Y=3] &= e^{5t}E[e^{tx}] \\
E[e^{tx}] &= \frac{1}{3}(e^{2t} + e^{3t}E[e^{tx}] + e^{5t}E[e^{tx}])) \\
E[e^{tx}] &= \frac{E^{2t}}{3 - e^{3t} - e^{5t}} = \psi_X(t) \\
E[X] &= \psi'(0) = 10 \text{ (hrs)} \\
\mathrm{Var}[x] &= E[X^2] - (Ex)^2 = \psi''(0) - 100
\end{aligned}
$$

1.6 Exponential Distribution

A continuous r.v. $X \geq 0$ is called to have an exponential distribution with parameter $\lambda > 0$ if its density function is

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

or the distribution function is

$$
F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}
$$

$$
\begin{aligned}
\psi(t) &= E[e^{tx}] = \int_0^\infty e^{tx}\lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t} \\
E[x] &= \psi'(0) = \frac{1}{\lambda} \qquad\qquad \mathrm{Var}[x] = \frac{1}{\lambda^2}
\end{aligned}
$$

Recall

$$
\begin{aligned}
\overline{F}(t) = 1 - F(t) = P(X > t) = e^{-\lambda t} \\
\overline{F}(t+s) = \overline{F}(t)\overline{F}(s) \ \forall t, s \geq 0 \\
P(X > t+s) = P(X > t)P(X > s)
\end{aligned}
$$

$$
\begin{aligned}
\Leftrightarrow P(X > s) &= \frac{P(X > t+s)}{P(X > t)} \\
&= \frac{P(X > t+s, X > t)}{P(X > t)} \\
&= P(X > t+s | X > t) \ \ \forall t, s > 0 \qquad \text{(ageless or memoryless)}
\end{aligned}
$$

Example 1.6.1. Consider a bank having two tellers, and suppose that customer A enters the bank he discovers that B is being served by one of the tellers and C by the other. Suppose that A will be served as soon as either B or C leaves. If the amount of time a teller spends with a customer is exponentially distributed with mean $1/\lambda$, what is the probability that, of the three customers, A is the last to leave the bank?

If B and C are independent, the probability is 1/2.

Theorem 1.6.1. *The exponential distribution is the only memoryless nonnegative r.v. whose distribution function is right continuous.*

Proof. Let $g(t) : P(X > t)$. Then $g(t + s) = g(t)g(s)\ \forall t, s \geq 0$

$$g(\frac{m}{n}) = \underbrace{g(\frac{1}{n})g(\frac{1}{n})\cdots g(\frac{1}{n})}_{m} = [g(\frac{1}{n})]^m = g(1)^{\frac{m}{n}}$$

$$\begin{aligned} g(x) &= g(1)^x\ \forall \text{ rational number } x \\ &\Downarrow \\ g(x) &= g(1)^x\ \forall x \in \mathbb{R}^+ \\ &= e^{x \ln g(1)} \end{aligned}$$

$P(X \leq x) = 1 - e^{-\lambda x}$ where $\lambda = -\ln g(1)$. □

Consider a continuous r.v. $X \geq 0$ having density f and distribution $F = 1 - \overline{F}$. The failure (or hazard) rate function,

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} = \frac{-\overline{F}'t}{\overline{F}(t)}$$

Interpretation.

$$P(x \in (t, t + dt)|x > t) = \frac{P(x \in (t, t + dt))}{P(X > t)} = \frac{f(t)dt}{\overline{F}(t)} = \lambda(t)dt$$

$\lambda(t)$: the probability intensity that a t-year old item will fail.
For an exponential r.v.

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

$$\begin{aligned} \lambda(t) &= -\frac{d}{dt} \ln \overline{F}(t) \\ \ln \overline{F}(t) - \ln \overline{F}(0) &= -\int_0^t \lambda(s)ds \\ \overline{F}(0) &= P(X \geq 0) = 1 \\ \Rightarrow \overline{F}(t) &= e^{-\int_0^t \lambda(s)ds} \\ \Rightarrow F(t) &= 1 - e^{-\int_0^t \lambda(s)ds} \end{aligned}$$

1.7 Some Important Inequalities

Lemma 1.7.1. *(Markov's inequality) If $X \geq 0$ is r.v., then $\forall a > 0$*

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

Proof.

$$\begin{aligned} a\,\mathbf{1}_{(x \geq a)} &\leq x && \text{(a.s.)} \\ E[a\,\mathbf{1}_{(x \geq a)}] &\leq E[X] \\ aP(x \geq a) &\leq E[x] \end{aligned}$$

□

Remark. $\boxed{P(|x| \geq a) \leq \frac{E[|X|]}{a}}$

Proposition 1.7.1. *(Chemoff bound) Let X be a r.v. with $M(t) = E[e^{tx}]$. Then $\forall a > 0$,*

$$\begin{aligned} P(X \geq a) &\leq e^{-at}M(t) && \forall t > 0 \\ P(X \leq a) &\leq e^{-at}M(t) && \forall t < 0 \end{aligned}$$

Proof of the case $t > 0$.

$$P(X \geq a) = P(e^{tx} \geq e^{ta}) \leq \frac{E[e^{tx}]}{e^{ta}} = e^{-ta}M(t)$$

□

Example 1.7.1. Let X be such that $P(X = x) = e^{-\lambda}\frac{\lambda^x}{\lambda!}$, $x = 0, 1, 2, \cdots$, $M(t) = e^{\lambda(e^t - 1)}$. Take $j > \lambda > 0$.

$$P(X \geq j) \leq e^{-jt}e^{\lambda(e^t-1)} \; \forall t > 0$$

Take $g(t) = \lambda e^t - jt$, $g'(t) = \lambda e^t - j = 0$, $g''(t) = \lambda e^t > 0 \Rightarrow t^* = \ln \frac{i}{\lambda}$.

$$P(X \geq j) \leq \exp(-j\ln(j/\lambda) + \lambda(e^{\ln(j/\lambda)} - 1)) = \frac{e^{-\lambda}(\lambda e)^j}{j^j}$$

Proposition 1.7.2. *(Jensen's Inequality) If f is a convex function, then*

$$E[f(X)] \geq f(E[X])$$

provided the expectation exists.

Proof. Assume f is differentiable. Take $\mu = E[X]$

$$\begin{aligned} f(X) &\geq f(\mu) + f'(\mu)(X - \mu) \\ E[f(X)] &\geq f(\mu) + f'(\mu)(E[X] - \mu) \\ &= f(\mu) = f(E[X]) \end{aligned}$$

□

Note. $$\boxed{\begin{gathered} f(x) = x^2 \\ E[X]^2 \geq (E[X])^2 \end{gathered}}$$

$$\boxed{\begin{gathered} f(x) = x^p, p \geq 1, x \geq 0 \\ E[|X|^p] \geq (E|X|)^p \\ E|X| \leq (E|X|^p)^{1/p} \; \forall p \geq 1 \end{gathered}}$$

Proposition 1.7.3. *(Hölder's Inequality)*

$$E|XY| \leq (E|X|^p)^{\frac{1}{p}} (E|Y|^q)^{\frac{1}{q}}$$

$\forall p, q > 1$ *with* $\frac{1}{p} + \frac{1}{q} = 1$

Proof. First we have

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad \forall a, b \geq 0$$

Take $f(x) = -\ln x$, $x > 0$, convex.

$$-\ln(\frac{1}{p}a^p + \frac{1}{q}b^q) \leq -\frac{1}{p}\ln a^p - \frac{1}{q}\ln b^q = -\ln(ab)$$

Take

$$\begin{aligned} U &= \frac{|X|}{(E|X|^p)^{\frac{1}{p}}} \\ V &= \frac{|Y|}{(E|Y|^q)^{\frac{1}{q}}} \\ UV &\leq \frac{U^p}{p} + \frac{V^q}{q} = \frac{|X|^p}{(E|X|^p)p} + \frac{|Y|^q}{(E|Y|^q)p} \\ E|UV| &\leq \frac{1}{p} + \frac{1}{q} = 1 \end{aligned}$$

□

Remark. $$\boxed{\begin{array}{l} p = q = 2, \\ |E[XY]| \leq E|XY| \leq (E|X|^2)^{\frac{1}{2}} (E|Y|^2)^{\frac{1}{2}} \\ \hfill \text{(Cauchy-Schwarz's Inequality)} \end{array}}$$

1.8 Stochastic Process

A stochastic process $X = \{X(t), t \in T\}$ is a collection of random variables.

T: index set. $t \in T$: time. $X(t)$: state of the process at time t.

X is discrete-time if T is countable.

X is continuous-time if T is continuum.

$\omega \in \Omega$

$X(t, \omega)$: mapping from $T \times \Omega \to \mathbb{R}$.

Any realization of X is called a sample path.

$$X(t, \omega_0) : T \to \mathbb{R}$$

X is called to have independent increments if for all $t_0 < t_1 < t_2 < \cdots < t_n$, the r.v.s, $X(t_1) - X(t_0)$, $X(t_2) - X(t_1)$, ..., $X(t_n) - X(t_{n-1})$ are mutually independent. It is said to have stationary increments if $X(t+s) - X(t)$ has the same distribution for all t.

Chapter 2

Poisson Process

2.1 Definition

A stochastic process $\{N(t), t \geq 0\}$, said to be a counting process if it satisfies

1. $N(t) \geq 0$.

2. $N(t)$ is integer valued.

3. $N(s) \leq N(t)$ if $s \leq t$.

4. For $s < t$, $N(t) - N(s)$ equals the number of events that have occurred in the interval $(s, t]$.

Definition 2.1.1. A counting process $N(t), t \geq 0$ is called a Poisson process having rate $\lambda > 0$, if

(i) $N(0) = 0$.

(ii) $N(t)$ has independent increments.

(iii) $\forall s, t > 0$,

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t}\,\frac{(\lambda t)^n}{n!}$$

where $n = 0, 1, 2, \cdots$

Theorem 2.1.1. *Definition 2.1.1 is equivalent to the following.*

(i') $N(0) = 0$.

(ii') $N(t)$ *has stationary and independent increments.*

(iii') $P(N(h) = 1) = \lambda h + o(h)$.

(iv') $P(N(h) \geq 2) = o(h)$.

Proof of (i') - (iv') ⇒ (i) - (iii). Let $P_n(t) = P(N(t) = n)$. Goal is to show

$$P_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

where $n = 0, 1, 2, \cdots$

$$\begin{aligned}
P_0(t+h) &= P(N(t+h) = 0) \\
&= P((N(t) = 0, N(t+h) - N(t) = 0) \\
&= P(N(t) = 0)P(N(h) = 0) \\
&= P_0(t)(1 - \lambda h) + o(h)
\end{aligned}$$

$$\begin{aligned}
\frac{P_0(t+h) - P_0(t)}{h} &= -\lambda P_0(t) + \frac{o(h)}{h} \\
P_0'(t) &= -\lambda P_0(t) + \frac{o(h)}{h} \\
P_0'(t) &= -\lambda p_0(t) \\
P_0(0) &= P(N(0) = 0) = 1 \\
P_0(t) &= e^{-\lambda t}
\end{aligned}$$

For $n \geq 1$,

$$\begin{aligned}
P_n(t+h) &= P(N(t+h) = n) \\
&= P(N(t) = n-1, N(t+h) - N(t) = 1) \\
&\quad + P(N(t) = n, N(t+h) - N(t) = 0) \\
&\quad + P(N(t+h) = n, N(t+h) - N(t) \geq 2) \\
&= P_{n-1}(t)P(N(h) = 1) + P_n(t)P(N(h) = 0) + o(h) \\
&= P_{n-1}(t)\lambda h + P_n(t)(1 - \lambda h) + o(h)
\end{aligned}$$

$$P_n(t+h) - P(t) = P_{n-1}(t)\lambda h - P_n(t)\lambda h + o(h)$$

$$\begin{cases} P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}'(t) \\ P_n(0) = P(N(0) = n) = 0 \end{cases}$$

□

2.2 Interarrival and Waiting Times

Consider a Poisson process $\{N(t), t > 0\}$.

X_1: the time of the first event.

X_n: the time between the $(n-1) < t$ and the nth events.

Sequence of r.v.s $\{X_n\}_{n=1}^{\infty}$: sequence of interarrival time.

$$P(X_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

$$\begin{aligned} P(X_2 > t | X_1 = s) &= P(N(t+s) - N(s) = 0 | N(s) = 1) \\ &= P(N(t) = 0) \\ &= e^{-\lambda t} \end{aligned}$$

Proposition 2.2.1. *X_n, $n = 1, 2, \cdots$, are independent identically distributed (i.i.d.) exponential r.v.s having mean $\frac{1}{\lambda}$.*

Remark. $\boxed{\lambda\text{: arrival rate.}}$

$S_n = \sum_{i=1}^{n} X_i$, $n \geq 1$: the arrived time of the nth event, or the waiting time until the nth event.

$$\begin{aligned} P(S_n \leq t) &= P(N(t) \geq n) \\ &= \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!}, \quad t > 0 \end{aligned}$$

which upon differentiation yields that the density function

$$\begin{aligned} f_n(t) &= \sum_{j=0}^{\infty} \left[-\lambda e^{-\lambda t} \frac{(\lambda t)^j}{j!} + \lambda e^{-\lambda t} \frac{\lambda^{j-1} t^{j-1}}{(j-1)!} \right] \\ &= -\lambda \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} + \lambda \sum_{j=n-1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!} \\ &= \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \qquad\qquad t \geq 0 \end{aligned}$$

Note. $\boxed{S_n \text{ follows gamma distribution.}}$

Another definition of the Poisson process.
Given $\{X_n\}_{n+1}^{\infty}$ i.i.d. exponential with mean $1/\lambda$.

$$\begin{aligned} S_n &= \sum_{i=1}^{n} X_i \\ N(t) &= \max\{n : S_n \leq t\} \end{aligned}$$

Example 2.2.1. Given $0 \le s < t$, compute $P(X_1 < s|N(t) = 1)$.

$$\begin{aligned}
P(X_1 < s|N(t) = 1) &= \frac{P(X_1 < s, N(t) = 1)}{P(N(t) = 1)} \\
&= \frac{P(N(s) = 1)P(N(t) - N(s) = 0)}{P(N(t) = 1)} \\
&= \frac{e^{-\lambda t}(\lambda s) \cdot e^{-\lambda(t-s)}}{e^{-\lambda t}\lambda t} \\
&= \frac{s}{t}
\end{aligned}$$

2.3 Nonhomogeneous Poisson Process

Definition 2.3.1. A counting process $\{N(t), t \ge 0\}$ is called a nonstationary or nonhomogeneous Poisson process which intensity function $\lambda(t)$, $t > 0$ if

1. $N(0) = 0$.
2. $N(t)$ has independent increments.
3. $P(N(t + h) - N(t) = 1) = \lambda(t)h + o(h)$.
4. $P(N(t + h) - N(t) \ge 2) = o(h)$

Theorem 2.3.1. *For a nonhomogeneous Poisson process* $\{N(t), t \ge 0\}$,

$$P(N(t+s) - N(t) = n) = \exp\left(-\int_t^{t+s} \lambda(r)dr\right) \frac{\left(\int_t^{t+s} \lambda(r)dr\right)^n}{n!}$$

where $n = 0, 1, 2, \cdots$

2.4 Compound Poisson r.v.s and Processes

Let $X_1, X_2, \cdots$ be a sequence of r.v.s, i.i.d., having distribution F, and let N be a Poisson r.v. with mean λ independent of $\{X_n\}_{n=1}^{\infty}$.

