Probability Theory II
Basics of Stochastic Processes; Markov Processes and Martingales

Seiji HIRABA
May 5, 2020

Contents

1 Definitions of Stochastic Processes \((X_n, P), (X_t, P)\) 1

2 Discrete-time Markov Chains 1
   2.1 Basic examples ................................. 1
   2.2 Time-homogeneous Markov chain ............... 1
   2.3 \(d\)-dimensional random walks .................. 6
   2.4 Galton-Watson process (GW process) .......... 8

3 Martingales 11
   3.1 Uniform integrability .......................... 11
   3.2 Radon-Nikodym theorem and conditional expectations ........ 12
   3.3 Definition and properties of martingales, Doob’s decomposition .... 14
   3.4 Stopping times and optional sampling theorem ........ 15
   3.5 Sub-martingale inequalities and convergence theorems .......... 16

4 Continuous-time Markov Chains 20
   4.1 Exponential times ............................. 20
   4.2 Poisson processes ............................ 21
   4.3 Continuous-time random walks .................. 25
   4.4 Continuous-time Galton-Watson processes ......... 25
   4.5 Continuous-time Markov chains & transition probabilities .... 26

In this text, we describe Markov properties and martingale properties on discrete/continuous-time stochastic processes.

We give examples of Markov processes, random walks, Galton-Watson processes and Poisson processes, and we investigate their properties.
1 Definitions of Stochastic Processes \((X_n, P), (X_t, P)\)

On a probability space \((\Omega, \mathcal{F}, P)\), a stochastic process is a family of random variables (RVs) \((X_n = X_n(\omega))\) or \((X_t = X_t(\omega))\) with time index \(n \in \mathbb{N}\) or \(Z_+\), i.e., \(n = 1, 2, \ldots\) or \(n = 0, 1, 2, \ldots\), or \(t \in [0, \infty)\). (It is called discrete-time or continuous-time).

A probability space \((\Omega, \mathcal{F}, P)\) is that \(\Omega \neq \emptyset\) is a non-empty set (a total set or a total event), \(\mathcal{F} \subset 2^\Omega\) is a \(\sigma\)-additive class, an element \(A \in \mathcal{F}\) is called an event, \(P = P(d\omega)\) is a probability measure. (where \(2^\Omega\) is a family of all subsets of \(\Omega\).)

\[(X_n) = \{X_n\} = \{X_n\}_{n \geq 0}\) is called a discrete-time stochastic process.
\[(X_t) = \{X_t\} = \{X_t\}_{t \geq 0}\) is called a continuous-time process.

When a filtration \((\mathcal{F}_n)\), i.e., a family of increasing sub \(\sigma\)-add. classes of \(\mathcal{F}\); \(\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}\) is given, for each time points \(n\), if \(X_n\) is \(\mathcal{F}_n\)-measurable, then \((X_n)\) is called an \((\mathcal{F}_n)\)-adapted stochastic process. In the following, we always assume this condition is satisfied. If the time index \(n\) is changed to \(t\), then it is a similar.

In this text, we first discuss on discrete-time processes, and at the end on continuous-time processes.

2 Discrete-time Markov Chains

In discrete-time stochastic processes we first investigate “Markov chains”.

2.1 Basic examples

A Markov chain is a stochastic process such that the future action depends only on the present state and it is independent of the past actions.

We give two examples.

The first one is called a random walk (RW) which has independent identically distributed increments, i.e., \(X_0, X_1 - X_0, X_2 - X_1, \ldots\) are independent and \(X_{n+1} - X_n \overset{\text{iid}}{=} X_1 - X_0\) \((n \geq 1)\). It is the simplest model and well investigated.

**Example 2.1** Let \((X_n, P)\) be a random walk on \(\mathbb{Z}\) starting from 0; \(X_0 = 0\) such that
\[
P(X_{n+1} = x + 1 \mid X_n = x) = p, \quad P(X_{n+1} = x - 1 \mid X_n = x) = q.
\]

The second one is a Gorton-Watson (GW) process which is a population model of generational change with respect to a family tree. Bienaymé, Galton, Watson noticed that many family trees go frequently lost and they calculated a survival probability of a family tree.

**Example 2.2** Galton-Watson process or Bienaymé-Galton-Watson process \((Z_n, P)\) is a number of males in each generations such that each males has \(Y\)-number of males, where \(Y\) satisfies \(P(Y = k) = p_k\) for \(k = 0, 1, 2, \ldots\) \((p_k)\) is a distribution; \(p_k \geq 0, \sum p_k = 1\). Let \(Z_n\) be a number of males of the \(n\)-th generation. Let the starting point be one ancestor; \(Z_0 = 1\). Then each born males remains boys independently according to the same probability of \(Y\).

In this model, we can show that it depends on a mean of number of descendants \(m = \sum kp_k\) that the survival probability of a family tree is positive or not.

2.2 Time-homogeneous Markov chain

In this subsection, we show the following result:

**Theorem 2.1** Let \(S\) be a countable set. An \(S\)-valued irreducible time-homogeneous Markov chain is recurrent or transient.

Now in general, so many people say “Mathematics is difficult, because sentences are unintelligible.”

The above sentence may be just so. The cause is simple that many people does not understand definitions of mathematical terms.
“irreducible”, “time-homogeneous”, “Markov chain”, “recurrent”, “transient”

A Markov chain a process such that future actions depend only on a present state and it is independent of past actions, however, roughly speaking, it may be called a hit-or-miss process or a stopgap process.

The exact definition is the following:

Let $S$ be a countable set. An $S$-valued stochastic process $(X_n, P) = (X_n(\omega), P(d\omega))$ $(n = 0, 1, 2, \ldots)$ is called a Markov Chain if it satisfies the following:

(M1) [Markov property] For $n \geq 1, j_0, j_1, \ldots, j_n, k \in S$,
$$P(X_{n+1} = k \mid X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n) = P(X_{n+1} = k \mid X_n = j_n).$$

Moreover, it is called a time-homogeneous Markov chain if it also satisfies that

(M2) [Time homogeneity] For $n \geq 1, j, k \in S$,
$$P(X_{n+1} = k \mid X_n = j) = P(X_1 = k \mid X_0 = j) =: q(j, k).$$

In this text, we don’t treat time-inhomogeneous type, so in the following we always say a Markov process as a time-homogeneous Markov chain.

The distribution of $X_0$: $\mu = \{\mu_j\} \mu_j = P(X_0 = j)$ is called an initial distribution, and especially, if for some $j \in S$, $P(X_0 = j) = 1$, then we denote $P$ as $P_j$ and $(X_n, P_j)$ is called a Markov chain starting from $j$. (This is equivalent to that when $P(X_0 = j) > 0$, $P_j$ is defined as $P_j(\cdot) := P(\cdot \mid X_0 = j)$. It is convenient in calculations.)

For $n \geq 0, j, k \in S$, let $q_n(j, k) = P(X_n = k \mid X_0 = j)$ and $Q_n = (q_n(j, k))$ is called an $n$-step transition probability (matrix), in particular, denote $Q_1$ as $Q = (q(j, k))$ and it is simply called a transition probability (matrix).

**Question 2.1** Show the following:
(i) $q_n(j, k) \geq 0, \sum_k q_n(j, k) = 1 \ (j \in S)$.
(ii) For $n \geq 1, j_0, j_1, \ldots, j_n \in S$,
$$P(X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n) = \mu_{j_0} q(j_0, j_1) \cdots q(j_{n-1}, j_n).$$
(iii) For $m, n \geq 1, j_1, \ldots, j_m, k_0, k_1, \ldots, k_n \in S$,
$$P(X_{n+m} = j_m \mid X_0 = k_0, X_1 = k_1, \ldots, X_n = k_n)$$
$$= q(k_n, j_1) q(j_1, j_2) \cdots q(j_{m-1}, j_m).$$
(iv) $Q_0 = I := (\delta_{jk})$ (unit matrix), $Q_n = Q^n \ (n \geq 1)$, where $\delta_{jk} = 1 \ (j = k)$, $= 0 \ (j \neq k)$.

**Question 2.2** Show that if $\mu = \{\mu_j\}$ is an initial distribution of a Markov chain $(X_n)$, then
$$P(X_n = k) = \sum_{j \in S} \mu_j q_n(j, k).$$

We defined a recurrence time $T_j$ to $j \in S$ as
$$T_j = \inf \{n \geq 1; X_n = j\} = \infty \ \text{if} \ \{\cdot\} = \emptyset.$$

We also define
- $j$ is recurrent $\iff P_j(T_j < \infty) = 1$,
- $j$ is transient=non-recurrent $\iff P_j(T_j < \infty) < 1$.
If all \( j \) are recurrent (or transient), then \((X_n)\) is called recurrent (or transient).

A Markov chain \( \{X_n\} \) or a transition probability \( Q = (q(j,k)) \) is \textit{irreducible} if for arbitrary \( j, k, \quad \exists n \geq 1; \quad q_n(j,k) > 0 \). This means it is possible to go to anywhere if it starts from anywhere. (In other word, there is no point that is a trap or transient or it can not go.)