$W = \sum_{i=1}^{N} X_i$ is called a compound Poisson r.v. with Poisson parameter λ and component distribution F.

$$\begin{aligned}
\psi_W(t) &= E\left[e^{t\sum_{i=1}^{N} X_i}\right] \\
&= \sum_{n=0}^{\infty} E\left[e^{t\sum_{i=1}^{n} X_i}\Big| N=n\right] P(N=n) \\
&= \sum_{n=0}^{\infty}\left[\prod_{i=1}^{n} E e^{tX_i}\right] e^{-\lambda}\frac{\lambda^n}{n!} \\
&= \sum_{n=0}^{\infty}[\psi_X(t)]^n e^{-\lambda}\frac{\lambda^n}{n!} \\
&= e^{-\lambda}\sum_{n=0}^{\infty}\frac{[\lambda\,\psi_X(t)]^n}{n!} \\
&= e^{-\lambda}e^{\lambda\psi_X(t)} = e^{\lambda(\psi_X(t)-1)}
\end{aligned}$$

$$\begin{aligned}
\psi_W'(t) &= e^{\lambda(\psi_X(t)-1)}\cdot\lambda\psi_X'(t) \\
E[W] &= \psi_W'(0) = \lambda\psi_X'[0] = \lambda E[X] \\
\mathrm{Var}[W] &= \lambda E[X^2]
\end{aligned}$$

Proposition 2.4.1. *Let $W=\sum_{i=1}^{N} X_i$ be a compound Poisson r.v. with Poisson parameter λ and compound distribution F, and X be r.v. having distribution F that is independent of W. Then for any measurable function $h(x)$.*

$$E[Wh(W)] = \lambda E[Xh(W+X)]$$

Proof.

$$\begin{aligned}
E[Wh(W)] &= \sum_{n=0}^{\infty} E[Wh(W)|N=n]P(N=n) \\
&= \sum_{n=0}^{\infty} E\left[\sum_{i=1}^{n} X_i h\Big(\sum_{j=1}^{n} X_j\Big)\right]\frac{\lambda^n}{n!}e^{-\lambda} \\
&= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\sum_{i=1}^{n} E\left[X_i h\Big(\sum_{j=1}^{n} X_j\Big)\right] \\
&= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\cdot nE\left[X_i h\Big(\sum_{j=1}^{n} X_j\Big)\right]
\end{aligned}$$

$$
\begin{aligned}
&= e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^n}{(n-1)!}\int E\Big[X_n h\Big(\sum_{j=1}^{n}X_j\Big)\Big|X_n=x\Big]dF(x)\\
&= \lambda e^{-\lambda}\sum_{n=0}^{\infty}\frac{\lambda^{n-1}}{(n-1)!}\int xE\Big[h\Big(\sum_{j=1}^{n-1}X_j+x\Big)\Big]dF(x)\\
&= \lambda e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!}\int xE\Big[h\Big(\sum_{j=1}^{m}X_j+x\Big)\Big]dF(x)
\end{aligned}
$$

$$
\begin{aligned}
\lambda E[Xh(W+X)] &= \lambda\int E[xh(W+x)|X=x]dF(x)\\
&= \lambda\int xE[h(W+x)]dF(x)\\
&= \lambda\int x\sum_{m=0}^{\infty}E\Big[h\Big(\sum_{j=1}^{m}X_j+x\Big)\Big|N=n\Big]P(N=m)dF(x)\\
&= \lambda e^{-\lambda}\sum_{m=0}^{\infty}\int xE\Big[h\Big(\sum_{j=1}^{m}X_j+x\Big)\Big]\frac{\lambda^m}{m!}dF(x)
\end{aligned}
$$

□

Proposition 2.4.2. *Under the assumption of Proposition 2.4.1, we have*

$$E[W^n]=\lambda\sum_{j=0}^{n-1}\binom{n-1}{j}E[W^j]E[X^{n-j}],\quad n=1,2,\cdots$$

Proof. Take $h(x)=x^{n-1}$.

$$
\begin{aligned}
E[W^n] &= E[Wh(W)]\\
&= \lambda E[X(W+X)^{n-1}]\\
&= \lambda E\left[\sum_{j=0}^{n-1}\binom{n-1}{j}W^jX^{n-j}\right]\\
&= \lambda\sum_{j=0}^{n-1}\binom{n-1}{j}E[W^j]E[X^{n-j}]
\end{aligned}
$$

□

Suppose now X_j's are positive integer valued.

$$
\begin{aligned}
P(X_j=n) &= \alpha_n \qquad\qquad n=1,2,\cdots\\
P_n &\triangleq P(W=n)\ ?
\end{aligned}
$$

Corollary 2.4.1.

$$P_0 = e^{-\lambda}$$
$$P_n = \frac{\lambda}{n}\sum_{j=1}^{n} j\alpha_j P_{n-j} \qquad n \geq 1$$

Proof. $P_0 = P(N=0) = e^{-\lambda}$. Fix $n \geq 1$. Define

$$h(x) = \begin{cases} 0 & \text{if } x \neq n \\ \frac{1}{n} & \text{if } x = n \end{cases} = \frac{1}{n}\chi_{\{W=n\}}$$

Remark. $\boxed{Wh(W) = W\frac{1}{n}\chi_{\{W=n\}} = \chi_{\{W=n\}}}$

$$\begin{aligned}
P_n &= P(W=n) = E[\chi_{\{W=n\}}] = E[Wh(W)] \\
&= \lambda E[Xh(W+X)] \\
&= \lambda \sum_{j=1}^{\infty} E[Xh(W+X)|X=j]\alpha_j && h(W+j) = \frac{1}{n}\chi_{\{W+j=n\}} \\
&= \lambda \sum_{j=1}^{n} \frac{1}{n} jP(W=n-j)\alpha_j && = \frac{1}{n}\chi_{\{W=n-j\}} \\
&= \frac{\lambda}{n}\sum_{j=1}^{n} j\alpha_j P_{n-j}
\end{aligned}$$

□

Example 2.4.1. Let $X_j = 1 \Rightarrow W = N$.

$$P(N=0) = e^{-\lambda}$$

$$\begin{aligned}
P_n = P(N=n) &= \frac{\lambda}{n}\sum_{j=1}^{n} j\alpha_j P_{n-j} \\
&= \frac{\lambda}{n}P_{n-1} = \frac{\lambda}{n}\cdot\frac{\lambda}{n-1}\cdot P_{n-2} = \cdots = \underbrace{\frac{\lambda}{n}\cdot\frac{\lambda}{n-1}\cdots\frac{\lambda}{1}}_{n}\cdot P_0 = \frac{\lambda^n}{n!}e^{-\lambda}
\end{aligned}$$

Definition 2.4.1. A stochastic process $\{X(t), t \geq 0\}$ is called a compound Poisson process if $X(t) = \sum_{i=1}^{N(t)} X_i$, where $\{N(t), t > 0\}$ is a Poisson process and $\{X_i\}_{i=1}^{\infty}$ is a family of i.i.d. r.v.s that independent of $\{N(t), t > 0\}$.

Example 2.4.2. An insurance company receives claims that arrive at a Poisson rate λ. Suppose that amounts of claims form a set of i.i.d. random variables that is independent of the claim arrival process. If $X(t)$ denotes the total amount of claims by time t, then $\{X(t), t \geq 0\}$ is a compound Poisson process.

2.5 Conditional Poisson Processes

Let $\Lambda > 0$ be a r.v. having distribution G and $\{N(t), t > 0\}$ be a counting process such that given $\Lambda = \lambda$, $\{N(t), t \geq 0\}$ is a Poisson process having rate λ.

$$\begin{aligned} P(N(t+s) - N(s) = n) &= \int_0^\infty P(N(t+s) - N(s)|\Lambda = \lambda)dG(\lambda) \\ &= \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!} dG(\lambda) \end{aligned}$$

The processes $N(t), t \geq 0$ is called a conditional Poisson process.

Note. $\boxed{\text{It is not a Poisson process.}}$

$$\begin{aligned} P(\Lambda \in (\lambda, \lambda + d\lambda)|N(t) = n) &= \frac{P(\Lambda \in (\lambda, \lambda + d\lambda, N(t) = n)}{P(N(t) = n} \\ &= \frac{\frac{e^{-\lambda t}}{n!}(\lambda t)^n dG(\lambda)}{\int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!} dG(\lambda)} \end{aligned}$$

$$\Rightarrow P(\Lambda \leq x|N(t) = n) = \frac{\int_0^x \frac{e^{-\lambda t(\lambda t)^n}}{n!} dG(\lambda)}{\int_0^\infty \frac{e^{-\lambda t}(\lambda t)^n}{n!} dG(\lambda)}$$

Example 2.5.1. Suppose that, depending on factors not at present understood, the average rate at which seismic shocks occur in a certain region over a given season is either λ_1 and λ_2. Suppose also that it is λ_1 with probability p and λ_2 with probability $1-p$. A simple model for such a situation would be to suppose that $\{N(t), t \geq 0\}$ is a conditional Poisson process such that Λ is either λ_1 or λ_2 with respective probabilities p and $1-p$. Given n shocks in the first t time units of a season, what is the probability that it is currently a λ_1 season and what is the distribution of the time from t until the next shock?

$$\begin{aligned} P(\Lambda = \lambda_1|N(t) = n) &= \frac{P(N(t) = n|\Lambda = \lambda_1)P(\Lambda = \lambda_1)}{P(N(t) = n} \\ &= \frac{pe^{-\lambda_1 t}\frac{(\lambda_1 t)^n}{n!}}{pe^{-\lambda_1 t}\frac{(\lambda_1 t)^n}{n!} + (1-p)e^{-\lambda_2 t}\frac{(\lambda_2 t)^n}{n!}} \\ &= \frac{pe^{-\lambda_1 t}(\lambda_t t)^n}{pe^{-\lambda_1 t}(\lambda_1 t)^n + (1-p)^{-\lambda_2 t}(\lambda_2 t)^n} \end{aligned}$$

$$P^{t,n}(.) \triangleq P(.|N(t) = n)$$

$$
\begin{aligned}
&P(\underbrace{\text{time from } t \text{ until next shock} \le x}_{A}|N(t) = n) \\
&= P^{t,n}(A) \\
&= P^{t,n}(A|\Lambda = \lambda_1)P^{t,n}(\Lambda = \lambda_1) + P^{t,n}(A|\Lambda = \lambda_2)P^{t,n}(\Lambda = \lambda_2) \\
&= \frac{(1 - e^{-\lambda_1 x})pe^{\lambda_1 t}(\lambda_1 t)^n + (1 - e^{-\lambda_2 x})(1 - p)e^{-\lambda_2 t}(\lambda_2 t)^n}{pe^{-\lambda_1 t}(\lambda_1 t)^n + (1 - p)e^{-\lambda_2 t}(\lambda_2 t)^n}
\end{aligned}
$$

2.6 Introduction of Renewal Processes

Let $\{X_n, n = 1, 2, \cdots\}$ be a sequence of nonnegative i.i.d. r.v.s with a common distribution F.

Assume $F(0) \equiv P(X_n = 0) < 1$,
$0 < \mu \triangleq E[X_n] < \infty$.

Define

$$
\begin{aligned}
S_0 &= 0 \\
S_n &= \sum_{i=1}^{n} X_i \qquad\qquad n = 1, 2, \cdots
\end{aligned}
$$

Let $N(t) \triangleq \sup\{n : S_n \le t\}$.

Definition 2.6.1. The counting process $\{N(t), t \ge 0\}$ is called a renewal process.

S_n : the time of the nth event or renewal.
X_n: the interarrival time between $(n-1)$th and nth events.

$$
\begin{aligned}
&\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu > 0 \text{ as } n \to \infty \text{ a.s.} \quad \text{by strong law of large number} \\
&\Rightarrow S_n \to \infty \text{ as } n \to \infty \\
&\Rightarrow N(t) = \max\{n : S_n \le t\}
\end{aligned}
$$

In a finite time of period, there are only a finite number of renewals.

$$
\begin{aligned}
N(t) \ge n &\Leftrightarrow S_n \le t \\
P(N(t) = n) &= P(N(t) > n) - P(N(t) \ge n + 1) \\
&= P(S_n \le t) - P(S_{n+1} \le t) \\
&= F_n(t) - F_{n+1}(t)
\end{aligned}
$$

where F_n is the distribution function of S_n.

$E[N(t)] \triangleq m(t)$ is the renewal function.

Chapter 3

Discrete-Time Markov Chains

3.1 Definition

Let $\{X_n, n = 0, 1, 2, \cdots\}$ be a (discrete-time) stochastic process taking on a finite or countable number of possible values, say, $\{0, 1, 2, \cdots\}$ — state space. If $X_n = i$, then the process is said to be in state i at time n. Suppose

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0) = P_{ij}$$

for all states $i_0, i_1, \cdots, i_{n-1}, i, j$ and all $n \geq 0$. Then $\{X_n, n \geq 0\}$ is called a (discrete-time) stationary Markov chain.

Markov property: The conditional distribution of any future state X_{n+1}, given the past state $X_0, X_1, \cdots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present states.

$\{P_{ij}\}$ satisfies

$$\begin{aligned} &P_{ij} > 0 \\ &\sum_{j=0}^{\infty} P_{ij} = 1 \qquad\qquad i = 0, 1, 2, \cdots \end{aligned}$$

(One-step) transition probability matrix

$$P = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots & P_{0j} & \dots \\ P_{10} & P_{11} & P_{12} & \dots & P_{1j} & \dots \\ \vdots & & & & & \\ P_{i0} & P_{i1} & P_{i2} & \dots & P_{ij} & \dots \\ \vdots & & & & & \end{bmatrix}_{\infty\times\infty}$$

Example 3.1.1. (General Random Walk) Let $\{X_i\}_{i=1}^{\infty}$ be i.i.d. with $P(X_i = j) = a_j$, $j = 0, \pm 1, \pm 2, \ldots$

Let $S_0 = 0$, $S_n = \sum_{i=1}^{n} X_i$, $\{S_n, n \geq 0\}$ is a Markov chain.

$$\begin{aligned} P_{ij} &= P(S_{n+1} = j | S_n = i) \\ &= P(X_n = j - i) \\ &= a_{j-i} \end{aligned}$$

Example 3.1.2. (The M/G/1 Queue) Customers arrive at a service center according to a Poisson process with rate λ. There is a single server and those arrivals finding the server free go immediately into service; all others wait in line until their service turn. The service time of successive customers are assumed to be independent random variables having a common distribution G; and they are also assumed to be independent of the arrival process.

X_n : # of customers left behind by the nth departure, $n = 1, 2, \ldots$
Y_n : # of customers arriving during the service period of $(n+1)$st customer.
S_n : service time of the nth customer.

$$X_{n+1} = \begin{cases} X_n - 1 + Y_n & \text{if } X_n > 0 \\ Y_n & \text{if } X_n = 0 \end{cases}$$

$\{X_n : n \geq 1\}$ is a Markov chain.

$$\begin{aligned} P(Y_n = j) &= \int_{-\infty}^{\infty} P(Y_n = j | S_{n+1} = x) dG(x) \\ &= \int_0^{\infty} \frac{(\lambda x)^n}{j!} e^{-\lambda x} dG(x) \qquad\qquad j = 0, 1, 2, \ldots \end{aligned}$$

$$\begin{aligned} &P_{0j} = P(X_{n+1} = j | X_n = 0) = P(Y_n = j) && j = 0, 1, 2, \ldots \\ &P_{ij} = P(X_{n+1} = j | X_n = i) = P(Y_n = j - i + 1) && i = 1, 2, \ldots \quad j \geq i - 1 \\ &P_{ij} = 0 && i = 1, 2, \ldots \quad j < i - 1 \end{aligned}$$

3.2 Chapman-Kolmogorov Equations

n-step transition probabilities

$$P_{ij}^{(n)} = P(X_{m+n} = j | X_m = i)$$

where $n \geq 0$, $i, j \geq 0$.