The following is a main result for a time-homogeneous Markov chain in this section.

\textbf{Theorem 2.2} \quad \text{Let} \ j, k \in S. \\
(i) \quad \text{The condition that} \ j \ \text{is recurrent is equivalent to the following:} \\
\quad \begin{align*}
& a) \quad \sum_{n=0}^{\infty} q_n(j,j) = \infty. \\
& b) \quad P_j(\{X_n\} \text{ is returns to} \ j \ \text{infinitely many times }) = 1.
\end{align*}
(ii) \quad \text{The condition that} \ j \ \text{is transient is equivalent to the following:} \\
\quad \begin{align*}
& a) \quad \sum_{n=0}^{\infty} q_n(j,j) < \infty. \\
& b) \quad P_j(\{X_n\} \text{ is returns to} \ j \ \text{infinitely many times }) = 0.
\end{align*}
(iii) \quad \text{If} \ \{X_n\} \ \text{is an irreducible Markov chain, then it is recurrent or transient.}

We first show b) of (i), (ii) and a), and (iii).

\textbf{Question O-1} \quad \text{Show that for} \ m, n \geq 1, j_1, \ldots, j_m, k_0, k_1, \ldots, k_n \in S, \\
\quad \begin{align*}
& P(X_{n+1} = j_1, \ldots, X_{n+m} = j_m \mid X_0 = k_0, X_1 = k_1, \ldots, X_n = k_n) \\
& = P(X_{n+1} = j_1, \ldots, X_{n+m} = j_m \mid X_n = k_n).
\end{align*}

\textbf{Question O-2} \quad \text{Let} \ \{B_k\}_{k=1}^n \ \text{be disjoint events and for a event} \ A, C \ \text{it satisfies that} \ P(A \mid B_k) = P(A \mid C) \ (1 \leq k \leq n). \ \text{Then show} \ P(A \mid \bigcup B_k) = P(A \mid C).

\textbf{Proposition 2.1} \quad (i) \quad \text{If} \ j \in S \ \text{is recurrent, then} \ P_j(\{X_n\} \ \text{is returns to} \ j \ \text{infinitely many times }) = 1.
\quad (ii) \quad \text{If} \ j \in S \ \text{is transient, then} \ P_j(\{X_n\} \ \text{is returns to} \ j \ \text{infinitely many times }) = 0.

\textbf{Proof.} \quad \text{Let} \ T^{(m)}_j \ \text{be the} \ m\text{-th return time to} \ j; \\
\quad \begin{align*}
& T^{(1)}_j = T_j, \quad T^{(m)}_j = \min\{n > T^{(m-1)}_j; X_n = j\} = \infty \ \text{if} \ \{\cdot\} = \emptyset.
\end{align*}

We first show \( P_j(T^{(m)}_j < \infty) = P_j(T_j < \infty)^m \). \text{For positive integers} \ s, t, \ \text{by time-homogeneous Markov property we can show} \\
\quad \begin{align*}
& P_j(T^{(m)}_j = s + t \mid T^{(m-1)}_j = s) = P_j(T_j = t).
\end{align*}

(In fact, \([\text{RHS}] = P(X_{s+t} = j, X_{s+u} \neq j \ (1 \leq u \leq t-1) \mid T^{(m-1)}_j = s) \) and by using \( \{X_u \neq j\} = \bigcup_{k_u \in S,k_u \neq j} \{X_u = k_u\} \) and noting that \( T^{(m-1)}_j = s \) is determined by the state of \( \{X_1, \ldots, X_{s}(= j)\} \), and by using the above questions O-1, O-2, we can get it.) \text{Moreover, by} \ P(A \cap B) = P(B \mid A)P(A) \text{we have} \\
\quad \begin{align*}
& P_j(T^{(m-1)}_j = s, T^{(m)}_j = s + t) = P_j(T^{(m-1)}_j = s)P_j(T_j = t).
\end{align*}

Hence, by \\
\quad \begin{align*}
& P_j(T^{(m)}_j < \infty) = P_j(T^{(m-1)}_j < T^{(m)}_j < \infty) \\
& = \sum_{s=m-1}^{\infty} \sum_{t=1}^{\infty} P_j(T^{(m-1)}_j = s, T^{(m)}_j = s + t) \\
& = P_j(T^{(m-1)}_j < \infty)P_j(T_j < \infty)\)
we have $P_j(T_j^{(m)} < \infty) = P_j(T_j < \infty)^m$. Therefore,

$$P_j(\{X_n\} \text{ returns to } j \text{ infinitely many times}) = P_j(\bigcap_{m} \{T_j^{(m)} < \infty\})$$

$$= \lim_{m \to \infty} P_j(T_j^{(m)} < \infty) = \lim_{m \to \infty} P_j(T_j < \infty)^m.$$

This is 1 if $P_j(T_j < \infty) = 1$, or 0 if otherwise.

We define some notations: For $j, k \in S$, let $f_m(j, k) := P_j(T_k = m) \ (m \geq 1)$ and

$$Q_{jk}(s) := \sum_{n=0}^{\infty} q_n(j, k)s^n \ (|s| < 1), \quad F_{jk}(s) := \sum_{m=1}^{\infty} f_m(j, k)s^m \ (|s| \leq 1).$$

Each is called a generating function of $\{q_n(j, k)\}_{n \geq 0}$ or $\{f_m(j, k)\}_{m \geq 1}$, respectively. Note that $\lim_{s \uparrow 1} Q_{jk}(s) = \sum_{n=0}^{\infty} q_n(j, k)$ and $F_{jk}(1) = P_j(T_k < \infty)$.

**Lemma 2.1** For $j, k \in S$, the following hold:

$$q_n(j, k) = \sum_{m=1}^{n} f_m(j, k)q_{n-m}(k, k) \ (n \geq 1), \quad Q_{jk}(s) = \delta_{jk} + F_{jk}(s)Q_{kk}(s) \quad (|s| < 1).$$

**Proof.** Noting that $\{T_k = m\} = \{X_m = k, X_s \neq k \ (1 \leq s \leq m - 1)\}$, we have

$$\sum_{m=1}^{n} f_m(j, k)q_{n-m}(k, k) = \sum_{m=1}^{n} P_j(T_k = m)P_j(X_n = k | X_m = k)$$

$$= \sum_{m=1}^{n} P_j(T_k = m)P_j(X_n = k | T_k = m)$$

$$= \sum_{m=1}^{n} P_j(X_n = k | T_k = m)$$

$$= P_j(X_n = k) = q_n(j, k).$$

Moreover, by this (also by change of sums $\sum_{n=1}^{\infty} \sum_{m=1}^{n} = \sum_{m=1}^{\infty} \sum_{n=m}^{\infty}$) we have

$$Q_{jk}(s) = \delta_{jk} + \sum_{n=1}^{\infty} q_n(j, k)s^n$$

$$= \delta_{jk} + \sum_{n=1}^{\infty} \sum_{m=1}^{n} f_m(j, k)q_{n-m}(k, k)s^n$$

$$= \delta_{jk} + F_{jk}(s)Q_{kk}(s).$$

**Proposition 2.2** $j \in S$ is recurrent $\iff \sum_{n=0}^{\infty} q_n(j, j) = \infty$.

**Proof.** By the above lemma, $Q_{jj}(s)(1 - F_{jj}(s)) = 1 \ (|s| < 1)$. Hence, by $F_{jj}(1) = P_j(T_j < \infty)$ and

$$\lim_{s \uparrow 1} Q_{jj}(s) = \sum_{n=0}^{\infty} q_n(j, j) \leq \infty,$$
and by letting $s \uparrow 1$ it is obtained. (In formal, it can be expressed as
\[
\sum_{n=0}^{\infty} q_n(j,j)(1 - P_j(T_j < \infty)) = 1.
\]
Thus, if $P_j(T_j < \infty) = 1$, then $\sum_{n=0}^{\infty} q_n(j,j) = 1$. If $P_j(T_j < \infty) < 1$, then $\sum_{n=0}^{\infty} q_n(j,j) < \infty.$) \hfill ■

**Question 2.3** By a similar way to the above proof and considering the case of $j \neq k$, show

\[
j \in S \text{ is transient} \Rightarrow \sum_{n=0}^{\infty} q_n(k,j) < \infty \text{ (} k \in S \text{)}
\]

(Then, the contraposition also holds: $[\exists k \in S: \sum_{n=0}^{\infty} q_n(k,j) = \infty \Rightarrow \text{ } j \text{ } \text{ is recurrent.})$

*Lemma 2.2* If $j$ is recurrent and $j \to k$, i.e., $\exists n; q_n(j,k) > 0$, then $P_k(T_j < \infty) = 1.$

*Proof.* For arbitrary $i, j \in S$, it holds that

\[
P_i(T_j < \infty) = q(i,j) + \sum_{k \in S, k \neq j} q(i,k)P_k(T_j < \infty).
\]

(In fact, by time-homogeneous Markov property [and also by $P_i(A \mid B) = P(A \mid B \cap \{X_0 = i\})$ (→ check this]), we have $P_i(T_j = n|X_1 = k) = P(T_j = n|X_0 = i, X_1 = k) = P(T_j = n|X_1 = k) = P_k(T_j = n - 1).$

Hence, $P_i(X_1 = k, T_j = n) = q(i,k)P_k(T_j = n - 1)$ and

\[
P_i(T_j < \infty) = \sum_{n=1}^{\infty} \sum_{k \in S} P_i(X_1 = k, T_j = n) = P_i(X_1 = j) + \sum_{n=2}^{\infty} \sum_{k \neq j} P_i(X_1 = k, T_j = n).
\]

These imply the above equation.) Now note that by the assumption of $q_n(j,k) > 0$, we have $\exists (k_1, \ldots, k_{n-1}); q(j,k_1)q(k_1,k_2)q(k_2,k_3)\cdots q(k_{n-1},k) < 0$. Hence, if we let $i = j$ in the above equation, then by the recurrence of $j$, i.e., $P_j(T_j < \infty) = 1$, we can get for $\forall k; q(j,k) > 0, P_k(T_j < \infty) = 1.$

Furthermore, let $k = k_1$, and again in the above equation let $i = k_1$. If $k = k_2$, then by $q(k_1,k_2) > 0$, we have $P_{k_2}(T_j < \infty) = 1.$ By continuing this operation, the desired result is obtained. \hfill ■

**Question 2.4** Show the following by the previous Question 2.3 and by the above lemma:

\[
j \text{ is recurrent and } j \to k \Rightarrow \sum_{n=0}^{\infty} q_n(k,j) = \infty.
\]

For $j, k \in S$, we denote $j \leftrightarrow k$ if $j \to k$ and $k \to j$.

**Proposition 2.3** For $j, k \in S; j \leftrightarrow k$, If $j$ is recurrent or transient, then $k$ is so, respectively. Therefore, an irreducible Markov chain is recurrent or transient.

*Proof.* By $j \leftrightarrow k$, $\exists \ell, m \geq 0; q_\ell(j,k) > 0, q_m(k,j) > 0$. Moreover, by

\[
q_{\ell+m+n}(j,j) \geq q_\ell(j,k)q_n(k,k)q_m(k,j) \quad (n \geq 0),
\]

we have

\[
Q_{jj}(s) = \sum_{n=0}^{\infty} q_n(j,j)s^n \geq \sum_{n=0}^{\infty} q_{\ell+m+n}(j,j)s^{\ell+m+n} \geq s^{\ell+m}q_\ell(j,k)q_m(k,j)Q_{kk}(s).
\]
Hence, if \( j \) is transient, then
\[
\lim_{s \to 1} Q_{jj}(s) = \sum_{n=0}^{\infty} q_n(j, j) < \infty
\]
and by the above inequality, we have \( \sum_{n=0}^{\infty} q_n(k, k) < \infty \), and hence, \( k \) is also transient. It is the same if we exchange \( j, k \).

**Remark 2.1** By the previous Question 2.3, 2.4 for an irreducible Markov chain,

- if it recurrent, then for \( \forall j, k \in S \), \( \sum_n q_n(j, k) = \infty \).
- if it transient, then for \( \forall j, k \in S \), \( \sum_n q_n(j, k) < \infty \).

Conversely, for some \( j, k \in S \), if \( \sum_n q_n(j, k) = \infty \), then it is recurrent, or if \( \sum_n q_n(j, k) < \infty \), then it is transient.

### 2.3 \( d \)-dimensional random walks

Let \( S = \mathbb{Z}^d \) (\( \geq j = (j_1, \ldots, j_d) \)) be a \( d \)-dimensional lattice.

\( \{p_k\}_{k \in \mathbb{Z}^d} \) is a distribution on \( \mathbb{Z}^d \) if \( \sum p_k = 1 \).

**Definition 2.1** \((X_n, P)\) is a \( d \)-dimensional random walk (\( d \)-dim. RW) if \( \{X_0, X_1 - X_0, X_2 - X_1, \ldots\} \) are independent and \( P(X_n - X_{n-1} = k) = p_k \) (\( n \geq 1, k \in \mathbb{Z}^d \)), where \( \{p_k\}_{k \in \mathbb{Z}^d} \) is a distribution on \( \mathbb{Z}^d \). (It is also called a RW with a one-step dist., \( \{p_k\} \)).

Especially, if \( p_k = 1/(2d) \) for all \( k \in \mathbb{Z}^d \), \( |k| = 1 \), then it is called a simple RW with a one-step dist., \( \{p_k\} \). Moreover, if \( p_k = 1/(2d) \) for all \( k \in \mathbb{Z}^d \), \( |k| = 1 \), then it is called a simple random walk, where \( k = (k_1, \ldots, k_d) \), \( |k| = \sqrt{k_1^2 + \cdots + k_d^2} \).

Moreover, define \( P_j \) by \( P_j(X_1 = k_1, \ldots, X_n = k_n) := P(X_1 = k_1, \ldots, X_n = k_n | X_0 = j) \), then \((X_n, P_j)\) is called a \( d \)-dim. RW starting from \( j \).

**Remark 2.2** If we denote a conditional probability \( P(A \mid B) := P(A \cap B)/P(B) \), then we always assume \( P(B) > 0 \).

**Question** Show that \( A, B \in \mathcal{F} \) are indep. \( \iff P(A \mid B) = P(A) \).

Clearly, a \( d \)-dim. RW is a Markov chain. Its transition probability \( Q = (q(j, k)) \) is given as \( q(j, k) = P_{k-j} \). Moreover, a simple RW is irreducible.

**Question 2.5** Check the above results: [time-homogeneous Markov property, transition probability, irreducibility]

**Question 2.5 Revision** Let \((X_n, P)\) be \( d \)-dim. RW.

1. Show \( X_{n+1} - X_n \) and \((X_0, X_1, \ldots, X_n)\) are indep., i.e.,
\[
P(X_{n+1} - X_n = j, X_0 = k_0, X_1 = k_1, \ldots, X_n = k_n) = P(X_{n+1} = j | X_n = k_n) P(X_0 = k_0, X_1 = k_1, \ldots, X_n = k_n).
\]

Especially, by summing on \( k_0, k_1, \ldots, k_{n-1} \in \mathbb{Z}^d \), it can be seen that \( X_{n+1} - X_n \) and \( X_n \) are indep.

2. Show \( P(X_{n+1} = j | X_0 = k_0, X_1 = k_1, \ldots, X_n = k_n) = P(X_{n+1} = j | X_n = k_n) = p_{j-k_n} \).

This implies \( \{X_n\} \) is a time-homogeneous Markov chain such that \( q(j, k) = p_{k-j} \).

3. Show that a simple RW is irreducible. (By using \( \|j - k\| := |j_1 - k_1| + \cdots + |j_d - k_d| \), divide the cases \( j \neq k, j = k \).)

Now we can discuss on recurrence and transience by using transition probabilities \( Q = (q(j, k)) = (p_{k-j}) \).

By using the results in the previous section, it is relatively easy to see that the following for a simple RW:
Theorem 2.3 A \(d\)-dimensional simple RW is

1. recurrent (i.e., \(P_j(T_j < \infty)\)) if \(d = 1, 2\).
2. transient if \(d \geq 3\).

In this text we show the case of \(d \geq 3\).

By the irreducibility it is recurrent or transient. It is enough to check the convergence or divergence of the sum of \(q_n(0, 0)\).

Since it is not return to starting point by odd steps, we have \(q_{2n+1}(0, 0) = 0\), and hence, it is enough to consider on \(q_{2n}(0, 0)\). We can show the following: (By this result the recurrence or transience is obtained by Theorem 2.2 in the previous section.)

Proposition 2.4 Let \(Q = (q(j, k))\) be a transition probability of a \(d\)-dim. simple RW.

1. If \(d = 1, 2\), then as \(n \to \infty\),
   \[
   q_{2n}(0, 0) \sim \begin{cases} \frac{1}{\sqrt{\pi n}} & (d = 1) \\ \frac{1}{(\pi n)} & (d = 2) \end{cases}
   \]
   where \(a_n \sim b_n\ (n \to \infty) \overset{\text{def}}{\Rightarrow} a_n/b_n \to 1\ (n \to \infty)\).

2. If \(d = 3\), then \(\exists C > 0; q_{2n}(0, 0) \leq Cn^{-3/2}\).

Question 2.6 Show that for positive numerical sequences \(\{a_n\}, \{b_n\}\), if \(a_n \sim b_n\ (n \to \infty)\), then \(\exists c_1, c_2 > 0; c_1b_n \leq a_n \leq c_2b_n\ (n \geq 1)\) holds.