Theorem 3.2.1.

$$P_{ij}^{(m+n)} = \sum_{k=0}^{\infty} P_{ik}^{(n)} P_{kj}^{(m)}$$

$\forall n, m \geq 0$, $\forall i, j \geq 0$

Proof.

$$\begin{aligned}
P_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) \\
&= \sum_{k=0}^{\infty} P(X_{m+n} = j | X_n = k, X_0 = i) P(X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} P(X_{m+n} = j | X_n = k) P(X_n = k | X_0 = i) \\
&= \sum_{k=0}^{\infty} P_{kj}^{(m)} P_{ik}^{(n)}
\end{aligned}$$

□

n-step transition matrix

$$\begin{gathered}
P^{(n)} = (P_{ij}^{(n)} \\
P^{(m+n)} = P^{(n)} P^{(m)} \\
P^{(n)} = P \cdot P^{(n-1)} = \cdots = P \cdot P \ldots P = P^n
\end{gathered}$$

3.3 Classification of States

State j is said to be accessible from state i if $P_{ij}^n > 0$ for some $n \geq 0$. Two state i and j accessible to each other are said to communicate and we write $i \leftrightarrow j$.

Proposition 3.3.1. *Communication is an equivalence relation. That is*

(i) $i \leftrightarrow i$;

(ii) if $i \leftrightarrow j$, *then* $j \leftrightarrow i$;

(iii) if $i \leftrightarrow j$, *and* $j \leftrightarrow k$, *then* $i \leftrightarrow k$.

Proof. (i),(ii) are trivial. For (iii) $\exists m, n \geq 0$, such that $P_{ij}^{(m)} > 0$, $P_{jk}^{(n)} > 0$.

$$\begin{aligned}
P_{ik}^{(m+n)} &= \sum_{r=0}^{\infty} P_{ir}^{(m)} P_{rk}^{(n)} \\
&\geq P_{ij}^{(m)} P_{jk}^{(n)} > 0
\end{aligned}$$

□

Two states that communicate are said to be in the same class. The Markov chain is said to be irreducible if there is only one class.

Given two state i and j. Define $f_{ij}^{(n)}$ the probability that starting in i, the first transition into j occurs at time n, $n = 0, 1, 2, \ldots$ ($f_{ij}^{(0)} = 0$, $i \neq j$; $f_{ii}^{(0)} = 1$)

$$f_{ij}^{(n)} = P(X_n = j, X_k \neq j, \forall k = 0, 1, \ldots, n-1 | X_0 = i)$$

$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$: the probability of ever making a transition into j, given starting in i. $f_{ij} > 0 \Leftrightarrow j$ is accessible from i, for $i \neq j$.

State j is called recurrence if $f_{jj} = 1$, and transient if $f_{jj} < 1$. A recurrence state j is called absorbing if $P_{jj} = 1$.

Theorem 3.3.1. *State j is recurrence iff $\sum_{n=1}^{\infty} P_{ij}^{(n)} = +\infty$.*

Proof. We want to show j is recurrent iff

$$E[\# \text{ of visits to } j | X_0 = j] = +\infty$$

If j is recurrent then w.p.1 the number of visits to j will be infinite.

If j is transient. At each time the process returns to j there is a positive probability $1 - f_{jj} > 0$ that it will never return again.

Bernoulli trial: "success" if it will never returns;
"failure" if it will return.
number of visits to j = the trial number on which the first success occurs.

$$E[\# \text{ of visits to j} | X_0 = j] = \frac{1}{1 - f_{jj}} < +\infty$$

$$\begin{aligned} E[\# \text{ of visits to j} | X_0 = j] &= E\left[\sum_{n=1}^{\infty} \chi_{\{X_n = j\}} | X_0 = j\right] \\ &= \sum_{n=1}^{\infty} E[\chi_{\{X_n = j\}} | X_0 = j] \\ &= \sum_{n=1}^{\infty} P(X_n = j | X_0 = j) \end{aligned}$$

□

Corollary 3.3.1. *With probability 1 a transient state will only be visited a finite number of times.*

Corollary 3.3.2. *A finite-state Markov chain has at least one recurrent state.*

Proof. Suppose the states are $1, 2, \ldots, M$. With probability 1, after a finite number of time T_i, state i will never be visited, $i = 1, 2, \ldots, M$.

Let $T = \sum_{i=1}^{M} T_i$. After T no state will be visited. Contradiction. □

Corollary 3.3.3. *If i is recurrent (resp. transient) and $i \leftrightarrow j$, then j is recurrent (resp. transient).*

Proof. Let m and n be such that $P_{ij}^{(n)} > 0$, $P_{ji}^{(m)} > 0$. For any $s \geq 1$,

$$\begin{aligned} P_{jj}^{(m+n+i)} &= \sum_{k,l} P_{jk}^{(m)} P_{kl}^{(s)} P_{lj}^{(n)} \\ &\geq P_{ji}^{(m)} P_{ii}^{(s)} P_{ij}^{(n)} \end{aligned}$$

$$\sum_{s=1}^{\infty} P_{jj}^{m+n+s} \geq P_{ji}^{(m)} \left(\sum_{s=1}^{\infty} P_{ii}^{(s)} \right) P_{ij}^{(n)} = +\infty$$

□

State i is said to have period d is d is the greatest common divisor of those $n \geq 1$ that $P_{ii}^{(n)} > 0$. (If $P_{ii}^{(n)} = 0 \forall n \geq 1$, then the period is defined to be $+\infty$.)

A state with $d = 1$ is said to be aperiodic.

Example 3.3.1. Consider a Markov chain with the one-step transition matrix.

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$P_{00}^{(1)} = 0$, $P_{00}^{(1)} = \frac{1}{3} > 0$, $P_{00}^{(3)} = 0$, $P_{00}^{(4)} > 0$.
$\{n \geq 1 | P_{00}^{(n)} > 0\} = \{2, 4, 6, \dots\}$. $d = 2$.

Example 3.3.2.

$$\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}$$

$P_{00}^{(1)} = 0.4 > 0$
$d = 1$. Aperiodic.

Example 3.3.3. (Simple Random Walk) Consider an Markov chain with state space $\{0, \pm 1, \pm 2, \dots\}$ and transition probabilities $P_{i,i+1} = p$, $P_{i,i-1} = 1 - p$, $i = 0, \pm 1, \pm 2, \dots$ where $0 < p < 1$. This is an irreducible Markov chain. $d = 2$.

$$P_{00}^{(2n+1)} = 0$$

$$P_{00}^{(2n)} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{(n!)^2} \cdot [p(1-p)]^n$$

Remark.

Sterling's estimation $$n! \sim n^{n+1/2}\, e^{-n} \sqrt{2\pi}$$

$$\begin{aligned}\frac{(2n)!}{(n!)^2}\cdot[p(1-p)]^n &\sim \frac{(2n)^{2n+1/2}\,e^{-2n}\sqrt{2\pi}}{n^{2n+1}\,e^{-2n}\sqrt{2\pi}}\cdot[p(1-p)]^n \\ &= \frac{1}{\sqrt{\pi n}}\cdot[4p(1-p)]^n \qquad (*)\end{aligned}$$

$4p(1-p) \leq 1$ and $4p(1-p) = 1$ iff $p = \frac{1}{2}$.

(i) If $p = \frac{1}{2}$, $(*) = \frac{1}{\sqrt{\pi n}} \Rightarrow 0$ is recurrent.

(ii) If $p \neq \frac{1}{2}$, $(*) = \frac{[4p(1-p)]}{\sqrt{\pi n}} \Rightarrow 0$ is transient.

3.4 Limit Theorem

For a state j, define

$$\mu_{jj} = \begin{cases} \infty & \text{if } j \text{ is transient.} \\ \sum_{n=1}^{\infty} n f_{jj}^{(n)} & \text{if } j \text{ is recurrent.} \end{cases}$$

μ_{jj} is expected number of transition needed to return to state j.

Theorem 3.4.1. *If $i \leftrightarrow j$ then*

(i)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} P_{ij}^{(k)} = \frac{1}{\mu_{jj}}$$

(ii) If j is aperiodic, then

$$\lim_{n\to\infty} P_{ij}^{(n)} = \frac{1}{\mu_{jj}} = \pi_j$$

Definition 3.4.1. If state j is recurrent, then it is said to be positive recurrent if $\mu_{jj} < +\infty$, and null recurrent if $\mu_{jj} = +\infty$. A positive recurrent, aperiodic state is called ergodic.

Definition 3.4.2. A probability distribution $\{P_j, j \geq 0\}$ is called stationary for the Markov chain if $P_j = \sum_{i=0}^{\infty} P_i P_{ij}$, $\forall j \geq 0$.

Corollary 3.4.1. *If $\{P_j, j > 0\}$ is stationary, then*

$$P_j = \sum_{i=0}^{\infty} P_i P_{ij}^{(x)}, \forall j \geq 0, \forall n \geq 1$$

Proof.

$$\begin{aligned}
\sum_{i=0}^{\infty} P_i P_{ij}^{(n)} &= \sum_{i=0}^{\infty} P_i \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)} \\
&= \sum_{k=0}^{\infty} \left(\sum_{i=0}^{\infty} P_i P_{ik} \right) P_{kj}^{(n-1)} \\
&= \sum_{k=0}^{\infty} P_k P_{kj}^{(n-1)} \\
&= \cdots = P_j
\end{aligned}$$

□

If the distribution of X_0 is stationary, $P_j = P(X_0 = j)$, then the distribution of X_n is stationary. The proof is omitted here.

Theorem 3.4.2. *A irreducible, aperiodic Markov chain belongs to one of the following two classes:*

(i) Either the states are all transient or all null recurrent; in this case, $P_{ij}^{(n)} \to 0$ as $n \to \infty$ for all i, j and there exists no stationary distribution.

(ii) Or else, all states are positive recurrent, that is $\pi_{ij} = \lim_{n\to\infty} P_{ij}^{(n)} > 0$. In this case $\{\pi_j, j \geq 0\}$ is a stationary distribution and there exits no other stationary distribution.

Proof. Step 1 Prove if j is positive recurrent, and $k \to j$, then k is also positive recurrent.

$$\pi_j = \lim_{n\to\infty} P_{ij}^{(n)} > 0$$

Let m be such that $P_{jk}^{(m)} > 0$. Then $P_{ik}^{(n+m)} \geq P_{ij}^{(n)} P_{jk}^{(m)}$

$$\pi_k = \lim_{n\to\infty} P_{ik}^{(n+m)} \geq \left(\lim_{n\to\infty} P_{ij}^{(n)} \right) P_{jk}^{(m)} > 0$$

Step 2 Prove that if the M.C. is positive recurrent then $\left\{ \frac{\pi_j}{\sum_{k=0}^{\infty} \pi_k}, j \geq 0 \right\}$

is a stationary distribution.

$$\sum_{j=0}^{M} P_{ij}^{(n)} \leq \sum_{j=0}^{\infty} P_{ij}^{(n)} = 1 \qquad \forall M$$

$$\sum_{j=0}^{M} \pi_j \leq 1 \qquad \forall M \qquad\qquad \Rightarrow \sum_{j=0}^{\infty} \pi_j \leq 1$$

$$P_{ij}^{(n+1)} = \sum_{k=0}^{\infty} P_{ik}^{(n)} P_{kj} \geq \sum_{k=0}^{M} P_{ik}^{(n)} P_{kj} \qquad \forall M$$

$$\pi_j \geq \sum_{k=0}^{M} \pi_k P_{kj} \qquad \forall M \qquad\qquad \Rightarrow \pi_j \geq \sum_{k=0}^{\infty} \pi_k P_{kj}$$

If $\pi_j \geq \sum_{j=0}^{\infty} \pi_k P_{kj}$ for some j,

$$\begin{aligned}
\sum_{j=0}^{\infty} \pi_j &> \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k P_{kj} \\
&= \sum_{k=0}^{\infty} \pi_k \left(\sum_{j=0}^{\infty} P_{kj} \right) \\
&= \sum_{k=0}^{\infty} \pi_k \Rightarrow \text{contradiction.}
\end{aligned}$$

Step 3 Let $\{\overline{P}_j, j \geq 0\}$ be any stationary distribution. Then by Corollary 3.4.1

$$\begin{aligned}
\overline{P}_j = \sum_{i=0}^{\infty} P_{ij}^{(n)} \overline{P}_i &\geq \sum_{i=0}^{M} P_{ij}^{(n)} \overline{P}_i \qquad\qquad \forall M \\
\Rightarrow \overline{P}_j &\geq \sum_{i=0}^{\infty} \pi_j \overline{P}_i = \pi_j
\end{aligned}$$

On the other hand

$$\begin{aligned}
\overline{P}_j &= \sum_{i=0}^{M} P_{ij}^{(n)} \overline{P}_i + \sum_{i=M+1}^{\infty} P_{ij}^{(n)} \overline{P}_i \\
&\leq \sum_{i=0}^{M} P_{ij}^{(n)} \overline{P}_i + \sum_{i=M+1}^{\infty} \overline{P}_i \qquad\qquad \forall M \\
\Rightarrow \overline{P}_i &\leq \sum_{i=0}^{M} \pi_j \overline{P}_j + \sum_{i=M+1}^{\infty} \overline{P}_i \qquad\qquad \forall M
\end{aligned}$$

Let $M \to \infty \Rightarrow$

$$\overline{P}_j \le \sum_{i=0}^{\infty} \pi_j \overline{P}_i = \pi_j$$
$$\frac{\pi_j}{\sum_{k=0}^{\infty} \pi_k} = \pi_j$$

Step 4 If all the states are transient or null recurrent, and $\{P_j, j \ge 0\}$ is a stationary distribution. By Corollary 3.4.1, $P_j = 0\ \forall j$. Contradiction.

□

Corollary 3.4.2. *For case (ii) in Theorem 3.4.2, the limiting probability are obtained by solving*

$$\begin{cases} \pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij} & j = 0, 1, 2, \cdots \\ \sum_{i=0}^{\infty} \pi_i = 1 \end{cases}$$

Example 3.4.1. (Weather Chain) Consider transition matrix

$$\begin{bmatrix} 0.4 & 0.6 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.7 & 0.2 \end{bmatrix}$$

$$\begin{aligned} \pi &= (\pi_0 + \pi_1 + \pi_2) \\ \pi_0 &= 0.4\pi_0 + 0.2\pi_1 + 0.1\pi_2 \\ \pi_1 &= 0.6\pi_2 + 0.5\pi_1 + 0.7\pi_2 \\ \pi_2 &= 0.3\pi_1 + 0.2\pi_2 \\ 1 &= \pi_0 + \pi_1 + \pi_2 \\ &\Rightarrow \ \pi_0 = \frac{19}{85}\ \pi_1 = \frac{48}{85}\ \pi_2 = \frac{18}{85} \end{aligned}$$

Example 3.4.2. (The M/G/1 Queue)

$$a_j = P(Y_n = j) = \int_0^{\infty} e^{-\lambda x} \frac{(\lambda x)^j}{j!} dG(x), j = 0, 1, 2, \cdots$$

$$\begin{cases} P_{0j} = a_j & j \ge 0 \\ P_{ij} = a_{j-i+1} & i \ge 1, j \ge i-1 \\ P_{ij} = 0 & i \ge 1, j < i-1 \end{cases}$$

$$\pi_j = \pi_0 a_j + \sum_{i=1}^{j+1} \pi_i a_{j-i+1},\ j = 0, 1, 2, \cdots$$

Introduce the generating function.

$$\pi(s) = \sum_{j=0}^{\infty} \pi_j s^j,\ A(s) = \sum_{j=0}^{\infty} a_j s^j$$

$$
\begin{aligned}
\pi(s) &= \sum_{j=0}^{\infty}\left(\pi_0 a_j + \sum_{i=1}^{j+1}\pi_i a_{j-i+1}\right)s^j \\
&= \pi_0 A(s) + \sum_{j=0}^{\infty}\sum_{i=1}^{j+1}\pi_i a_{j-i+1}s^j \\
&= \pi_0 A(s) + \sum_{i=1}^{\infty}\sum_{j=i-1}^{\infty}\pi_i a_{j-i+1}s^j \\
&= \pi_0 A(s) + s^{-1}\sum_{i=1}^{\infty}\pi_i s^i \underbrace{\left(\sum_{j=i-1}^{\infty} a_{j-i+1}s^{j-i+1}\right)}_{A(s)} \\
&= \pi_0 A(s) + \frac{A(s)(\pi(s)-\pi_0)}{s}
\end{aligned}
$$

$$
\pi(s) = \frac{(s-1)\pi_0 A(s)}{s-A(s)}
$$

$$
\begin{aligned}
\lim_{s\to 1}\pi(s) &= \lim_{s\to 1}\frac{(s-1)\pi_0 A(s)}{s-A(s)} \\
&= \pi_0 \lim_{s\to 1}\frac{A(s)+(s-1)A'(s)}{1-A'(s)} \\
&= \pi_0 \frac{1}{1-\rho}
\end{aligned}
$$

where $\rho = A'(1) = \sum_{j=0}^{\infty} j a_j = E[Y_n]$.

Therefore the stationary distribution exists if and only if $\rho < 1$. In this case $\pi_0 = 1 - \rho$.

$$
\begin{aligned}
\rho = E[Y_n] &= \int_0^{\infty} E[Y_n|S_{n+1}=x]dG(x) \\
&= \int_0^{\infty} \lambda x dG(x) \\
&= \lambda \int_0^{\infty} x dG(x) = \lambda E[s]
\end{aligned}
$$

$$
\rho < 1 \Leftrightarrow \lambda < \frac{1}{E[s]}
$$

Chapter 4

Continuous-Time Markov Chains

4.1 Definitions

Let $\{X(t), t \geq 0\}$ be a continuous-time stochastic process taking values in the set of nonnegative integers. If

$$P\{X(t+s) = j|X(s) = i, X(u) = x(u), 0 \leq u < s\} = P(X(t+s) = j|X(s) = i)$$

for all $s, t \geq 0$, and nonnegative integers $i, j, x(u), 0 \leq u < s$, then $\{X(t), t \geq 0\}$ is called a continuous-time Markov chain.

If $P(X(t+s) = j|X(s) = i)$ is independent of s, then the Markov chain is called stationary or homogeneous.

Let τ_i denote the amount of time that the process stays in state i before making a transition into a different states. τ_i is exponentially distributed with parameter v_i ($E\tau_i = \frac{1}{v_i}$). A state i with $v_i = +\infty$ is called instantaneous.

Assume throughout that $0 \leq v_i < +\infty$, $\forall i$.

The state i is called absorbing if $v_i = 0$. An Markov chain is called regular if, w.p.1, the number of transitions in any finite length of time is finite.

$$P_{ij}(t) = P(X(t+s) = j|X(s) = i)$$

Transition intensity

$$q_{ij} = \begin{cases} P'_{ij}(0) = \lim\limits_{t \to \infty} \dfrac{P_{ij}(t)}{t} \geq 0 & \text{if } i \neq j \\ P'_{ij}(0) = \lim\limits_{t \to \infty} \dfrac{P_{ii}(t) - 1}{t} \leq 0 & \text{if } i = j \end{cases}$$

Remark.
$$\boxed{\begin{gathered}\sum_j q_{ij} = 0 \\ q_{ii} = -\sum_{j\neq i} q_{ij}\end{gathered}}$$

Note. $\boxed{Q = (q_{ij})\text{: generator of Markov chain.}}$

$$P_{ij}(\Delta t) = P_{ij}(0) + P'_{ij}(0)\Delta t = q_{ij}\Delta t, \quad i \neq j$$

q_{ij}: transition rate from i to j.

$$P_{ii}(\Delta t) = P_{ii}(0) + P'_{ii}(0)\Delta t = 1 + q_{ii}\Delta t$$
$$v_i = -q_{ii} = \sum_{j\neq i} q_{ij}$$

v_i: rate at which the process makes a transition in state i.
Probability that a transition from i to j occurs

$$P_{ij} = q_{ij}\left(-\frac{1}{q_{ii}}\right) = \frac{q_{ij}}{v_i} \Leftrightarrow q_{ij} = v_i P_{ij}$$

4.2 Birth and Death Processes

Definition 4.2.1. A continuous-time Markov chain with states, 0,1,2,..., for which $q_{ij} = 0$ whenever $|i - j| > 1$ is called a birth and death process.

$$\begin{aligned}
\lambda_i &= q_{i,i+1} && \text{birth rate} \\
\mu_i &= q_{i,i-1} && \text{death rate} \\
v_i &= \lambda_i + \mu_i \\
P_{i,i+1} &= \frac{\lambda_i}{\lambda_i + \mu_i} \\
P_{i,i-1} &= \frac{\mu_i}{\lambda_i + \mu_i}
\end{aligned}$$

For BAD process, whenever the process is in state i the time until the next birth is exponential with rate λ_i, and is independent of the time until the next death, which is exponential with rate μ_i.

Example 4.2.1. (The M/M/s Queue) Suppose that customers arrive at an s-server service station in accordance with a Poisson process having rate λ. That is, the times between successive arrivals are independent exponential random variables having mean $1/\lambda$. Each customer, upon arrival, goes directly into service if any of the servers are free, and if not, then the customer joins the queue (that is, he waits in line). When a server finishes serving a customer, the customer leaves the system, and the next customer in line, if there are any waiting, enters the service. The successive service times are assumed to be independent exponential random variables having mean $1/\mu$.

Let $X(t)$ denote the number of customers at time t.

Then $\{X(t), t \geq 0\}$ is a BAD process.

$$\lambda_n = \lambda$$

$$\mu_n = \begin{cases} s\mu & n \geq s \\ n\mu & n < s \end{cases} = \min(n, s)\mu$$

A BAD process is called a pure birth process if $\mu_n = 0\ \forall n$.

Example 4.2.2. (Yule Process) Consider a pure birth process resulting from a population where each member acts independently and gives birth at an exponential rate λ. No one ever dies.

Let $X(t)$ denote the population size at t.

$\{X(t), t \geq 0\}$ is a pure birth process with $\lambda_n = n\lambda$, $n \geq 1$.

Consider the case when $i = 1$, $X(0) = 1$.

$$P_{ij}(t) = P(X(t) \geq j | X(0) = 1) - P(X(t) \geq j + 1 | X(0) = 1)$$

Let T_i denote the time between the $(i-1)$st and the ith birth.

$$P(X(t) \geq j | X(0) = 1) = P(T_1 + T_2 + \cdots + T_{j-1} \leq t | X(0) = 1)$$

$\{T_i, i \geq 1\}$ are independt and T_i is exponential with rate $i\lambda$.

$$P(T_1 \leq t) = 1 - e^{-\lambda t}$$

$$\begin{aligned} P(T_1 + T_2 \leq t) &= \int_0^t P(T_1 + T_2 \leq t | T_1 = x)\lambda e^{-\lambda x} dx \\ &= \int_0^t P(T_2 \leq t - x)\lambda e^{-\lambda x} dx \\ &= \int_0^t (1 - e^{-2\lambda(t-x)})\lambda e^{-\lambda x} dx \\ &= (1 - e^{-\lambda t})^2 \end{aligned}$$

$$P(T_1 + T_2 + \cdots + T_j \leq t) = (1 - e^{-\lambda t})^j$$

$$\begin{aligned} P_{ij}(t) &= (1 - e^{-\lambda t})^{j-1} - (1 - e^{-\lambda t})^j \\ &= (1 - e^{-\lambda t})^{j-1} e^{-\lambda t} \qquad \forall j = 1, 2, \cdots \end{aligned}$$

This is a geometric distribution with mean $e^{\lambda t}$.

Now if the population starts with i individuals, $X(0) = i$, the population size at t will be the sum of i i.i.d. geometric r.v.s, hence is a negative binomial distribution

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-\lambda t i}(1 - e^{-\lambda t})^{j-i}, \quad j \geq i \geq 1$$

with mean $ie^{\lambda t}$.

4.3 The Kolmogorov Differential Equations

Lemma 4.3.1. *For all $s, t \geq 0$*

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s)$$

Proof left as an exercise. □

Theorem 4.3.1. *(Kolmogrov's Backward Equations) For all i, j and $t \geq 0$*

$$P'_{ij}(t) = \sum_{k=0}^{\infty} q_{ik} P_{kj}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - V_i P_{ij}(t)$$
$$P'(t) = QP(t)$$

Proof. Assume there are finite states.

$$\begin{aligned} \frac{1}{h}[P_{ij}(t+h) - P_{ij}(t)] &= \frac{1}{h}\left[\sum_{k=0}^{N} P_{ik}(h) P_{kj}(t) - P_{ij}(t)\right] \\ &= \frac{1}{h}\left[\sum_{k \neq i} P_{ik}(h) P_{kj}(t) + (P_{ii}(h) - 1) P_{ij}(t)\right] \end{aligned}$$

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) + q_{ii} P_{ij}(t)$$

□

Theorem 4.3.2. *(Kolmogrov's Forward Equations) Under suitable conditions (including BAD process and finite state chains).*

$$P'_{ij}(t) = \sum_{k} q_{kj} P_{ik}(t) = \sum_{k \neq i} q_{kj} P_{ik}(t) - v_j P_{ij}(t)$$
$$P'(t) = P(t)Q$$

Example 4.3.1. Consider a two state Markov chain with

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} P = \begin{bmatrix} P_{00}(t) & P_{01}(t) \\ P_{10}(t) & P_{11}(t) \end{bmatrix}$$

$$\begin{aligned} P'_{00}(t) &= -\lambda P_{00}(t) + \mu P_{01}(t) \\ &= -\lambda P_{00}(t) + \mu(1 - P_{00}(t)) \\ &= -(\lambda + \mu) P_{00}(t) + \mu \\ P_{00}(0) &= 1 \end{aligned}$$

$$\frac{d}{dt}[e^{(-\lambda+\mu)t} P_{00}(t)] = \mu e^{(\lambda+\mu)t}$$

$$\Rightarrow \begin{cases} P_{00}(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} \\ P_{11}(t) = \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} \end{cases}$$

Example 4.3.2. For a pure birth process the forward equations are

$$\begin{aligned} P'_{ii}(t) &= \sum_{k \neq i} q_{ki} P_{ik}(t) - v_i p_{ii}(t) \\ &= -\lambda_i P_{ii}(t) \\ P_{ii}(0) &= 1 \\ \Rightarrow P_{ii}(t) &= e^{-\lambda_i t} \\ P'_{ij}(t) &= \sum_{k \neq j} q_{kj} P_{ik}(t) - v_j P_{ij}(t) \\ &= q_{j-1,j}\ P_{i,i-1}(t) - \lambda_j P_{ij}(t) \end{aligned}$$

$$\begin{aligned} \frac{d}{dt}[e^{\lambda_j t} P_{ij}(t)] &= \lambda - j - 1 e^{\lambda_j t P_{i,j-1}(t)} \\ e^{\lambda_j t} P_{ij}(t) &= \int_0^t \lambda_{j-1}\ e^{\lambda_j s} P_{i,j-1}(s) ds \\ P_{ij}(t) &= \lambda_{j-1}\ e^{-\lambda_j t} \int_0^t e^{\lambda_j s} P_{i,j-1}(s) ds \end{aligned}$$

For Yule processes, $\lambda_j = j\lambda$.

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-\lambda t i} (1 - e^{-\lambda t})^{j-1}$$

4.4 Limiting Probabilities

If $\lim_{t\to\infty} P_{ij}(t)$ exists and equals P_j, then P_j is called the limiting probability (stationary distribution) of state j.

Theorem 4.4.1. *If the limiting probability $(P_1, P_2, \cdots, P_k, \cdots) = P$ exists, then it satisfies*

$$\begin{cases} PQ = 0 \\ \sum_k P_k = 1 \end{cases}$$

Proof. Assume there are *finite* states.

$$P'_{ij}(t) = \sum_k q_{kj} P_{ik}(t) \qquad \text{(forward)}$$

$$P'_{ij}(t) = \sum_k q_{ik} P_{kj}(t) \qquad \text{(backward)}$$

$$\lim_{t\to\infty} P'_{ij}(t) = \sum_k q_{ik} P_j = P_j \cdot 0 = 0$$

$\square$

P_j is long run proportion of time the process is in state j.

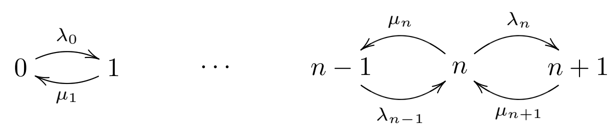
$$\sum_{k\neq j} q_{kj} P_k = v_j P_j \quad \forall j \qquad \text{(Balance equations)}$$

rate at which the process enters state j in the long run

rate at which the process leaves state j in the long run

Example 4.4.1. Consider a BAD processes

State	Rate process leave	Rate process enter
0	$\lambda_0 P_0$	$\mu_1 P_1$
$n = 1, 2, \cdots$	$(\lambda_n + \mu_n) P_n$	$\lambda_{n-1} P_{n-1} + \mu_{n-1} P_{n-1}$



$$\begin{aligned}
\lambda_0 P_0 &= \mu_1 P_1 \\
\lambda_n P_n &= \lambda_0 P_0 + \mu_{n+1} P_{n+1} - \mu_n P_n \\
\lambda_1 P_1 &= \lambda_0 P_0 + \mu_2 P_2 - \mu_1 P_1 = \mu_2 P_2 \\
&\vdots \\
\lambda_n P_n &= \mu_{n+1} P_{n+1}
\end{aligned}$$

$$\begin{aligned}
P_1 &= \frac{\lambda_0}{\mu_1} P_0 \\
P_2 &= \frac{\lambda_1}{\mu_2} P_1 = \frac{\lambda_1 \lambda_2}{\mu_1 \mu_2} P_0 \\
&\vdots \\
P_n &= \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} P_0
\end{aligned}$$

$$P_0 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} P_0 = 1$$

$$P_0 = \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \right)^{-1}$$

Provide that

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < +\infty$$

Example 4.4.2. Consider M/M/1 queue.

$$\lambda_n = \lambda \qquad \mu_n = \mu \qquad \forall n \geq 1$$

$$\sum_{n=1}^{\infty} \frac{\lambda^n}{\mu_n} = \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu} \right)^n < +\infty \Rightarrow \lambda < \mu$$

$$P_n = \frac{\left(\frac{\lambda}{\mu} \right)^n}{1 + \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu} \right)^n} = \left(\frac{\lambda}{\mu} \right)^n \frac{\mu - \lambda}{\mu}, \quad \forall n \geq 1$$

Example 4.4.3. Consider a job shop consisting of M independent machines and a repairman. Then breakdown rate of a machine is λ and the repair rate is μ. Let $X(t)$ denote the number of machines down at time $t \in \{0, 1, 2, \cdots, n\}$. This is a BAD process with parameters $\mu_n = \mu$, $\lambda_n = \lambda$, $0 \leq n \leq M$.

$$P_0 = \frac{1}{1 + \sum_{n=1}^{M} \frac{M\lambda(M-1)\lambda \cdots (M-n+1)\lambda}{\mu^n}}$$

$$= \frac{1}{1 + \sum_{n=1}^{M} \left(\frac{\lambda}{\mu}\right)^n \frac{M!}{(M-n)!}}$$

$$P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n \frac{M!}{(M-n)!}}{1 + \sum_{n=1}^{M} \left(\frac{\lambda^n}{\mu}\right) \frac{M!}{(M-n)!}}$$

What is long-run proportion of time that a given machine is working?