Remark 2.3 It is known that (the constant is \(\sqrt{3/\pi^3}/4\) if \(d = 3\))

\[
q_{2n}(0, 0) \sim 2^{1-d}d^{d/2}(\pi n)^{-d/2} \quad (n \to \infty).
\]

We give an important formula:

**[Stirling’s formula]** \(n! \sim \sqrt{2\pi n^{n+1/2}}e^{-n}\ (n \to \infty)\).
where \( c_n = \max_{j,k,m\geq 0, j+k+m=n} (j!k!m!)^{-1} \). Moreover, for \( c_n \), the following holds and by Stirling formula we get the result.

\[
(2.1) \quad c_n \leq c 3^{n+3/2} n^{-3/2} e^{-n} \quad (c > 0 \text{ is independent of } n \geq 1).
\]

In fact, by dividing \( n \) by 3 and dividing into the cases of remains, we see that

\[
(2.2) \quad c_n \leq \begin{cases} 
(m!)^{-3} & (n = 3m), \\
(m!)^{-2}((m+1)!)^{-1} & (n = 3m + 1), \\
(m!)^{-1}((m+1)!)^{-2} & (n = 3m + 2).
\end{cases}
\]

hence, by Stirling formula, there exist constants \( c_1, c_2 > 0 \) such that

\[
c_1 n^{n+1/2} e^{-n} \leq n! \leq c_2 n^{n+1/2} e^{-n}
\]

and if we substitute this to the above, then we can get the result.

**Question 2.7** Calculate the cases of 1-dimensional and 2-dimensional by using Stirling formula.

**Question 2.8** Show the above inequation (2.2) and by using it, show (2.1), and verify the proof (calculation) of \( d = 3 \).

### 2.4 Galton-Watson process (GW process)

A Galton-Watson process is a family tree model. Let \( Z_n \in \mathbb{Z}_+ = \{0, 1, 2, \ldots\} \) be a number of males of the \( n \)-th generation and let \( Z_0 = 1 \). In general, we call the object as a particle and each particles born \( Y \) number of particles as the next generation, where \( Y \) is a \( \mathbb{Z}_+ \)-valued RV with a dist. \((p_k)_{k \geq 0}\), i.e., \( P(Y = k) = p_k \) \((k \geq 0)\). Clearly, \( \{Z_n\} \) is a Markov process with a transition probability such that

\[
p(i, j) = P(Z_{n+1} = j \mid Z_n = i) = P(\sum_{k=1}^{i} Y_k = j) \quad (i \geq 1, j \geq 0),
\]

where \( \{Y_k\} \) is i.i.d. with a dist. \((p_k)\). If \( Z_n = 0 \), then a family tree is lost. Hence,

\[
p(0, i) = 0 \quad (i \geq 1), \quad p(0, 0) = 1
\]

We assume existence of a mean of \((p_k)\):

\[
m := \sum_{k=1}^{\infty} kp_k \in (0, \infty).
\]

Now let \( q \) be an extinction probability of the GW-process starting from 1.

Then, by \( Z_1 \overset{d}{=} Y \) and conditioned on \( \{Z_1 = k\} \), we have

\[
q = P(\text{extinction} \mid Z_0 = 1) = P(\exists n \geq 1; Z_n = 0 \mid Z_0 = 1) = \sum_{k \geq 0} P(\text{extinction} \mid Z_1 = k) P(Y = k) = \sum_{k \geq 0} q^k p_k.
\]

\( q = 1 \) is one solution of this equation, however, what is the condition of \( q \in [0, 1) \)?

In order to answer the question, we introduce the following generating function \( f \): \( q \) is a solution to a equation \( s = f(s) \enspace (s \in [0, 1]) \)

\[
f(s) = E[s^Y] = \sum_{k=0}^{\infty} p_k s^k \quad (|s| \leq 1).
\]

This series converges absolutely in \(|s| \leq 1\), and hence, it is infinitely differentiable. Moreover, it holds that

\[
f(0) = p_0, \quad f(1) = 1, \quad f'(1) = \sum_{k \geq 1} kp_k = m.
\]
\textbf{Theorem 2.4} A GW-process satisfies the following: Denote $P_1(\cdot) = P(\cdot | Z_0 = 1)$.

$m < 1 \text{ or } [m = 1, p_0 > 0] \implies P(\gamma n \geq 1, Z_n \geq 1 | Z_0 = 1) = 0$, \textit{i.e.}, $q = 1$

$m > 1 \implies P(\gamma n \geq 1, Z_n \geq 1 | Z_0 = 1) > 0$, \textit{i.e.}, $q < 1$

Moreover, when $m > 1$, the extinction probability $q$ is a unique solution to the equation $f(s) = s$ on $[0, 1)$.

Note that if $p_0 = 0$, then it surely leaves offspring, and hence $q = 0$ (in this case $m \geq 1$ holds). Especially, if $p_1 = 1$, then $m = 1$ and $q = 0$.

We first show some propositions. Since $f$ is increasing in $s \in [0, 1]$ from $f(0) = p_0 \geq 0$ to $f(1) = 1$, we can consider the composition itself; we define $f_1 = f$, $f_{n+1} = f \circ f_n$ ($n \geq 1$).

\textbf{Proposition 2.5} For each $n \geq 1$, the generating function of $Z_n$ is $f_n$ under the condition $Z_0 = 1$, \textit{i.e.}, $E_s[Z_n] = f_n(s)$.

\textbf{Proof.} We denote $P_1 = P, E_1 = E$. Let $g_n(s) = E_s[Z_n] = \sum_{k=0}^{\infty} s^k P(Z_n = k)$. If $n = 1$, then under $\{Z_0 = 1\}$, by $Z_1 \overset{(d)}{=} Y$, clearly $g_1(s) = E_s[Y] = f(s)$. We assume for $n \geq 1$, $g_n = f_n$. Under $\{Z_n = k\}$, $Z_{n+1} \overset{(d)}{=} \sum_{i=1}^{k} Y_i$ and $\{Y_i\}$ are i.i.d. and $Y \overset{(d)}{=} Y$. Hence

$$E_s[Z_{n+1}|Z_n = k] = E_s[\prod_{i=1}^{k} Y_i | Z_n = k] = \prod_{i=1}^{k} E_s[Y_i] = f(s)^k.$$

Therefore,

$$g_{n+1}(s) = \sum_{k=0}^{\infty} E_s[Z_{n+1}|Z_n = k] P(Z_n = k) = \sum_{k=0}^{\infty} f(s)^k P(Z_n = k) = g_{n}(f(s)).$$

By the assumption of the induction, $g_{n+1}(s) = g_n(f(s)) = f_n(f(s)) = f_{n+1}(s)$. \hfill \blacksquare

\textbf{Proposition 2.6} $E_s[Z_0] = m^n$ ($n \geq 0$).

\textbf{Proof.} We denote $P_1 = P, E_1 = E$. Note that $m = E[Y] = E[Z_1]$ and $E[Z_0|Z_{n-1} = k] = E[E_s[Z_n|Z_{n-1} = k] = km$, we have

$$E_s[Z_n] = \sum_{k \geq 1} E_s[Z_n | Z_{n-1} = k] P(Z_{n-1} = k) = \sum_{k \geq 1} km P(Z_{n-1} = k) = m E_s[Z_{n-1}].$$

By continuing this, we have $E_s[Z_n] = m^{n-1} E_s[Z_1] = m^n$. \hfill \blacksquare

\textbf{[Proof of Theorem 2.4]}

Since under $P_1$, the generating function of $Z_n$ is $f_n$, it holds $P_1(Z_n = 0) = f_n(0)$; note that $\{Z_n = 0\} \uparrow$, hence, $q = P_1(\exists n \geq 1; Z_n = 0) = P_1(\bigcup_{n \geq 1} \{Z_n = 0\}) = \lim_{n \to \infty} f_n(0)$.

Hence, by $f_{n+1}(0) = f(f_n(0))$, letting $n \to \infty$ and by the continuity of $f$, we have $q = f(q)$.

\textbf{(Case: }$m < 1$\textbf{)} By $P_1(Z_n \geq 1) \leq E_1[Z_n] = m^n$, noting that $\{Z_n \geq 1\} \downarrow$, we have

$$0 = \lim_{n \to \infty} P_1(Z_n \geq 1) = P_1(\bigcap_{n \geq 1} \{Z_n \geq 1\}) = P_1(\exists n \geq 1, Z_n \geq 1) \text{ i.e.}, q = 1.$$

\textbf{(Case: }$m = 1$\textbf{)} If $p_0 > 0$, then $p_0 + p_1 < 1$. (In fact, if we assume $p_0 + p_1 = 1$, then $m = p_1 = 1 - p_0 < 1$ and this contradicts.) Hence, noting that $\exists k \geq 2; p_k > 0$, we have

$$f'(s) = \sum_{k \geq 1} kp_k s^{k-1} < f'(1) = \sum_{k \geq 1} kp_k = m = 1 \quad (0 < s < 1).$$
By mean value theorem for \( s \in (0, 1) \), \( 3s \in (s, 1) \); \( f(1) - f(s) = f'(c)(1 - s) < 1 - s \). By \( f(1) = 1 \), we get \( f(s) > s \) (0 < \( s < 1 \)). Moreover, by \( f(0) = p_0 > 0 \), the solution for \( f(s) = s \) in \([0, 1]\) is \( s = 1 \) only. Thus, \( q = 1 \).