$P(\text{the machine is working}) =$

$$\sum_{n=0}^{M} P(\text{the machine is working}|n \text{ machines are down}) P_n$$

$$= \sum_{n=0}^{M} \frac{M-n}{M} P_n = \frac{1 + \sum_{n=1}^{M} \left(\frac{\lambda}{\mu}\right)^n \frac{(M-1)!}{(M-n-1)!}}{1 + \sum_{n=1}^{M} \left(\frac{\lambda}{\mu}\right)^n \frac{M!}{(M-n)!}}$$

Chapter 5

Martingales

5.1 Definitions

Definition 5.1.1. A stochastic process $\{Z_n, n \geq 1\}$ is called a martingale if

$$E|Z_n| < +\infty \quad \forall n$$

and

$$E[Z_{n+1}|Z_1, Z_2, \cdots, Z_n] = Z_n$$

Lemma 5.1.1. *For any random variables X, Y, Z.*

(a) $E[X|Y] = E\{E[X|Y, Z]|Y\}$

(b) $E[X|X, Y] = X$

(c) $E[XZ|X, Y] = XE[Z|X, Y]$

Definition 5.1.2. Given random variables X and Y, and an event A. X is said to be *determined* by Y if the value of X is completely determined by that of Y. A is said to be determined by Y if χ_A is determined by Y.

Lemma 5.1.2. *(a)* $E[X|Y] = X$ *if X is determined by Y.*

(b) $E[X|Y, Z] = E[X|Y]$ *if Z is determined by Y.*

Proposition 5.1.1. *For a martingale $\{Z_n, n \geq 1\}$*

$$E[Z_n] = E[Z_1] \quad \forall n$$

Proof.

$$E[Z_n] = E[E[Z_{n+1}|Z_1, Z_2, \cdots, Z_n]] = E[Z_{n+1}]$$

□

Example 5.1.1. Let $X_1, X_2, \cdots$ be independent and identically distributed random variables with 0 mean, and $E|X_i| < +\infty$. Let $Z_n = \sum_{i=1}^{n} X_i$.

$$E|Z_n| \leq \sum_{i=1}^{n} E|X_i| < +\infty \quad \forall n$$

$$\begin{aligned} E[Z_{n+1}|Z_1, Z_2, \cdots, Z_n] &= E[Z_n + X_{n+1}|Z_1, Z_2, \cdots, Z_n] \\ &= E[Z_n|Z_1, Z_2, \cdots, Z_n] + E[X_{n+1}|Z_1, Z_2, \cdots, Z_n] \\ &= Z_n + E[X_{n+1}] = Z_n \end{aligned}$$

Example 5.1.2. Let $X_1, X_2, \cdots$ be independent and identically distributed random variables with $E[X_i] = 1$ and $E|X_i| < +\infty$. Let $Z_n = \prod_{i=1}^{n} X_i$, $n \geq 1$.

$$E|Z_n| = E\left[\prod_{i=1}^{n} |X_i|\right] = \prod_{i=1}^{n} E|X_i| < +\infty$$

$$\begin{aligned} E[Z_n|Z_1, Z_2, \cdots, Z_n] &= E[Z_n X_{n+1}|Z_1, Z_2, \cdots, Z_n] \\ &= Z_n E[X_{n+1}|Z_1, Z_2, \cdots, Z_n] = Z_n \end{aligned}$$

Example 5.1.3. (Doob type martingale) Let $X, Y_1, Y_2, \cdots$ be random variables with $E|X| < +\infty$ and let $Z_n = E[X|Y_1, Y_2, \cdots, Y_n]$ $\forall n \geq 1$.

$$\begin{aligned} E|Z_n| &= E|E[X|Y_1, Y_2, \cdots, Y_n]| \\ &\leq E[E[|X||Y_1, Y_2, \cdots, Y_n]] \\ &= E|X| < +\infty \end{aligned}$$

$$\begin{aligned} E[Z_{n+1}|Y_1, Y_2, \cdots, Y_n] &= E\{E[X|Y_1, Y_2, \cdots, Y_{n+1}]|Y_1, Y_2, \cdots, Y_n\} \\ &= E[X|Y_1, Y_2, \cdots, Y_n] = Z_n \end{aligned}$$

$$\begin{aligned} &E[Z_{n+1}|Z_1, Z_2, \cdots, Z_n] = \\ &\qquad E\{E[Z_{n+1}|Y_1, Y_2, \cdots, Y_n, Z_1, Z_2, \cdots, Z_n]|Z_1, Z_2, \cdots, Z_n\} \\ &\qquad\qquad = E[Z_{n+1}|Y_1, Y_2, \cdots, Y_n] = Z_n \end{aligned}$$

5.2 Stopping Times

Definition 5.2.1. The positive, integer valued, possibly infinite, random variable N is said to be a *random time* for the process $\{Z_n, n \geq 1\}$ if the event $\{N = n\}$ is terminated by $Z_1, Z_2, \cdots, Z_n$ $\forall n$ if $P(N < +\infty) = 1$, then the random time N is said to be a *stopping time*.

Let N be a random time for $\{Z_n, n \geq 1\}$ and $\overline{Z}_n = \begin{cases} Z_n & \text{if } n \leq N \\ Z_N & \text{if } n > N \end{cases}$

$\{\overline{Z}_n, n \geq 1\}$ is called the stopped process of $\{Z_n, n \geq 1\}$.
$\{\overline{Z}_1, \overline{Z}_2, \cdots, \overline{Z}_n\}$ is determined by $\{Z_1, Z_2, \cdots, Z_n\}$.

Proposition 5.2.1. *If N is a random time for the martingale $\{Z_n, n \geq 1\}$, then the stopped process $\{\overline{Z}_n, n \geq 1\}$ is also a martingale.*

Proof. Define $I_n = \chi_{n \leq N}$, which is determined by $Z_1, Z_2, \cdots, Z_{n-1}$.
We have $\overline{Z}_n = \overline{Z}_{n-1} + I_n(Z_n - Z_{n-1})$

$$\begin{aligned} E[\overline{Z}_n | Z_1, Z_2, \cdots, Z_{n-1}] &= E[\overline{Z}_{n-1} + I_n(Z_n - Z_{n-1} | Z_1, Z_2, \cdots, Z_{n-1})] \\ &= \overline{Z}_{n-1} + I_n \underbrace{E[Z_n - Z_{n-1} | Z_1, Z_2, \cdots, Z_{n-1}]}_{0} \\ &= \overline{Z}_{n-1} \end{aligned}$$

□

Theorem 5.2.1. *(The Martingale Stopping Theorem/Optimal Sampling Theorem) Let N be a stopping time for the martingale $\{Z_n, n \geq 1\}$. If either*

(i) $\overline{Z}_n$ are uniformly bounded; or

(ii) N is bounded; or

(iii) $E[N] < +\infty$ and there is $M < +\infty$ such that

$$E[|Z_{n-1} - Z_n| \; |Z_1, \cdots, Z_n] \leq M$$

Then $E[Z_N] = E[Z_1]$.

Proof. Assume (i). By Proposition 5.2.1, $E[\overline{Z}_n] = E[\overline{Z}_1] = E[Z_1]$ $\forall n$. However $P(N < +\infty) = 1$, $P(\overline{Z}_n \to Z_N$, as $n \to +\infty) = 1$

$$\begin{aligned} E[Z_N] = E[\lim_{n\to\infty} \overline{Z}_n] &= \lim_{n\to\infty} E[\overline{Z}_n] \qquad \text{(dominated convergence theorem)} \\ &= E[Z_1] \end{aligned}$$

□

Corollary 5.2.1. *(Wald's Equation) If X_i, $i \geq 1$, are independent and identically distributed random variables with $E|X_i| < +\infty$, and N is a stopping time for $X_1, X_2, \cdots$ with $E[N] < +\infty$. Then*

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X_1]$$

Proof. Let $M = E[X_1] < +\infty$, and

$$Z_n = \sum_{i=1}^{n}(X_i - \mu)$$

$\{Z_n, n \geq 1\}$ is a martingale.

$$\begin{aligned} E[|Z_{n+1} - Z_n| \; |Z_1, \cdots, Z_n] &= E[|X_{n+1} - \mu| \; |Z_1, \cdots, Z_n] \\ &\leq E|X_{n+1}| + \mu < +\infty \end{aligned}$$

$$\begin{aligned} E[Z_N] &= E[X_1 - \mu] = 0 \\ \| \\ E\left[\sum_{i=1}^{N}(X_i - \mu)\right] &= E\left[\sum_{i=1}^{N} X_i - N\mu\right] \\ &= E\left[\sum_{i=1}^{N} X_i\right] - \mu E[N] \end{aligned}$$

□

Example 5.2.1. (Simple Random Walk) Consider an individual who starts at 0 and at each step either moves 1 position to the right with probability p or one to the left with probability $1 - p$. Assume that the successive movements are independent. If $p > 1/2$ find the expected number of steps it takes until the individual reaches position i, $i > 0$.

Solution.

$$X_j = \begin{cases} 1 & \text{step } j \text{ is to the right} \\ -1 & \text{step } j \text{ is to the left} \end{cases}$$

$$N: \; \# \text{ of steps to reach } i$$

$$\sum_{j=1}^{N} X_j = i \Rightarrow i = E[N]E[X_i] = E[N](2p-1)$$

$$E[N] = \frac{i}{(2p-1)}$$

If $p = \frac{1}{2}$, then 0 is recurrent. □

Example 5.2.2. (Three Players) Players X, Y and Z contest the following game. At each stage two of them are randomly chosen in sequence, with the first one chosen being required to give 1 coin to the other. All of the possible choices are equally likely and successive choices are independent of the past.

This continues until one of the players has no remaining coins. At this point that player departs and the other two continue playing until one of them has all the coins. If the players initially have x, y and z coins, respectively, find the expected number of plays until one of them has all the $s = x + y + z$ coins.

Solution. Consider an equivalent game: each player may hold "negative" number of coins. For example, at some time, X, Y and Z have 0, -4 and $s+4$ coins respectively. Let X_i, Y_i and Z_i denote the number of coins X, Y and Z have after ith round. Let T denotes the first time that two of the values X_n, Y_n, Z_n are 0. The question is to find $E[T]$.

First show $M_n = X_nY_n + Y_nZ_n + Z_nX_n + n$ is a Martingale.

Case 1 $X_nY_nZ_n \neq 0$.

$$
\begin{aligned}
&E[X_{n+1}Y_{n+1}|X_n = X, Y_n = Y] \\
&\quad = [(x+1)y + (x+1)(y-1) + x(y+1) + x(y-1) \\
&\quad\quad +(x-1)y + (x-1)(y+1)] \cdot \frac{1}{6} \\
&\quad = x - y - \frac{1}{3}
\end{aligned}
$$

$$
\begin{aligned}
E[M_{n+1}|X_n = x, Y_n = y, Z_n = z] &= (xy - \frac{1}{3}) + (yz - \frac{1}{3}) + (xz - \frac{1}{3}) + (n+1) \\
&= xy + yz + zx - 1 + (n+1) \\
&= xy + yz + zx + n
\end{aligned}
$$

$$
\Rightarrow E[M_{n+1}|X_n, Y_n, Z_n] = M_n
$$

Case 2 $X_nY_nZ_n = 0$. (Assume $X_n = 0$, i.e. one player has quitt ed.)

$$
\begin{aligned}
E[Y_{n+1}Z_{n+1}|Y_n = y, Z_n = z] &= [(y+1)(z-1) + (y-1)(z+1)] \cdot \frac{1}{2} \\
&= yz - 1
\end{aligned}
$$

$$
\Rightarrow E[M_{n+1}|X_n, Y_n, Z_n] = Y_nZ_n - 1 + (n+1) = Y_nZ_n + n = M_n
$$

Since two of X_T, Y_T and Z_T are 0, it follows that $E[T] = E[M_T]$. Consider T is a stopping time, from Theorem 5.2.1

$$
E[M_T] = E[M_0] = xy + yz + zx
$$

□

5.3 Azuma's Inequality

Lemma 5.3.1. *Let X be such that with $E[X]=0$ and $P(-\alpha \le X \le \beta)=1$. Then for any convex function f*

$$E[f(x)] \le \frac{\beta}{\alpha+\beta} f(-\alpha) + \frac{\alpha}{\alpha+\beta} f(\beta)$$

Proof. Let $X = (1-\lambda)(-\alpha) + \lambda\beta$.

$$\lambda = \frac{X+\alpha}{\alpha+\beta} \qquad 1-\lambda = \frac{\beta - X}{\alpha+\beta}$$

$f(x)$ is convex function.

$$\Rightarrow f(x) \le (1-\lambda) f(-\alpha) + \lambda f(\beta)$$
$$\Rightarrow E[f(x)] \le \frac{\beta}{\alpha+\beta} f(-\alpha) + \frac{\alpha}{\alpha+\beta} f(\beta)$$

□

Lemma 5.3.2. *For any $0 \le \theta \le 1$ and $X \in \mathbb{R}$, we have*

$$\theta e^{(1-\theta)x} + (1-\theta)e^{-\theta x} \le e^{x^2/8}$$

Proof. Let $\theta = \frac{1+\alpha}{2}$, $x = 2\beta$, we need to show for $\forall -1 \le \alpha \le 1$, $\beta \in \mathbb{R}$,

$$(1+\alpha)e^{\beta(1-\alpha)} + (1-\alpha)e^{-\beta(1+\alpha)} \le 2e^{\beta^2/2}$$

or, equivalently,

$$e^{\beta} + e^{-\beta} + \alpha(e^{\beta} - e^{-\beta}) \le 2e^{\alpha\beta + \beta^2/2}$$

It is true when $\alpha = \pm 1$ or $|\beta|$ is large, intuitively.
Let $f(\alpha,\beta) = e^{\beta} + e^{-\beta} + \alpha(e^{\beta} - e^{-\beta}) - 2e^{\alpha\beta+\beta^2/2}$.
We need to show $f(\alpha,\beta) \le 0$, for $-1 < \alpha < 1$ and $|\beta| < 100$.
Otherwise, there is a maximum point, $\alpha^* \in (-1,1)$, $\beta^* \in (-100,100)$, such that $f(\alpha^*,\beta^*) > 0$. ($\beta^* \ne 0$, otherwise $f(\alpha^*, 0) = 0$)

$$\Rightarrow \begin{cases} 0 = \left.\dfrac{\partial f}{\partial \alpha}\right|_{\substack{\alpha=\alpha^* \\ \beta=\beta^*}} = e^{\beta^*} - e^{-\beta^*} - 2\beta^* e^{\alpha^*\beta^* + \beta^{*2}/2} \\ 0 = \left.\dfrac{\partial f}{\partial \beta}\right|_{\substack{\alpha=\alpha^* \\ \beta=\beta^*}} = e^{\beta^*} - e^{-\beta^*} + \alpha^*(e^{\beta^*} + e^{-\beta^*}) - 2(\alpha^* + \beta^*)e^{\alpha^*\beta^* + \beta^{*2}/2} \end{cases}$$

$$\Rightarrow 1 + \alpha^* \frac{e^{\beta^*} + e^{-\beta^*}}{e^{\beta^*} - e^{-\beta^*}} = \frac{\alpha^*}{\beta^*} + 1$$
$$\Rightarrow \alpha^* \frac{e^{\beta^*} + e^{-\beta^*}}{e^{\beta^*} - e^{-\beta^*}} = \frac{\alpha^*}{\beta^*}$$

If $\alpha^* = 0$, from $0 = \left.\dfrac{\partial f}{\partial \beta}\right|_{\substack{\alpha=\alpha^* \\ \beta=\beta^*}}$ we have

$$e^{\beta^*} - e^{-\beta^*} = 2\beta^* e^{\beta^{*2}/2}$$

But

$$e^{\beta^*} - e^{-\beta^*} = \sum_{i=0}^{\infty} \frac{\beta^{*i}}{i!} - \sum_{i=0}^{\infty} \frac{(-\beta^*)^i}{i!} = \sum_{i=0}^{\infty} \frac{2\beta^{*(2i+1)}}{(2i+1)!}$$
$$2\beta^* e^{\beta^*/2} = 2\beta^* \sum_{i=0}^{\infty} \frac{(\beta^{*2}/2)^i}{i!} = 2\beta^* \sum_{i=0}^{\infty} \frac{(\beta^*)^{2i}}{2^i \cdot i!} = \sum_{i=0}^{\infty} \frac{(\beta^*)^{2i+1}}{2^{i+1} \cdot i!}$$

Consider $(2i+1)!$ and $2^i \cdot i!$, we have $(2i+1)^i > 2^i \cdot i!$ $\forall i$.
So $e^{\beta^* - e^{-\beta^*}} < 2\beta^* e^{\beta^*/2}$ and $\alpha^* \neq 0$.