(Case: \( m > 1 \)) Note that \( p_0 + p_1 < 1 \) (because if \( p_0 + p_1 = 1 \), then \( m = p_1 \leq 1 \)). By \( f'(1) = m > 1 \) and by the continuity of \( f' \),

\[
3\eta > 0; 1 - \eta < s < 1, 1 < f'(s) < f'(1) = m
\]

Hence, if \( 1 - s < \eta < 1 \), then \( f(s) > s \). Since \( f(0) = p_0 \geq 0 \), and by using intermediate value theorem for \( g(s) = f(s) - s \) we have \( 3s_1 \in [0, 1] \); \( f(s_1) = s_1 \). We show the uniqueness of the solution for this. If \( 3s_2 \in [0, 1] \); \( s_1 < s_2 \), \( f(s_2) = s_2 \), then \( g(s_1) = 0 \) and \( g(1) = 0 \) by \( f(1) = 1 \). By Roll’s theorem \( 0 \leq s_1 < 3s_1 < s_2 < 3s_2 < 1 \); \( g'(s_1) = g'(s_2) = 0 \), i.e., \( f'(s_1) = f'(s_2) = 1 \). Moreover, by \( p_0 + p_1 < 1 \),

\[
s \in (0, 1) \implies f''(s) = \sum_{k \geq 2} k(k-1)p_k s^{k-2} > 0.
\]

Hence, \( f''(s) \) is strictly increasing for \( s \in (0, 1) \). This contradicts \( f'(s_1) = f'(s_2) = 1 \). Thus, the solution of \( f(s) = s \) is only \( q = s_1 \) or \( q = 1 \). Furthermore, if \( q = 1 \), then \( 1 = q = \lim_{n \to \infty} f_n(0) \), and for sufficiently large \( n \gg 1 \), \( f_n(0) > 1 - \eta \). By the result shown as above, \( f_n(0) = f(f_n(0)) < f_n(0) \). This contradicts \( f_n \) is increasing (in \( n \)). Therefore, \( q = s_1 \in [0, 1] \).

**Example 2.3** If \( p_0 = p_2 = 1/2 \), then the mean is \( m = 1 \), however, this family tree becomes extinct someday.

**Example 2.4** Lotka found in 1939 the distribution of male descendants of Americans is a geometric distribution.

\[
P(Y = 0) = \frac{1}{2}, \quad P(Y = k) = \frac{1}{5} \left( \frac{3}{5} \right)^{k-1} \quad (k \geq 1).
\]

In this case,

\[
m = \frac{1}{5} \sum_{k \geq 1} k \left( \frac{3}{5} \right)^{k-1} = \frac{5}{4} > 1.
\]

Hence, the extinction probability \( q \) is a solution \( s = q < 1 \) for

\[
s = f(s) = \frac{1}{2} + \frac{1}{5} \sum_{k \geq 1} \left( \frac{3}{5} \right)^{k-1} s^k, \quad \text{that is,} \quad \frac{3}{5}s^2 = -\frac{11}{10}s + \frac{1}{2} = 0
\]

Then \( s = 5/6, 1 \) and \( q = 5/6 \). Therefore, the survival probability of a family tree is 1/6.

**Question 2.9** Calculate \( m = 5/4 \) and a solution \( s = 5/6, 1 \) for \( s = f(s) \).
3 Martingales

Let \( \{M_n\}_{n \geq 1} \) be a stochastic process with a filtration \( (F_n) \) and \( \{N_n\} \) be \( (F_n) \)-adapted. (We often use a notation \( \{M_n\} \) as a martingale.)

- A \{M_n\} is a martingale, more exactly, an \( (F_n) \)-martingale

\[
\begin{align*}
\text{def} & \quad M_n \in L^1, E[M_{n+1} | F_n] = M_n \text{ a.s., } \forall n \geq 1. \\
\iff & \quad M_n \in L^1, \forall n \geq 0, \forall A \in F_n, E[M_{n+1}; A] = E[M_n; A]
\end{align*}
\]

In general, for a RV \( X \) and sub-\( \sigma \)-add. class \( \mathcal{G} \subset \mathcal{F} \), \( E[X | \mathcal{G}] \) is called a conditional expectation of \( X \) under \( \mathcal{G} \), which is defined by Radon-Nikodym Theorem (R-N Th.).

3.1 Uniform integrability

Let \( (\Omega, \mathcal{F}, P) \) be a probability space. (However, we may set \( P \) be a finite measure in this subsection.)

- A sequence of measurable functions \( \{X_n\} \) is uniformly integrable (simply, we say UI.)

\[
\begin{align*}
\text{def} & \quad \lim_{n \to \infty} \sup_{|X_n| \geq a} E[X_n] = 0 \\
\iff & \quad (U1) \sup_{n} E[|X_n|] < \infty \iff \text{boundedness of means} \\
& \text{(U2)} \quad P(A) \to 0 \iff \sup_{n} E[|X_n|; A] \to 0 \iff \text{uniform absolute continuity of integrals},
\end{align*}
\]

i.e.,

\[
\forall \varepsilon > 0, \exists \delta > 0; \forall A \in \mathcal{F}; P(A) < \delta, E[|X_n|; A] < \varepsilon.
\]

**Proof.** (\( \Rightarrow \)) \((U1)\) follows from the following. (The finiteness of \( P(\Omega) \) is needed.)

\[
E[|X_n|] = E[|X_n|; |X_n| \geq a] + E[|X_n|; |X_n| < a] \leq \sup_n E[|X_n|; |X_n| \geq a] + aP(\Omega).
\]

(U2) is immediately obtained by the following:

\[
E[|X_n|; A] = E[|X_n|; A \cap \{X_n \geq a\}] + E[|X_n|; A \cap \{X_n < a\}] \leq E[|X_n|; |X_n| \geq a] + aP(A).
\]

It is enough to fix a sufficiently large \( a > 0 \) such that the last 1st term is lower than \( \varepsilon/2 \), and take \( \delta = \varepsilon/(2a) \).

\((\Leftarrow)\) By \( P(|X_n| \geq a) \leq E[|X_n|]/a \) and \((U1)\), letting \( a \to \infty \), this probability converges to 0 uniform. Hence, by \((U2)\), the desired result holds.

The following proposition is immediately follows:

**Proposition 3.1**

1. If \( \exists \forall Y \in L^1; |X_n| \leq Y \text{ a.s.}, \text{ then } \{X_n\} \text{ is UI}. \\
2. If \( \exists p > 1; \sup_n E[|X_n|^p] < \infty \), then \{X_n\} is UI.
3. If \( \{X_n\}; UI \text{ and } X_n \to X, \text{ a.s.}, \text{ then } X \in L^1. \\
4. If \( \{X_n\}; UL, Y \in L^1, \text{ then } X_n + Y \text{ is UI.} \\
5. If each \{X_n\}, \{Y_n\} is UI, then \{Z_n = X_n + Y_n\} is so.

**Proof.**

1. It is enough to show the case without a.s. and then, by \( \{X_n \geq a \} \subset \{Y \geq a\} \),

\[
E[|X_n|; |X_n| \geq a] \leq E[|Y|; Y \geq a], Y \text{ is integrable and absolute continuity of integrals implies the result.}
\]

2. Let \( K := \sup_n E[|X_n|^p] < \infty \). By Chebychev, \( P(|X_n| \geq a) \leq K/a^p \), and by Hölder, noting that \( 1/q + 1/p = 1 \), we have \( E[|X_n|; |X_n| \geq a] \leq E[|X_n|^p]^{1/p} E[1_{|X_n| \geq a}]^{1/q} \leq K^{1/p} P(|X_n| \geq a)^{1/q} \leq K/a^p \to 0 \text{ (a.e. \( \to \infty \).)}
\]

3. It is obvious by Fatou’s Lem. and boundedness of means.

4. (4), (5) are clear by checking \((U1),(U2)\).

The important result is the following convergence theorem.

**Theorem 3.1**

If \( X_n \to X, \text{ a.s. and } \{X_n\} \text{ is UI}, \text{ then } X_n \to X \text{ in } L^1, \text{ i.e. }, E[X_n - X] \to 0.

This is obtained as a corollary of the following result.

**Theorem 3.2**

Let \( X_n \in L^1, X_n \to X, \text{ a.s.} \) The following are equivalent:

1. \( \{X_n\}; UI \), \( \{X_n \to X \text{ in } L^1, \text{ i.e. }, E[X_n - X] \to 0, \ (3) E[|X_n|] \to E[|X|]. \)
This result is an extension of Lebesgue’s convergence theorem, because, if a sequence of functions is estimated by an integrable function, then it is UI.

**Proof.**

(1) ⇒ (2) Since \( \{X_n\} \) is UI, by the assumption and by the above prop., \( X \in L^1 \). Hence, again by the above prop. \( \{X_n-X\} \) is also UI. Moreover, \( X_n \to X \), a.s. implies in pr. For \( \forall \varepsilon > 0, P(\|X_n - X\| > \varepsilon) \to 0. \)

\[
E[|X_n - X|] \leq E[|X_n - X|; |X_n - X| \geq \varepsilon] + \varepsilon P(|X_n - X| < \varepsilon) \leq E[|X_n - X|; |X_n - X| \geq \varepsilon] + \varepsilon. 
\]

Therefore, by letting \( n \to \infty \) and by uniform abs. continuity of integrals, the 1st term of the above goes to 0 and since \( \varepsilon > 0 \) is arbitrary, we have \( \lim E[|X_n - X|] = 0. \)

(2) ⇒ (3) is clear (by \( |E[X_n] - E[X]| \leq E[|X_n - X|] \)).

(3) ⇒ (1) For a continuous point \( a > 0 \) of the distribution of \( |X| \), i.e., \( P(|X| = a) = 0 \), we can show \( E[|X_n|; |X_n| \geq a] \to E[|X|; |X| \geq a] (n \to \infty). \) If it is true, then by the abs. continuity of integral, sup \( n E[|X_n|; |X_n| \geq b] \to 0 (b \to \infty) \) can be shown. We show these in order.

For \( \forall \varepsilon > 0 \), set \( X^a = X_{\{X|<a\}} \). Then \( |X^a| \leq a \) and if \( |X(\omega)| \neq a \), then \( X^a(\omega) \to X^a(\omega) (n \to \infty) \) (→ show it). Therefore, For \( a > 0 \), \( P(|X| = a) = 0 \), it holds \( X^a_n \to X^a \), a.s. Since \( |X^a_n| \leq a \) and by bounded convergence theorem, we have \( E[X^a_n] \to E[X^a] \). Thus, by the assumption of (3),

\[
E[|X_n|; |X_n| \geq a] = E[|X_n| - E[X_n^a] \to E[|X| - E[X^a]] = E[|X|; |X| \geq a].
\]

Moreover, by the abs. continuity of integrals, For \( \forall \varepsilon > 0, \exists a > 0 \) (sufficiently large); The last term of the above = \( E[|X|; |X| \geq a] < \varepsilon/2. \) Furthermore, since this \( a \) can be taken as a continuous point of the dist. of \( |X| \), we have \( \exists N \forall n \geq N, E[|X_n|; |X_n| \geq a] < \varepsilon. \) On the other hand, for \( n < N \), by the abs. continuity of integrals, \( \exists b_k, k = 1, 2, \ldots, N; E[|X_k|; |X_k| \geq b_k] < \varepsilon. \) we get \( \forall b \geq \max\{a, b_1, \ldots, b_N\}, \sup_n E[|X_n|; |X_n| \geq b] \leq \varepsilon, \) and the desired result is obtained.

**Question 3.1** Show an example of \( \{X_n\} \) such that \( X_n \to X \) a.s. and \( EX_n \to EX \), however, \( \{X_n\} \) is not UI.

On a Lebesgue prob. sp. \( (0,1) \), let \( X_n \) be n on \( (0,1/n) \), \(-n\) on \( (1-1/n,1) \), 0 on otherwise. Then \( X_n \to 0 \), \( EX_n \to 0 \), however, if \( n \geq a > 0 \), then \( E[|X_n|; |X_n| \geq a] = E|X_n| = 2 \), and hence, it is not UI.

### 3.2 Radon-Nikodym theorem and conditional expectations

In general, for a finite measure \( \mu \) on \( (\Omega, \mathcal{G}) \) and non-negative integrable \( \mathcal{G}\)-measurable \( f; f \geq 0, \mu\text{-a.e.}, f \in L^1 \), set \( dv = f d\mu \), i.e., \( \nu(A) = \int_A f d\mu \text{ (} A \in \mathcal{F} \) ), then it holds that \( \mu(A) = 0 \implies \nu(A) = 0. \) It is denoted as \( \nu \ll \mu \) and \( \nu \) is called **absolute continuous** with respect to \( \mu \), and \( f \) is called a **density function** of \( \nu \) (w.r.t. \( \mu \)).

Radon-Nikodym Theorem is that the inverse holds.

**Theorem 3.3** (Radon-Nikodym Theorem) Let \( \mu, \nu \) be finite measures on a measurable space \( (\Omega, \mathcal{G}) \). If \( \nu \ll \mu \), then \( \exists ! f; f \geq 0, \mu\text{-a.e.}, f \in L^1(\mu) \) : \( dv = f d\mu \), i.e., \( \nu(A) = \int_A f d\mu \text{ (} A \in \mathcal{G} \) ). The uniqueness means \( \mu\text{-a.e.}, \text{i.e., if } \tilde{f} \text{ satisfies the same conditions, then } f = \tilde{f}, \mu\text{-a.e.}\)

The difference of two finite measures is called a **finite signed measure**. The above theorem holds for a finite signed measure \( \nu \). (Of course, in this case, the condition \( f \geq 0 \mu\text{-a.e.} \) is omitted.)

For the proof, we only describe a construction of a density \( f \). Let \( h \in H \overset{def}{=} \int_A hd\mu \leq \nu(A) \) (\( \forall A \in \mathcal{G} \)) (since a constant \( f \) satisfies this condition, \( H \neq \emptyset \)). Then

\[
\exists h_n \in H; \lim_{n \to \infty} \int_\Omega h_n d\mu = \sup_{h \in H} \int_\Omega hd\mu.
\]
Thus, if we set $f_n := \max_{k \leq n} h_k$, then we can see $f_n \in H$, and hence, by letting $f := \lim f_n$, we have $f \in H$ and $\int \Omega f d\mu = \sup_{b \in H} \int \Omega b d\mu$. This is the desired one. To show this we need a further result; Hahn-Jordan decomposition. However, we don’t describe the detail.

Now by using this result we define a conditional expectation.

Let $X$ be a RV and $\mathcal{G} \subset \mathcal{F}$ be a sub $\sigma$-add. class.

A conditional expectation $Y(\omega) = E[X \mid \mathcal{G}](\omega)$ of $X$ under $\mathcal{G}$ is defined such that $Y$ is $\mathcal{G}$-measurable and $\forall A \in \mathcal{G}, E[Y \mid A] = E[X \mid A]$. Then, $E[E[X \mid \mathcal{G}] \mid A] = E[X \mid A] \ (\forall A \in \mathcal{G})$ holds.

Let $Q(A) := E[X \mid A] (A \in \mathcal{G})$. This is a finite signed measure on $(\Omega, \mathcal{G})$. Clearly, if $P(A) = 0$, then $Q(A) = 0$, i.e., $Q \ll P$; $Q$ is absolute continuous w.r.t. $P$. Hence, by Radon-Nikodym Theorem, $\exists ! Y = Y(\omega); \mathcal{G}$-measurable; $Q(A) = \int A Y dP = E[Y \mid A]$, i.e., $E[X \mid A] = E[Y \mid A] \ (A \in \mathcal{G})$ and this is unique $P$-a.s. Then, we denote $Y$ as $Y(\omega) = E[X \mid \mathcal{G}](\omega)$.

[Properties of a conditional expectations]

Proposition 3.2 Let $X, X_n$ be $\mathcal{F}$-measurable and integrable RVs and $\mathcal{G} \subset \mathcal{F}$ be a sub $\sigma$-add. class. The following hold:

1. $X \in \mathcal{G} \implies E[X \mid \mathcal{G}] = X \ a.s.$
2. For $a_1, a_2 \in \mathbb{R}$, $E[a_1 X_1 + a_2 X_2 \mid \mathcal{G}] = a_1 E[X_1 \mid \mathcal{G}] + a_2 E[X_2 \mid \mathcal{G}] \ a.s.$
3. $X_1 \leq X_2 \ a.s. \implies E[X_1 \mid \mathcal{G}] \leq E[X_2 \mid \mathcal{G}] \ a.s.$
4. $|E[X \mid \mathcal{G}]| \leq E[|X| \mid \mathcal{G}] \ a.s.$
5. $Y \in \mathcal{G}, XY, X \in L^1 \implies E[XY \mid \mathcal{G}] = YE[X \mid \mathcal{G}] \ a.s.$
6. $0 \leq X_n \uparrow X \ a.s. \implies 0 \leq E[X_n \mid \mathcal{G}] \uparrow E[X \mid \mathcal{G}] \ a.s.$
7. Let $1 \leq p, q \leq \infty; 1/p + 1/q = 1$. $X \in L^p, Y \in L^q \implies E[XY \mid \mathcal{G}] \leq E[|X|^p \mid \mathcal{G}]^{1/p} E[|Y|^q \mid \mathcal{G}]^{1/q} \ a.s. \text{ if } 1 < p, q < \infty.$

If $p = 1, q = \infty$, then $E[XY \mid \mathcal{G}] \leq E[|X| \mid \mathcal{G}]||Y||_{\infty}.$ Especially, $1 \leq p_1 < p_2 \leq \infty, X \in L^{p_2} \implies E[|X|^{p_1} \mid \mathcal{G}]^{1/p_1} \leq E[|X|^{p_2} \mid \mathcal{G}]^{1/p_2} \ a.s.$
8. Let $1 \leq p \leq \infty$. $X_n \rightarrow X$ in $L^p \implies E[X_n \mid \mathcal{G}] \rightarrow E[X \mid \mathcal{G}]$ in $L^p$.
9. $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$; sub $\sigma$-add. classes $\implies E[E[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = E[X \mid \mathcal{G}_1] \ a.s.$
10. (Jensen’s inequality) $X \in L^1, \varphi \text{ is convex on } \mathbb{R} \implies \varphi(E[X \mid \mathcal{G}]) \leq E[\varphi(X) \mid \mathcal{G}] \ a.s.$

Proof. We always let $\forall A \in \mathcal{G}$. (1) is clear by $E[X \mid A] = E[X \mid A]$ and $X \in \mathcal{G}$.