So $\dfrac{e^{\beta^*} + e^{-\beta^*}}{e^{\beta^*} - e^{-\beta^*}} = \dfrac{1}{\beta^*}$, i.e. $\beta^*(e^{\beta^*} + e^{-\beta^*}) = e^{\beta^*} - e^{-\beta^*}$.
or, expanding in a Taylor series,

$$\sum_{i=0}^{\infty} \frac{(\beta^*)^{2i+1}}{(2i)!} = \sum_{i=0}^{\infty} \frac{(\beta^*)^{2i+1}}{(2i+1)!}$$

which is clearly not possible when $\beta^* \neq 0$. □

Theorem 5.3.1. *(Azuma's Inequality) Let $\{Z_n, n \geq 1\}$ be a martingale with $\mu = E[Z_n]$. Let $Z_0 = \mu$. Assume for $\alpha_n \geq 0$, $\beta_n \geq 0$, $n \geq 1$,*

$$-\alpha_n \leq Z_n - Z_{n-1} \leq \beta_n$$

Then $\forall n \geq 0$, $a > 0$

$$\begin{cases} P(Z_n - \mu \geq a) \leq \exp\Big[-2a^2 / \sum_{i=1}^n (\alpha_i + \beta_i)^2\Big] \\ P(Z_n - \mu \leq -a) \leq \exp\Big[-2a^2 / \sum_{i=1}^n (\alpha_i + \beta_i)^2\Big] \end{cases}$$

Proof. First assume $\mu = 0$, for any $c \geq 0$.

$$P(Z_n \geq a) = P(e^{cZ_n} \geq e^{ca}) \leq e^{-ca} \cdot E[e^{cZ_n}] \qquad \text{(Markov's inequality)}$$

Let $W_n = e^{cZ_n}$, $n > 0$.
Then $W_n = e^{cZ_{n-1}} e^{c(Z_n - Z_{n-1})} = W_{n-1} e^{c(Z_n - Z_{n-1})}$.

$$E[W_n | Z_{n-1}] = W_{n-1} E[e^{c(Z_n - Z_{n-1})} | Z_{n-1}]$$

Noticing $E[Z_n - Z_{n-1} | Z_{n-1}] = 0$, apply Lemma 5.3.1 to get

$$E[W_n | Z_{n-1}] \leq W_{n-1} \left[\frac{\beta_n}{\alpha_n + \beta_n} e^{-c\alpha_n} + \frac{\alpha_n}{\alpha_n + \beta_n} e^{c\beta_n} \right] \triangleq W_{n-1} Y_n$$

So

$$E[W_n] \le Y_n E[W_{n-1}] \le \cdots \le \prod_{i=1}^{n} Y_i E[W_0] = \prod_{i=1}^{n} Y_i$$
$$= \prod_{i=1}^{n} \left[\frac{\beta_n}{\alpha_n+\beta_n} e^{-c\alpha_n} + \frac{\alpha_n}{\alpha_n+\beta_n} e^{c\beta_n} \right]$$

Let $\theta_i = \frac{\alpha_i}{\alpha_i+\beta_i}$, $x = c(\alpha_i+\beta_i)$, applying Lemma 5.3.2, we have

$$E[W_n] \le \prod_{i=1}^{n} \exp\left(\frac{c^2(\alpha_i+\beta_i)^2}{8} \right)$$

Hence

$$P(Z_n \ge a) \le \exp\left[-ca + \sum_{i=1}^{n} \frac{(\alpha_i+\beta_i)^2}{8} c^2 \right] \triangleq f(c)$$

$f(c)$ achieves minimum at $c^* = \dfrac{4a}{\sum_{i=1}^{n}(\alpha_i+\beta_i)^2} > 0$.

So

$$\begin{cases} P(Z_n \ge a) \le \exp\left[-2a^2/\sum_{i=1}^{n}(\alpha_i+\beta_i)^2\right] \\ P(Z_n \le -a) \le \exp\left[-2a^2/\sum_{i=1}^{n}(\alpha_i+\beta_i)^2\right] \end{cases}$$

For $\mu \ne 0$, consider $\{Z_n - \mu, n \ge 1\}$ and $\{\mu - Z_n, n \ge 1\}$. □

Example 5.3.1. Let $X_1, X_2, \cdots$ be random variables with $E[X_1] = 0$, $E[|X_i|] < +\infty$ and $E[X_i | X_1, \cdots, X_{i-1}] = 0$ for $i \ge 1$.

1. $Z_n = \sum_{i=1}^{n} X_i$ is Martingale (from definition).
2. Now suppose $-\alpha_i \le X_i \le \beta_i$, $\forall i$, $\mu = E[Z_n] = 0$. By Theorem 5.3.1

$$P\left(\sum_{i=1}^{n} X_i \ge a \right) \le \exp\left[-2a^2 / \sum_{i=1}^{n} (\alpha_i+\beta_i)^2 \right]$$
$$P\left(\sum_{i=1}^{n} X_i \le -a \right) \le \exp\left[-2a^2 / \sum_{i=1}^{n} (\alpha_i+\beta_i)^2 \right]$$

Example 5.3.2. Let h be a function such that if the vectors $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$ differ in at most one coordinate, then $|h(x) - h(y)| \le 1$. Let $X_1, \cdots, X_n$ be independent random variables. Then with $X = (X_1, \cdots, X_n)$, we have $\forall a > 0$

(i) $P[h(x) - E[h(x)] \ge a] \le e^{-a^2/2n}$

(ii) $P[h(x) - E[h(x)] \le -a] \le e^{-a^2/2n}$

Proof. Consider the Doob type martingale (Example 5.1.3)

$$Z_i = E[h(X)|X_1, \cdots, X_i]$$

$$\begin{aligned}
&| \ E[h(X)|X_1 = x_1, \cdots, X_i = x_i] - E[h(X)|X_1 = x_1, \cdots, X_{i-1} = x_{i-1}]| \\
&\quad = | \ E[h(x_1, \cdots, x_i, X_{i+1}, \cdots, X_n)] - E[h(x_1, \cdots, x_{i-1}, X_i, \cdots, X_n)]| \\
&\quad = | \ E[h(x_1, \cdots, x_i, X_{i+1}, \cdots, X_n) - h(x_1, \cdots, x_{i-1}, X_i, \cdots, X_n)]| \\
&\quad = E[|h(x_1, \cdots, x_i, X_{i+1}, \cdots, X_n) - h(x_1, \cdots, x_{i-1}, X_i, \cdots, X_n)|] \\
&\quad \leq 1 \quad \text{(only } X_i \text{ may be different in two items)}
\end{aligned}$$

So $|Z_i - Z_{i-1}| \leq 1$ or $-1 \leq Z_i - Z_{i-1} \leq 1$, applying Azuma's inequality with $\alpha_i = \beta_i = 1$ and noticing $Z_n = h(x)$ and $E[Z_n] = E[h(X)]$,

$$P(h(x) - E[h(x)] \geq a) \leq \exp\left(-\frac{2a^2}{\sum_{i=1}^n 2^2}\right) = e^{-a^2/2n}$$

□

5.4 Martingale Convergence Theorem

Definition 5.4.1. A stochastic process $\{Z_n, n \geq 1\}$ with $E[|Z_n|] < +\infty$, $\forall n \geq 1$, is called a *submartingale* if

$$E[Z_{n+1}|Z_1, \cdots, Z_n] \geq Z_n$$

and a *supermartingale* if

$$E[Z_{n+1}|Z_1, \cdots, Z_n] \leq Z_n$$

Theorem 5.4.1. *If N is a stopping time for $\{Z_n, n \geq 1\}$ such that any one of the three condition (i) - (iii) of Theorem 5.2.1 satisfies, then*

$$\begin{cases} E[Z_N] \geq E[Z_1] & \textit{for a submartingale} \\ E[Z_N] \leq E[Z_1] & \textit{for a supermartingale} \end{cases}$$

Lemma 5.4.1. *If $\{Z_i, i \geq 1\}$ is a submartingale and N be a stopping time such that $P(N \leq n) = 1$ for a given n, then $E[Z_1] \leq E[Z_N] \leq E[Z_n]$.*

Proof. Only need to show $E[Z_N] \leq E[Z_n]$.

$$\begin{aligned}
E[Z_n] &= E[E[Z_n|Z_1, \cdots, Z_N]] \\
&= \sum_{k \leq n} E[E[Z_n|Z_1, \cdots, Z_N]|N = k]P(N = k) \\
&= \sum_{k \leq n} E[E[Z_n|Z_1, \cdots, Z_k]|N = k]P(N = k) \\
&\geq \sum_{k \leq n} E[Z_k|N = k]P(N = k) \\
&= E[Z_N]
\end{aligned}$$

□

Lemma 5.4.2. *If $\{Z_n, n \geq 1\}$ is a martingale and f is a convex function with $E[|f(Z_n)|] < +\infty$, $\forall n$, then $\{f(Z_n), n \geq 1\}$ is a submartingale.*

Proof.

$$E[f(Z_{n+1})|Z_1, \cdots, Z_n] \geq f(E[Z_n|Z_1, \cdots, Z_n]) = f(Z_n)$$

□

Theorem 5.4.2. *(Kolmogorov's Inequality for Submartingale) If $\{Z_n, n \geq 1\}$ is a nonnegative submartingale, then $\forall a > 0$*

$$P(\max(Z_1, \cdots, Z_n) \geq a) \leq \frac{E[Z_n]}{a}$$

Proof. Define $N = \min\{i : Z_i > a, i \leq n\}$ ($N = n$, if $Z_i \leq a$, $\forall i \leq n$). Then N is a stopping time with $P(N \leq n) = 1$. Now from Lemma 5.4.1

$$P(\max\{Z_1, \cdots, Z_n\} \geq a) \leq P(Z_n \geq a) \leq \frac{E[Z_N]}{a} \leq \frac{E[Z_n]}{a}$$

□

Corollary 5.4.1. *Let $\{Z_n, n \geq 1\}$ be a martingale, then $\forall a > 0$*

(i) $P(\max\{|Z_1|, \cdots, |Z_n|\} > a) \leq \dfrac{E|Z_n|}{a}$

(ii) $P(\max\{|Z_1|, \cdots, |Z_n|\} > a) \leq \dfrac{E[Z_n^2]}{a^2}$

Proof. Parts (i) and (ii) follow from Lemma 6.4.4 and Kolmogrov's inequality since the functions $f(x) = |x|$ and $f(x) = x^2$ are both convex. □

Theorem 5.4.3. *(The Martingale Convergence Theorem) If $\{Z_n, n \geq 1\}$ is a martingale such for some $M < +\infty$*

$$E|Z_n| \leq M \quad \forall n$$

then, with probability 1, $\lim_{n\to\infty} Z_n$ exists and is finite.

Proof. Assume that $E|Z_n|^2 \leq M$ $\forall n$. By Lemma 5.4.2, $\{Z_n^2, n \geq 1\}$ is a submartingale. So $E|Z_n|^2$ is nondecreasing. However $E|Z_n|^2 \leq M$, hence, $\lim_{n\to\infty} |Z_n|^2 = \mu$ exists. Now we want to show that $\{Z_n, n \geq 1\}$ is a Cauchy sequence with probability 1, i.e. with probability 1

$$|Z_{m+k} - Z_m| \to 0 \qquad \text{as } m, k \to \infty$$

Now fix $\varepsilon > 0$ and $n > 1$

$$
\begin{aligned}
& P(|Z_{m+k} - Z_m| > \varepsilon \text{ for some } k \leq n) \\
& \leq \frac{E|Z_{m+n} - Z_n|}{\varepsilon^2} \qquad \text{(Corollary 5.4.1 - (ii))} \\
& = \frac{1}{\varepsilon^2} E[Z_{m+n}^2 - 2Z_{m+n}Z_m + Z_m^2] \qquad (*)
\end{aligned}
$$

$$
\begin{aligned}
E[Z_{m+n}Z_n] &= E[E[Z_{m+n}Z_m|Z_m]] \\
&= E[Z_m E(Z_{m+n}|Z_n)] \\
&= E[Z_m^2]
\end{aligned}
$$

$$(*) \leq \frac{1}{\varepsilon^2} E[Z_{m+n}^2 - Z_m^2]$$

Leaving $n \to \infty$

$$P(|Z_{m+k} - Z_m| > \varepsilon \text{ for some } k) \leq \frac{1}{\varepsilon^2}(\mu - E[Z_m^2])$$

Leaving $m \to \infty$

$$P(|Z_{m+k} - Z_m| > \varepsilon \text{ for some } k) \to 0$$

Thus, with probability 1, Z_n will be a Cauchy sequence, and thus $\lim_{n\to\infty} Z_n$ will exists and be finite. □

Lemma 5.4.3. *If $\{Z_n, n \geq 1\}$ is a nonnegative martingale, then, with probability 1, $\lim_{n\to\infty} Z_n$ exists and is finite.*

Proof.

$$E|Z - n| = E[Z_n] = E[Z_1] \leq +\infty \qquad \forall n$$

□

Example 5.4.1. Consider a gamble playing a fair game whose fortune is Z_n after nth play. On each gamble at least 1 unit is either lost or won and no credit is given.

Let $N = \min\{n : Z_n = Z_{n+1}\}$ denoting the number of play until the gamble is broke. Since $\{Z_n, n \geq 1\}$ is a nonnegative martingale, $\lim_{n\to\infty} Z_n$ exists and is finite with probability 1. However $|Z_{n+1} - Z_n| \geq 1$, $\forall n < N$

$$\Rightarrow P(N < \infty) = 1$$

Chapter 6

Brownian Motions

6.1 Definitions and Basic Properties

Definition 6.1.1. A continuous-time stochastic process $\{X(t), t \geq 0\}$ is called a Brownian motion (or Wiener process) if

(i) $X(0) = 0$

(ii) $\{X(t), t \geq 0\}$ has stationary and independent increments.

(iii) $X(t) \sim N(0, c^2 t)\ \forall t > 0$

It is called a standard Brownian motion if $c = 1$ (which we will assume throughout).

$$X(t) = \Delta x (X_1 + X_2 + \cdots + X_{[\frac{t}{\Delta t}]})$$

where $X_i = \begin{cases} +1 & \text{if the } i\text{th step of length } \Delta x \text{ is to the right} \\ -1 & \text{if it to the left} \end{cases}$

$$E[X(t)] = 0$$
$$\mathrm{Var}[X(t)] = (\Delta x)^2 \cdot 1 \cdot \left[\frac{t}{\Delta t}\right]$$

Take $\Delta x = \sqrt{\Delta t}$, then $\mathrm{Var}[X(t)] = t$.