(2) Let $Y_1 = E[X_1 \mid \mathcal{G}], Y_2 = E[X_2 \mid \mathcal{G}], Y = E[a_1 X_1 + a_2 X_2 \mid \mathcal{G}]$. Then $E[X_1 \mid A] = E[Y_1 \mid A]$, $E[X_2 \mid A] = E[Y_2 \mid A]$. $E[a_1 X_1 + a_2 X_2 \mid A] = E[a_1 Y_1 \mid A] + a_2 E[X_2 \mid A] = a_1 E[Y_1 \mid A] + a_2 E[X_2 \mid A]$. Then, we have $Y = a_1 Y_1 + a_2 Y_2$ a.s.

(3) $E[E[X_1 \mid \mathcal{G}] \mid A] = E[X_1 \mid A] \leq E[X_2 \mid A] = E[E[X_2 \mid \mathcal{G}] \mid A]$ and both are $\mathcal{G}$-measurable. By the arbitrary of $A \in \mathcal{G}$ we have the result. $\leftarrow$ the next question.

(4) By $-|X| \leq X \leq |X|$ and by the previous result, we have $-E[|X| \mid \mathcal{G}] \leq E[X \mid \mathcal{G}] \leq E[|X| \mid \mathcal{G}]$ a.s.

(5) It is enough to show the case $Y = 1_B \ (B \in \mathcal{G})$. By $E[Y \mid A] = E[X \mid A \cap B] = E[E[X \mid \mathcal{G}] \mid A \cap B] = E[E[X \mid \mathcal{G}] \mid A]$; it is obvious.

(6) By (3) the non-negativity and the monotonicity are clear. It is enough to show $\lim E[X_n \mid \mathcal{G}] = E[X \mid \mathcal{G}]$ a.s. By MCT, we have $E\lim E[X_n \mid \mathcal{G}] \mid A] = \lim E[E[X_n \mid \mathcal{G}] \mid A] = \lim E[X_n \mid A] = E\lim X_n \mid A] = E[X \mid A] = E[E[X \mid \mathcal{G}] \mid A]$. Since both insides are $\mathcal{G}$-measurable and $A \in \mathcal{G}$ is arbitrary, it is clear.
(7) It is possible to show by the same way as in the proof of Hölder’s inequality, because a conditional expectation satisfies linearity, monotonicity and \(|E[X| \mathcal{G}]| \leq E[|X| | \mathcal{G}]| a.s.

(8) is clear, by Hölder the following holds: \(E|E[X_n | \mathcal{G}]-E[X| \mathcal{G}]|^p \leq E[E[X_n - X]^p | \mathcal{G}] = E|X_n - X|^p\).

(9) It is enough to show that for \(\forall A \in \mathcal{G}_1, E[E[X | \mathcal{G}]_2]; A| = E[X; A], \) however it is clear by \(A \in \mathcal{G}_2\).

(10) A convex function can be shown by the supremum of linear functions which are lower than or equal to it. Hence, let \(\phi(x) \geq ax + b = \psi(x)\), then \(E[\phi(X) | \mathcal{G}] \geq E[aX + b | \mathcal{G}] = aE[X | \mathcal{G}] + b = \psi(E[X | \mathcal{G}])\)

Therefore, by taking the supremum on \(\psi (\leq \phi)\) of the last term we get the result.

**Question 3.2** Let \(\mathcal{G} \subset \mathcal{F}\) and \(X, Y \in \mathcal{G}\) and \(\in L^1\). Show that if for \(\forall A \in \mathcal{G}, E[X; A] \leq E[Y; A], \) then \(X \leq Y\) a.s.

**Ans.** It is enough to show the case \(X = 0\) by considering \(Y - X\) as \(Y\). That is, we show that if for \(\forall A \in \mathcal{G}, E[Y; A] \geq 0\), then \(Y \geq 0\) a.s. Set \(A_n = \{Y \leq -1/n\}\), then \(0 \leq E[Y; A_n] \leq -(1/n)P(A_n)\), and \(P(A_n) = 0\). Hence, \(P(Y < 0) = P(\bigcup A_n) = \sum P(A_n) = 0\).

**Proposition 3.3** If \(X \in L^1\) is indep. of \(\mathcal{G}\), then \(E[X | \mathcal{G}] = EX\) a.s. Furthermore, if \(X \in \mathcal{G}\), then \(X = EX\) (constant) a.s. Here note that \(X \) is indep. of \(\mathcal{G} \) \(\iff P(\{X \leq a\} \cap A) = P(X \leq a)P(A)\) \((\forall a \in \mathbb{R}, A \in \mathcal{G})\)

**Proof.** By independence of \(X, \mathcal{G}\), it is easy to see that for \(\forall A \in \mathcal{G}, E[X1_A] = EXE[1_A] = EXP(A)\) holds. This is equivalent to the desired result.

### 3.3 Definition and properties of martingales, Doob’s decomposition

\((M_n, \mathcal{F}_n)_{n \geq 1}\) is a **martingale** if \(\mathcal{F}_n \subset \mathcal{F}\) sub \(\sigma\)-add.classes, \(M_n \in \mathcal{F}_n\) is integrable and satisfies \(E[M_{n+1} | \mathcal{F}_n] = M_n\) a.s. \(\forall n \geq 1\), i.e., \(\forall n \geq 1, \forall A \in \mathcal{F}_n, E[M_{n+1}; A] = E[M_n; A]\).

Instead of this condition, if \(E[M_{n+1} | \mathcal{F}_n] \geq M_n\) a.s. \(\forall n \geq 1\), then it is called a **sub-martingale**. If the inequality is reverse, then it is called a **super-martingale**.

Clearly, if \((M_n)\) is a martingale, then the means are constant, i.e., \(\forall n \geq 1, EM_n = EM_1\). If it is a sub-martingale, then the means are increasing i.e., \(EM_n \uparrow\).

In case of \(n \geq 0\), it is the same we may change \(M_1\) to \(M_0\).

· For a sequence of independent RVs \(\{X_n\}_{n \geq 1}\), let \(M_n = \sum_{k=1}^{n}X_k\) and \(\mathcal{F}_n = \sigma(X_1, \ldots, X_n)\). If \(EX_n = 0\) \((n \geq 1)\), then \((M_n, \mathcal{F}_n)\) is a martingale. (If \(EX_n \geq 0\) \((n \geq 1)\), then it is a sub-martingale.)

· For an integrable RV \(X\) and a filtration \(\{\mathcal{F}_n\}\), let \(M_n := E[X | \mathcal{F}_n]\), then this is a martingale.

**Question 3.3** Check the above two results.

In the following, a filtration \(\{\mathcal{F}_n\}\) is already given and we don’t denote it.

**Proposition 3.4** (1) If \(\{M_n\}\) is a martingale, \(\varphi\) is convex on \(\mathbb{R}\) and \(\varphi(M_n) \in L^1\) \((\forall n \geq 1)\), then \(\{\varphi(M_n)\}\) is a sub-martingale.

(2) If \(\{X_n\}\) is a sub-martingale, \(\varphi\) is convex and increasing on \(\mathbb{R}\) and \(\varphi(X_n) \in L^1\) \((\forall n \geq 1)\), then \(\{\varphi(X_n)\}\) is also a sub-martingale.

(3) For each \(k \geq 1\), let \(\{X_n^{(k)}\}\) be a martingale. Fix \(K < \infty\). Then \(M_n^{(k)} := \max_{k \leq K} X_n^{(k)}\) is a sub-martingale.

**Proof.** (1) By Jensen’s inequality for conditional expectations,

\[E[\varphi(M_{n+1}) | \mathcal{F}_n] \geq \varphi(E[M_{n+1} | \mathcal{F}_n]) = \varphi(M_n)\] a.s.

(2) In the above, the last equation is changed to inequality “≥”, because \(\varphi\) is increasing.

(3) It is enough to show the case of two \((X_n), (Y_n)\). Let \(Z_n := X_n \vee Y_n\). We have \(E|Z_n| \leq E|X_n| + E|Y_n| < \infty\) and

\[X_n \leq E[X_{n+1} | \mathcal{F}_n] \leq E[Z_{n+1} | \mathcal{F}_n], \quad Y_n \leq E[Y_{n+1} | \mathcal{F}_n] \leq E[Z_{n+1} | \mathcal{F}_n],\]

Hence, we have the desired result.
Theorem 3.4 (Doob’s decomposition theorem) If \((X_n)\) is a sub-martingale, then \(\exists \{M_n\}, \{A_n\}; X_n = M_n + A_n\). \((M_n)\) is a martingale, \((A_n)\) is a predictable increasing process starting from 0, i.e., 0 = \(A_1 \leq A_2 \leq \cdots \leq A_n \leq \cdots \leq A = \lim_{n \to \infty} A_n\) almost surely, \(A_n\) is \(\mathcal{F}_{n-1}\)-measurable (that is, predictable.)