Basic properties of Brownian motion $\{X(t), t \geq 0\}$

(a) $X(t)$ is a continuous function of t with probability 1.

(b) $X(t)$ is nowhere differentiable with probability 1.

(c) $X(t)$ is Markovian.

$$\begin{aligned}
&P(X(t+s) \le a | X(s) = x, X(u), 0 \le u < s) \\
&= P(X(t+s) - X(s) \le a - x | X(s) = x, X(u), 0 \le u < s) \\
&= P(X(t+s) - X(s) \le a - x) \\
&= P(X(t+s) - X(s) \le a - x | X(s) = x) \\
&= P(X(t+s) \le a | X(s) = x)
\end{aligned}$$

(d) Consider the joint distribution of $X(t_1), X(t_2), \cdots, X(t_n)$ where $0 \le t_1 < t_2 < \cdots < t_n$.

First of all, the density function of $X(t)$

$$f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$$

The joint density of $X(t_1), X(t_2), \cdots, X(t_n)$

$$f(x_1, x_2, \cdots, x_n) = f_{t_1}(x_1) f_{t_2 - t_1}(x_2 - x_1) \cdots f_{t_n - t_{n-1}}(x_n - x_{n-1})$$

(e) For $s \le t$, the covariance

$$\begin{aligned}
\operatorname{cov}(X(s), X(t)) &= \operatorname{cov}(X(s), X(s) + X(t) - X(s)) \\
&= \operatorname{cov}(X(s), X(s)) + \operatorname{cov}(X(s), X(t) - X(s)) \\
&= s
\end{aligned}$$

(f) Fix $t_1 \ge 0$, $X(t + t_1) - X(t)$ is still a Brownian motion.

Example 6.1.1. Compute the conditional distribution of $X(s)$ given $X(t) = B$ where $s < t$.

Solution. The conditional density

$$\begin{aligned}
f(x|B) &= \frac{f(x, B)}{f_{x(t)}(B)} = \frac{f_s(x) f_t(B - x)}{f_t(B)} \\
&= \frac{\frac{1}{2\pi\sqrt{s(t-s)}}}{\frac{1}{\sqrt{2\pi t}}} \exp\left\{ -\frac{x^2}{2s} - \frac{(B-x)^2}{2(t-s)} + \frac{B^2}{2t} \right\} \\
&= \frac{1}{\sqrt{2\pi \frac{s}{t}(t-s)}} \exp\left\{ -\frac{(x - \frac{BS}{t})^2}{2\frac{s}{t}(t-s)} \right\}
\end{aligned}$$

It is a normal distribution with

$$\begin{aligned}
E[X(s)|X(t) = B] &= B\frac{s}{t} \\
\Rightarrow E[X(s)|X(t)] &= \frac{s}{t} X(t) \\
\operatorname{Var}[X(s)|X(t) = B] &= \frac{s}{t}(t-s)
\end{aligned}$$

□

6.2 Hitting Times

Given a Brownian motion $\{X(t), t \geq 0\}$ and $a \in R^+$, $T_a = \inf\{t \geq 0 : X(t) = a\}$.

If $a > 0$

$$\begin{aligned} P(X(t) \geq a) &= P(X(t) \geq a | T_a \leq t) P(T_a \leq t) \\ &\quad + P(X(t) \geq a | T_a < t) P(T_a < t) \\ &= \frac{1}{2} P(T_a \leq t) \end{aligned}$$

$$\begin{aligned} P(T_a \leq t) &= 2P(X(t) \geq a) \\ &= 2 \cdot \frac{1}{\sqrt{2\pi t}} \int_a^{+\infty} e^{-y^2/2t} dy \\ &= \frac{2}{\sqrt{2\pi t}} \int_{a/\sqrt{t}}^{+\infty} e^{-y^2/2} dy \end{aligned}$$

$$\begin{aligned} E[T_n] &= \int_0^{\infty} P(T_a > t) dt \\ &= \int_0^{\infty} \left(1 - \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^{+\infty} e^{-y^2/2} dy \right) dt \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \int_0^{a/\sqrt{t}} e^{-y^2/2} dy\, dt \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-y^2/2} \int_0^{a^2/y^2} dt\, dy \\ &= \frac{2a^2}{\sqrt{2\pi}} \int_0^{\infty} \frac{e^{-y^2/2}}{y^2} dy \\ &\geq \frac{2a^2}{\sqrt{2\pi}} \int_0^{1} \frac{e^{-y^2/2}}{y^2} dy \\ &\geq \frac{2a^2}{\sqrt{2\pi}} e^{-1/2} \int_0^{1} \frac{1}{y^2} dy \\ &= \infty \end{aligned}$$

If $a < 0$, then by symmetry

$$\begin{aligned} P(T_a \leq t) &= P(T_{-a} \leq t) \\ &= \frac{2}{\sqrt{2\pi}} \int_{|a|/\sqrt{t}}^{+\infty} e^{-y^2/2} dy \end{aligned}$$

For $a > 0$

$$P\left(\max_{0 \leq s \leq t} X(s) \geq a \right) = P(T_a \leq t)$$

Let $0(t_1, t_2)$, where $t_2 > t_1 \geq 0$, denote the event that the Brownian motion takes on 0 at least once in the interval (t_1, t_2).

$$P(0(t_1, t_2)) = \frac{1}{\sqrt{2\pi t_1}} \int_{-\infty}^{\infty} P(0(t_1, t_2)|X(t_1) = x) e^{-x^2/2t_1} dx$$

Since $Y(t) = X(t_1 + t) - X(t_1)$ is a Brownian motion

$$P(0(t_1, t_2)|X(t_1) = x) = P(T_{|x|} \leq t_2 - t_1)$$

So

$$\begin{aligned}
&P(0(t_1, t_2)|X(t_1) = x) \\
&= \frac{2}{\sqrt{2\pi t_1}} \int_0^{\infty} \frac{2}{\sqrt{2\pi(t_1 - t_2)}} \int_x^{\infty} e^{-y^2/2(t_2 - t_1)} dy \cdot e^{-x^2/2t_1} dx \\
&\qquad = 1 - \frac{2}{\pi} \arcsin \sqrt{\frac{t_1}{t_2}}
\end{aligned}$$

Proposition 6.2.1. *For $0 < x < 1$ and $t > 0$,*

$$P(\textit{Brownian motion takes no zero value in } (xt, t)) = \frac{2}{\pi} \arcsin \sqrt{x}$$

6.3 Variation on Brownian Motion

Model 1. (Brownian Motion Absorbed at a Point) Let $\{X(t), t \geq 0\}$ be a Brownian motion and $x > 0$. Define

$$Z(t) = \begin{cases} X(t) & \text{if } t < T_x \\ x & \text{if } t \geq T_x \end{cases}$$

then $\{Z(t), t \geq 0\}$ is Brownian motion that when it hits x remains there forever.

$$P(Z(t) = x) = P(T_x \leq t) = \frac{2}{\sqrt{2\pi t}} \int_x^{\infty} e^{-y^2/2t} dy$$

$$\begin{aligned}
P(Z(t) \leq y) &= P\left(X(t) \leq y, \max_{0 \leq s \leq t} X(s) < x\right) \\
&= P(X(t) \leq y) - P\left(X(t) \leq y, \max_{0 \leq s \leq t} X(s) \geq x\right)
\end{aligned}$$

$$\begin{aligned}
&P\left(X(t) \leq y, \max_{0 \leq s \leq t} X(s) \geq x\right) \\
&\qquad = \underbrace{P\left(X(t) \leq y \middle| \max_{0 \leq s \leq t} X(s) \geq x\right)}_{I} P\left(\max_{0 \leq s \leq t} X(s) \geq x\right)
\end{aligned}$$

$$I = P\left(X(t) \geq x + (x-y) \Big| \max_{0\leq s\leq t} X(s) \geq x\right)$$

So

$$\begin{aligned}
P\left(X(t) \leq y, \max_{0\leq s\leq t} X(s) \geq x\right) &= P\left(X(t) \geq 2x - y, \max_{0\leq s\leq t} X(s) \geq x\right) \\
&= P(X(t) \geq 2x - y) \\
&= P(X(t) \leq y - 2x)
\end{aligned}$$

$$\begin{aligned}
P(Z(t) \leq y) &= (y - 2x < X(t) \leq y) \\
&= \frac{1}{\sqrt{2\pi t}} \int_{y-2x}^{y} e^{-u^2/2t} du
\end{aligned}$$

Model 2. (Brownian Motion Reflected at the Origin) If $\{X(t), t \geq 0\}$ is a Brownian motion, then $\{Z(t), t \geq 0\} = \{|X(t)|, t \geq 0\}$ is called a Brownian motion reflected at origin.

$$\begin{aligned}
P(Z(t) \leq y) &= P(|X(t)| \leq y) \\
&= P(-y \geq X(t) \leq y) \\
&= P(X(t) \leq y) - P(X(t) \leq -y) \\
&= P(X(t) \leq y) - \underbrace{P(X(t) \geq y)}_{1-P(X(t)\leq y)} \\
&= 2P(X(t) \leq y) - 1 \\
&= \frac{2}{\sqrt{\pi t}} \int_{-\infty}^{y} e^{-x^2/2t} dx - 1
\end{aligned}$$

$$\begin{aligned}
E[Z(t)] &= \sqrt{\frac{2t}{\pi}} \\
\mathrm{Var}[Z(t)] &= (1 - \frac{2}{\pi})t \quad (< t)
\end{aligned}$$

Model 3. (Geometric Brownian motion) If $\{X(t), t \geq 0\}$ is Brownian motion, then $\{Y(t), t \geq 0\} = \{e^{X(t)}, t \geq 0\}$ is called geometric Brownian motion.

Moment generation function of $X(t)$

$$\psi(s) = Ee^{sX(t)} = e^{ts^2/2}$$

and therefore

$$E[Y(t)] = E[e^{X(t)}] = e^{t/2}$$

$$\begin{aligned}\text{Var}[Y(t)] &= E[Y(t)^2] - e^t \\ &= Ee^{2X(t)} - e^t \\ &= e^{2t} - e^t\end{aligned}$$

Model 4. (Integrated Brownian motion) If $\{X(t), t \geq 0\}$ is Brownian motion, then

$$\{Z(t), t \geq 0\} = \left\{ \int_0^t X(s)ds, t \geq 0 \right\}$$

is called integrated Brownian motion.

$$E[Z(t)] = E\int_0^t X(s)ds = \int_0^t E[X(s)]ds = 0$$

For $s \leq t$

$$\begin{aligned}\text{cov}(Z(s), Z(t)) &= E[Z(s)Z(t)] \\ &= E\left[\int_0^s X(y)dy \int_0^t X(u)du\right] \\ &= \int_0^s \int_0^t E[X(y)X(u)]dy\,du \\ &= \int_0^s \int_0^t \min(y,u)dy\,du \\ &= \int_0^s \left[\int_0^u ydy + \int_u^t udy\right] du \\ &= \int_0^s \left[\frac{1}{2}u^2 + u(t-u)\right] du \\ &= s^2(\frac{t}{2} - \frac{s}{6})\end{aligned}$$

$$\text{Var}[Z(t)] = t^2(\frac{t}{2} - \frac{t}{6}) = \frac{t^3}{3}$$

6.4 Brownian Motion with Drift

Definition 6.4.1. A stochastic process $\{X(t), t \geq 0\}$ is a Brownian motion with drift coefficient μ if

(i) $X(0) = 0$;

(ii) $\{X(t), t \geq 0\}$ has stationary and independent increments;

(iii) $X(t) \sim N(\mu t, t)$.

Or $X(t) = \mu t + B(t)$ where $\{B(t), t \geq 0\}$ is standard Brownian motion.

Example 6.4.1. Consider a drifted Brownian motion $\{X(t), t \geq 0\}$ with drift μ. Compute the probability that the process hits A before $-B$, where $A, B > 0$ are given.

Solution. Let $P(x) = P\{X(t) \text{ hits } A \text{ before } -B | X(0) = x\}$, where $-B < x < A$. Then

$$\begin{aligned}
P(x) &= \int_{-\infty}^{\infty} P(X(t) \text{ hits } A \text{ before } -B|(X(0) = x, Y = y) dF_Y(y) + o(h) \\
&= \int_{-\infty}^{\infty} P(X(t) \text{ hits } A \text{ before } -B|X(0) = x, X(h) = x + y) dF_Y(y) + o(h) \\
&= \int_{-\infty}^{\infty} P(x + y) dF_Y(y) + o(h) \\
&= E[P(x + y)] + o(h) \\
&= E[P(x) + P'(x)Y + \frac{P''(X)}{2} Y^2 + \cdots] + o(h) \\
&= P(x) + \mu h P'(x) + P''(x) \frac{\mu^2 h^2 + h}{2} + o(h)
\end{aligned}$$

$$\begin{aligned}
\mu P'(x) + \frac{1}{2} P''(x) &= 0 \\
2\mu P(x) + P'(x) &= c_1 \\
\frac{d}{dx}[e^{2\mu x} P(x)] &= c_1 e^{2\mu x} \\
P(x) &= C_1 + C_2 e^{2\mu x}
\end{aligned}$$

With the boundary conditions that $P(A) = -1$, $P(B) = 0$, we have

$$C_1 = \frac{e^{2\mu B}}{e^{2\mu B} - e^{-2\mu A}} \quad C_2 = \frac{-1}{e^{2muB} - e^{-2\mu A}}$$

$$P(x) = \frac{e^{2\mu B} - e^{-2\mu x}}{e^{2\mu B} - e^{-2\mu A}}$$

Starting at $x = 0$, the probability of reaching A before $-B$

$$P(0) = \frac{e^{2\mu B} - 1}{e^{2\mu B} - e^{-2\mu A}}$$

If $\mu < 0$, letting $B \to \infty$

$$P(\text{the process ever goes up to } A) = e^{2\mu A}$$

$$P\left(\max_{t \geq 0} X(t) < y\right) = 1 - P\left(\max_{t \geq 0} X(t) \geq y\right) = 1 - e^{2\mu A}$$

Therefore $\max_{t \geq 0} X(t)$ is an exponential random variable with parameter -2μ. $\square$

Example 6.4.2. A stock call option with an exercise price A. The current price of the stock of the stock is 0 and the price follows Brownian motion with drift $-d$, $d > 0$. When should we exercise the option?

Solution. Suppose we exercise when the price is $x(> A)$. The expected gain is $P(x)(x - A)$, where $P(x)$ is the probability that the price ever hits A.

The gain function

$$f(x) = P(x)(x - A) = e^{-2dx}(x - A)$$

Letting $f'(x) = 0$ we have the maximum point $x = A + \frac{1}{2d}$.

□

6.5 Martingale and Brownian motion

Definition 6.5.1. A continuous-time process $\{X(t), t \geq 0\}$ is called a martingale if $E|X(t)| < \infty \ \forall t > 0$ and $E[X(t)|X(u), 0 \leq u \leq s] = X(s) \ \forall t \geq s$.

Proposition 6.5.1. *Let $\{B(t), t \geq 0\}$ be a standard Brownian motion. Then all the following processes are martingale:*

(a) $Y(t) = B(t)$

(b) $Y(t) = B(t)^2 - t$

(c) $Y(t) = \exp\{cB(t) - \frac{c^2 t}{2}\} \ \forall c \in R$

Proof. (a)

$$\begin{aligned} E[B(t)|B(u), 0 \leq u \leq s] &= E[B(s) + B(t) - B(s)|B(u), 0 \leq u \leq s] \\ &= B(s) + E[B(t) - B(s)] \\ &= B(s) \end{aligned}$$

(b)

$$\begin{aligned} E[B(t)^2|B(u), 0 \leq u \leq s] &= E[B(s)^2 + 2B(s)(B(t) - B(s)) \\ &\quad + (B(t) - B(s))^2|B(u), 0 \leq u \leq s] \\ &= B(s)^2 + 2B(s)E[B(t) - B(s)|B(u), 0 \leq u \leq s] \\ &\quad + E[B(t) - B(s)]^2 \\ &= B(s)^2 + t - s \end{aligned}$$

(c) Left as as exercise.

□

Definition 6.5.2. A random variable $\tau \geq 0$ is called a stopping time for a stochastic process $\{X(t), t \geq 0\}$ if the event $\{\tau \leq t\}$ is determined by $\{X(s), 0 \leq s \leq t\} \ \forall t \geq 0$.

Theorem 6.5.1. *(Martingale Stopping Theorem / Optimal Sampling Theorem) Let τ be a stopping time for a martingale $\{X(t), t \geq 0\}$ satisfying either*

(i) τ is uniformly bounded, or

(ii) $P(\tau < +\infty) = 1$ and $\forall t \geq 0$

$$|X\underbrace{(t \wedge \tau)}_{\min(t,\tau)}| \leq K$$

Then $E[X(s)] = E[X(0)]$.

Example 6.5.1. $X(t) = B(t) + \mu t$ where $B(t)$ is a standard Brownian motion. For $A, B > 0$ define stopping time $T = \min\{t > 0 : X(t) = A$ or $X(t) = -B\}$. What is $P_A = P(X(\tau) = A)$?

Solution. By optimal sample, $E\exp\{cB(T) - \frac{c^2T}{2}\} = 1$.

$$E[\exp\{cX(T) - c\mu T - \frac{cT^2}{2}\}] = 1$$

Take $c = -2\mu$, $E\exp\{-2\mu X(T)\} = 1$.

$$e^{-2\mu A}P_A + e^{-2\mu(-B)}(1 - P_A) = 1$$

$$P_A = \frac{e^{2\mu B} - 1}{e^{-2\mu B} - e^{-2\mu A}}$$

Now $\{B(t), t \geq 0\}$ is martingale.

$$0 = EB(T) = E[X(t) - \mu T]$$

$$\begin{aligned} E[T] &= \frac{1}{\mu}(AP_A - BP_B) \\ &= \frac{Ae^{2\mu B} + Be^{-2\mu A} - A - B}{\mu[e^{2\mu B} - e^{-2\mu A}]} \end{aligned}$$

□

Chapter 7

Introduction to Itô's Calculus

7.1 Stochastic Integration

Given a (standard) Brownian motion $\{B(t), t \geq 0\}$. For fixed $T > 0$, define

$$L^2_{\mathcal{F}}(0,T;\mathbb{R}) = \Big\{\{X(t), 0 \leq t \leq T\}\Big| \forall t \in [0,T],$$
$$X(t) \text{ is determined by } B(s),\ 0 \leq s \leq t, \text{ and } E\int_0^T |X(s)|^2 ds < +\infty\Big\}$$

For X_n, $n = 1, 2, \cdots$, $X \in L_{\mathcal{F}}(0,T:\mathbb{R})$, we say $X_n \xrightarrow{L^2} X$ as $n \to \infty$ if $E\int_0^T |X_n(s) - X(s)|^2 ds \to 0$ as $n \to \infty$. We are now to define *stochastic integration* $\int_0^t f(s)dB(s)$ for $f \in L^2_{\mathcal{F}}(0,T;\mathbb{R})$.

Step 1. If f is a *simple process*

$$f(t) = f_0\chi_{\{t=0\}}(t) + \sum_{i=0}^{k-1} f_i\chi_{(t_i,t_{i+1}]}(t)$$

where $0 = t_0 < t_1 < \cdots < t_k = T$, f_i is a random variable determined by $B(s)$, $0 \leq s \leq t_i$. Then define

$$\int_0^t f(s)dB(s) = \sum_{i=0}^{j-1} f_i[B(t_{i+1}) - B(t_i)] + f_j[B(t) - B(t_j)]$$

if $t \in (t_j, t_{j+1}]$.

Step 2. For any $f \in L_{\mathcal{F}}(0,T;\mathbb{R})$, $\exists\{f_n\}$ that are simple process such that $f_n \xrightarrow{L^2} f$ as $n \to \infty$. Define

$$\int_0^t f(s)dB(s) = \lim_{n\to\infty} \int_0^t f_n(s)dB(s)$$

$\int_0^t f(s)dB(s)$ is called the Itô's integral. For any $s \le t \in [0,T]$

$$\int_s^t f(r)dB(r) = \int_0^t f(r)dB(r) - \int_0^s f(r)dB(r)$$

Theorem 7.1.1. $\forall f, g \in L^2_{\mathcal{F}}(0,T;\mathbb{R})$

(i) $E\Big[\int_s^t f(r)dB(r)\Big|B(u), 0 \le u \le s\Big] = 0$, $\forall 0 \le s \le t \le T$

(ii) $\forall 0 \le s \le t \le T$

$$E\Big[\int_s^t f(r)dB(r) \cdot \int_s^t g(r)dB(r)\Big|B(u), 0 \le u \le s\Big] \\ = E\Big[\int_s^t f(r)g(r)dr\Big|B(u), 0 \le u \le s\Big]$$

(iii) $E\int_0^t f(r)dB(r) = 0$, $\forall t \in [0,T]$

(iv) $E\Big[\int_0^t f(r)dB(r)\int_0^t g(r)dB(r)\Big] = E\int_0^t f(r)g(r)dr$, $\forall t \in [0,T]$

(v) $E\Big|\int_0^t f(r)dB(r)\Big|^2 = E\int_0^t |f(r)|^2 dr$

Remark. By (i), $X(t) = \int_0^t f(r)dB(r)$ is a martingale.

$$E[X(t)|B(u), 0 \le u \le s] = E\left[\int_0^s f(r)dB(r) + \int_s^t f(r)dB(r)\Big|B(u), 0 \le u \le s\right] \\ = \int_0^s f(r)dB(r) = X(s)$$

Definition 7.1.1. A multi-dimensional stochastic process $\{B(t) = (B_1(t), \cdots, B_m(t))^T\}$ is called a Brownian motion if

(i) $B(0) = 0$;

(ii) $\{B(t), t \ge 0\}$ has stationary and independent increments;

(iii) $EB(t) = 0$, $E[B(t)B(t)^T] = tI_{m\times m}$ $\forall t$.

Given an m-dim Brownian motion $B(t)$ and $f = (f_1, \cdots, f_m) \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^m)$

$$\int_0^t f(s)dB(s) = \sum_{i=1}^m \int_0^t f_i(s)dB_i(s)\ \forall t$$

Similarly for $\int_0^t \underbrace{\sigma(s)}_{n\times m} d\underbrace{B(s)}_{m\times 1}$ where $\sigma(s)$ is $n \times m$ matrix-valued process.

7.2 Itô's Formula

Recall the deterministic case. If $X(t) = X(0) + \int_0^t b(s)ds$ (or $dX(t) = b(t)dt$), given $F(t,x) \in C^1([0,T] \times \mathbb{R})$

$$\begin{aligned} dF(t,X(t)) &= \frac{\partial F}{\partial t}(t,X(t))dt + \frac{\partial F}{\partial X}(t,X(t))dX(t) \\ &= \left(\frac{\partial F}{\partial t} + \frac{\partial F}{\partial X}b(t)\right)dt \end{aligned}$$

or equivalently

$$F(t,X(t)) = F(0,X(0)) + \int_0^t \left(\frac{\partial F}{\partial t}\big(s,X(s)\big) + \frac{\partial F}{\partial X}\big(s,X(s)\big)b(s)\right)ds$$

Theorem 7.2.1. *(Itô's Formula) Let an n-dim stochastic process $\{X(t), t \geq 0\}$ be given as*

$$X(t) = X(0) + \int_0^t b(s)ds + \underbrace{\int_0^t \sigma(s)dB(s)}_{\text{diffusion term}}$$

where $b \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^2)$, $\sigma \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^{n\times m})$. Let $F \in C^{1,2}([0,T]\times\mathbb{R}^2)$ be given.

$$\begin{aligned} dF(t,X(t)) = F_t\big(t,X(t)\big)dt + F_X\big(t,X(t)\big)dX(t) \\ + \frac{1}{2}\Big[\sigma(t)^T F_{XX}\big(t,X(t)\big)\sigma(t)dt\Big]dt \end{aligned}$$

or

$$\begin{aligned} F(t,X(t)) = F\big(0,X(0)\big) + \int_0^t \Big\{F_t\big(t,X(t)\big) + F_X\big(t,X(t)\big) \\ + \frac{1}{2}\operatorname{Tr}\Big[\sigma(t)^T F_{XX}\big(t,X(t)\big)\sigma(t)dt\Big]\Big\}ds + \int_0^t F_X\big(s,X(s)\big)^T\sigma(s)dB(s) \end{aligned}$$

where $F_X = \left(\frac{\partial F}{\partial X_1} \cdots \frac{\partial F}{\partial X_n}\right)^T \quad F_{XX} = \left(\frac{\partial^2 F}{\partial X_i \partial X_j}\right)_{n\times n}$

Example 7.2.1. Calculate $E|\int_0^t \sigma(s)dB(s)|^2$.

Solution. Let $X(t) = \int_0^t \sigma(s)dB(s)$, $F(t,x) = x^2$.

$$X(t)^2 = F(t,X(t)) = \int_0^t \sigma(s)^2 ds + \int_0^t 2X(s)\sigma(s)dB(s)$$

$$E|X(t)^2| = E\int_0^t \sigma(s)^2 ds$$

□

Example 7.2.2. Calculate $Ee^{B(t)}$.

Solution. Let $X(t) = B(t) = \int_0^t dB(s)$, $F(t,x) = e^x$.

$$e^{X(t)} = F(t, X(t)) = 1 + \int_0^t \frac{1}{2} e^{X(s)} ds + \int_0^t e^{X(s)} dB(s)$$

$$E[e^{X(t)} = 1 + \frac{1}{2} \int_0^t E[e^{X(s)}]ds]$$

Let $y(t) = E[e^{X(t)}]$

$$y(t) = 1 + \frac{1}{2} \int_0^t y(s) ds$$

$$\Rightarrow \begin{cases} y'(t) = \frac{1}{2} y(t) \\ y(0) = 1 \end{cases}$$

$$y(t) = e^{\frac{1}{2}t}$$

□

Example 7.2.3. Let $dX_i(t) = b_i(t)dt + \sigma_i(t)dB(t)$, where $i = 1, 2$. Calculate $d[X_1(t)X_2(t)]$.

Solution. Take $X(t) = \big(X_1(t), X_2(t)\big)^T$, $F(t, x_1, x_2) = x_1 x_2$.

$$\begin{aligned} &d[X_1(t)X_2(t)] \\ &= dF(t, X_1(t), X_2(t)) \\ &= \left[X_2(t)b_1(t) + X_1(t)b_2(t) + \frac{1}{2}\Big(\sigma_1(t) + \sigma_2(t)\Big) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_1(t) \\ \sigma_2(t) \end{pmatrix}\right] dt \\ &\quad + \big[X_2(t)\sigma_1(t) + X_1(t)\sigma_2(t)\big] dB(t) \\ &= X_1(t)dX_2(t) + X_2(t)dX_1(t) + \sigma_1(t)\sigma_2(t)dt \end{aligned}$$

□

7.3 Stochastic Differential Equations

Consider the following stochastic differential equation

$$(*) = \begin{cases} dX(t) = b(t, X(t))dt + \sigma(t, X(t))dB(t) \\ X(0) = X_0 \end{cases}$$

or

$$X(t) = x_0 + \int_0^t b\big(s, t(s)\big) ds + \int_0^t \sigma\big(s, X(s)\big) dB(s)$$

Definition 7.3.1. A stochastic process $\{X(t) \in \mathbb{R}^n, t \geq 0\}$ is called a solution to $(*)$ if

(i) $X(t)$ is $\mathcal{F}_t$-adapted.

(ii) $P\big(X(0) = X_0\big) = 1$

(iii) $E\int_0^t \Big[\big|b(s), x(s)\big| + \big|\sigma(s, X(s)\big|^2\Big] < \infty$, $\forall t > 0$

(iv) $P\left(X(t) = x_0 + \int_0^t b\big(s, X(s)\big)ds + \int_0^t \sigma\big(s, X(s)\big)dB(s)\ \forall t \geq 0\right) = 1$

Moreover, $(*)$ is said to have a *unique solution* if for any two solutions $x(t), y(t)$ one must have $P(x(t) = y(t)\ \forall t) = 1$.

Theorem 7.3.1. *If there exists $L > 0$ such that*

$$\begin{aligned} \big|b(t,x) - b(t,y)\big| &\leq L|x-y| \\ \big|\sigma(t,x) - \sigma(t,y)\big| &\leq L|x-y| \qquad \forall x, y \qquad \textit{(Lipschitz condition)} \\ \big|b(t,x)\big| + \big|\sigma(t,x)\big| &\leq L(1+|x|) \qquad \forall x \qquad \textit{(linear growth)} \end{aligned}$$

Then $(*)$ admits a unique solution $\{X(t), t \geq 0\}$ satisfying $E \sup\limits_{0\leq t\leq T} |X(t)|^l \leq K_{T,L}$, $\forall\ 0 < T < \infty$

$$E\big|X(t) - X(s)\big|^l \leq K_{T,l}|t-s|^{l/2}$$

where $l \geq 1$.

Remark. (i) If the linear growth condition fail, we can have

$$\begin{cases} dX(t) = X(t)^2 dt \\ X(0) = 1 \end{cases}$$

The only possible solution is $X(t) = \frac{1}{1-t}$, which explodes at $t = 1$. Therefore there is no solution.

(ii) If Lipschitz condition fails, we can have

$$\begin{cases} dX(t) = 2X(t)^{2/3} dt \\ X(0) = 0 \end{cases}$$

$X(t) = (t-c)^3$ satisfies $dX(t) = 2X(t)^{2/3}dt\ \forall c > 0$.

Take $X(t) = \begin{cases} 0 & t < c \\ (t-c)^3 & t \geq c \end{cases}$ which is also solution.

Therefore there is no unique solution.

Example 7.3.1. Solve the Ornstein-Uhlenbeck equation

$$\begin{cases} dx(t) = \mu x(t)dt + \sigma dB(t) \\ X(0) = x_0 \end{cases}$$

Solution.

$$\begin{aligned}
dX(t) - \mu X(t)dt &= \sigma dB(t) \\
e^{-\mu t}\big(dX(t) - \mu X(t)dt\big) &= \sigma e^{-\mu t} dB(t) \\
d\big[e^{-\mu t}X(t)\big] &= \sigma e^{-\mu t} dB(t) \\
e^{-\mu t}X(t) - X_0 &= \int_0^t \sigma e^{-\mu s} dB(s) \\
X(t) &= X_0 e^{-\mu t} + \sigma e^{\mu t} \int_0^t e^{-\mu s} dB(s)
\end{aligned}$$

□

A Challenge Left as an Exercise

$$\begin{cases} d\Phi(t) = A(t)\Phi(t)dt + c(t)\Phi(t)dB(t) \\ \Phi(0) = I \end{cases}$$

Assume $|A(t)| + |c(t)| \leq K$.

1. Is $\Phi^{-1}(t)$ exists?
2. $\Phi^{-1}(t)$ satisfies what equation?