Proof. We decompose \(X_n\) as follows, and we may set \(\{\cdot\}\) part be \(M_n\), the rest part be \(A_n\).

\[
X_n = X_1 + \sum_{k=1}^{n-1} (X_{k+1} - X_k) = \left\{ X_1 + \sum_{k=1}^{n-1} (X_{k+1} - E[X_{k+1} | \mathcal{F}_k]) \right\} + \sum_{k=1}^{n-1} (E[X_{k+1} | \mathcal{F}_k] - X_k)
\]

For the uniqueness, if \(X_n = M_n + A_n\), then \(M_n - \tilde{M}_n = \tilde{A}_n - A_n\). This is a martingale and predictable, i.e., \(\mathcal{F}_{n-1}\)-measurable. Therefore, for every \(n \geq 1\),

\[
\tilde{A}_{n+1} - A_{n+1} = E[M_{n+1} - \tilde{M}_{n+1} | \mathcal{F}_n] = M_n - \tilde{M}_n = \tilde{A}_n - A_n \text{ a.s.}
\]

Hence, \(\tilde{A}_n - A_n = \tilde{A}_0 - A_0 = 0. \text{ a.s.}\)

3.4 Stopping times and optional sampling theorem

Let \((\mathcal{F}_n)_{n \geq 0}\) be a filtration, i.e., \(\mathcal{F}_n\) is an increasing sub \(\sigma\)-add. class.

Note that if \(\{X_n\}_{n \geq 0}\) is a stochastic process, then by \((\mathcal{F}_n)\)-adaptability, it holds \(\sigma(X_0, X_1, \ldots, X_n) \subset \mathcal{F}_n\) (\(\forall n \geq 0\)).

An \(Z_\tau = \{0, 1, 2, \ldots\}\)-valued RV \(\tau = \tau(\omega)\) is a stopping time; ST if \(\tau \in \mathcal{F}_n\).

Question 3.4 Show that the following are also ST’s.

1. \(\tau \equiv n\) (a constant time)
2. If \(\sigma, \tau\) are ST’s, then \(\sigma \wedge \tau, \sigma \vee \tau\) are so.
3. For a real-valued process \(\{X_n\}_{n \geq 0}\) and \(B \in \mathcal{B}^1\), a hitting time to \(B\):

\[
\tau_B := \inf\{n \geq 0; X_n \in B\} \quad (= \infty \text{ if } \{\cdot\} = \emptyset).
\]

Note that if we omit the starting point \(X_n\), then changing \(n \geq 0\) to \(n \geq 1\) in the above definition.

Note. In (3), an exit time \(\sigma_B = \sup\{n \geq 0; X_n \in B\} = \infty\) if \(\{\cdot\} = \emptyset\) is not a ST in general (go explain why.)

In the following we assume that

[Assumption] \((\Omega, \mathcal{F}, P)\) is complete and \(\mathcal{F}_0\) contains all null sets.

For a ST \(\tau\), set

\[
\mathcal{F}_\tau := \{ A \in \mathcal{F}; \forall n \geq 0, A \cap \{\tau \leq n\} \in \mathcal{F}_n \}.
\]

(It is possible to change the above “\(A \cap \{\tau \leq n\} \in \mathcal{F}_n\)” to “\(A \cap \{\tau = n\} \in \mathcal{F}_n\)”.

Question 3.5 (1) Show the above \(\mathcal{F}_\tau\) is a \(\sigma\)-add. class and \(\tau\) is \(\mathcal{F}_\tau\)-measurable.

(2) For ST’s \(\sigma, \tau\), if \(\sigma \leq \tau\) a.s., then \(\mathcal{F}_\sigma \subset \mathcal{F}_\tau\).

In (2) by “\(\sigma \leq \tau\) a.s.”, we need the assumption that \(\mathcal{F}_0\) contains all null sets. Because \(A \cap \{\tau \leq n\} = \left( (A \cap \{\sigma \leq n\}) \cap \{\tau \leq n\} \right) \) holds except the difference of a null set \(\{\sigma > \tau\}\),

A martingale means an equitable game. If it is stopped at a stopping time, then what is happened? Does the equity unchange?

The following result ensure it, however, the boundedness of ST’s are needed.

Theorem 3.5 (Optional Sampling Th.) Let \(\{X_n\}\) be a sub-martingale. If ST’s \(\sigma, \tau\) are bounded a.s. and \(\sigma \leq \tau\) a.s., then \(E[X_\tau | \mathcal{F}_\sigma] \geq X_\sigma\) a.s.
The boundedness of ST’s can not be omitted. In fact, let \( \{X_n\} \) be a one-dimensional simple RW starting from 0, which is a martingale. For \( a \leq -1 \), set \( \sigma = 0 \) and \( \tau = \inf\{n \geq 0; X_n \geq a\} \), then they are ST’s such that \( \sigma \leq \tau \). However \( \tau \) is not bounded and \( EX_\sigma = EX_0 = 0, EX_\tau = a < 0 \).

**Question 3.6** Show \( X_\tau \) in the theorem is \( F_\tau \)-measurable and integrable.

We may set \( \sigma \leq \tau \leq 3N \) a.s. By \( X_\tau = \sum_{n \leq N} X_n 1_{\{\tau = n\}} \), the integrability is clear. Moreover, by \( \{X_n \leq a\} \cap \{\tau = n\} \in F_n \), measurability is clear.

**[Proof of Optional Sampling Theorem]**. We may set \( \sigma \leq \tau \leq 3N \) a.s. It is enough to show that for \( \forall A \in F_\sigma, E[X_\tau; A] \geq E[X_\sigma; A] \). Let \( 0 \leq n \leq N \) and set \( A_n := A \cap \{\sigma = n\} \), then \( A_n \in F_n \). Noting that for each \( n \leq k \leq N \), \( A_n \cap \{\tau \geq k + 1\} = A_n \cap \{\tau \leq k\} \) = \( A_n \setminus (A_n \cap \{\tau \leq k\}) \) \( \in F_k \), by sub-martingale property, we have

\[
E[X_k; A_n \cap \{\tau \geq k\}] = E[X_k; A_n \cap \{\tau = k\}] + E[X_k; A_n \cap \{\tau \geq k + 1\}] \\
\leq E[X_\tau; A_n \cap \{\tau = k\}] + E[X_{k+1}; A_n \cap \{\tau \geq k + 1\}] .
\]

Continuing the same calculation for the 2nd term, we have

\[
\leq \sum_{j=k}^N E[X_j; A_n \cap \{\tau = j\}] = E[X_\tau; A_n \cap \{\tau \geq k\}] .
\]

Let \( k = n \), then by \( \sigma \leq \tau \) a.s., we have \( A_n = A \cap \{n = \sigma \leq \tau\} = A_n \cap \{\tau \geq n\} \) except a difference of a null set. Hence, \( E[X_\sigma; A_n] = E[X_n; A_n] \leq E[X_\tau; A_n] \). Thus, by summing on \( 0 \leq n \leq N \), \( E[X_\sigma; A] \leq E[X_\tau; A] \).

**Corollary 3.1 (Optional stopping th.)** Let \((X_n, F_n)\) be a sub-martingale and \( \tau \) be a ST. Then \((X_\tau, F_\tau) := (X_{n \wedge \tau}, F_{n \wedge \tau})\) is also a sub-martingale.

\( n \wedge \tau \) is also a ST, and if \( m < n \), then \( m \wedge \tau \leq n \wedge \tau \leq n \) is bounded. Hence, by OST it is clear.

### 3.5 Sub-martingale inequalities and convergence theorems

Kolmogorov’s maximal inequality for a sum of independent RVs which is used in the proof of string LLN, can be extended to the following for a martingale:

**Theorem 3.6 (sub-martingale inequality)** (1) Let \((X_n)\) be a sub-martingale. Then

\[
\forall a > 0, aP(\max_{k \leq n} X_k \geq a) \leq E[X_n; \max_{k \leq n} X_k \geq a] \leq EX_n^+ .
\]

(2) Especially, if \((X_n)\) is a sub-martingale and \( X_n \geq 0 \) a.s., then \( aP(\max_{k \leq n} X_k \geq a) \leq EX_n^+ \) \( \forall a > 0 \). Moreover, if for some \( p > 1 \), \( X_n \in L^p \), then

\[
\left[ E \max_{k \leq n} X_k^p \right]^{1/p} \leq \frac{p}{p-1} ||X_n||_p .
\]

If \( \{M_n\} \) is a martingale, then \( \{|M_n|\} \) is a sub-martingale. On the application, the following is useful:

**Corollary 3.2** If \((M_n)\) is a martingale, then \( aP(\max_{k \leq n} |M_k| \geq a) \leq E|M_n| \) \( \forall a > 0 \).

Moreover, if \( M_n \in L^p \), then

\[
\left[ E \max_{k \leq n} |M_k|^p \right]^{1/p} \leq \frac{p}{p-1} ||M_n||_p .
\]

Furthermore, the following also holds:
Corollary 3.3 Let \( n \geq 0 \). If \( (X_n) \) is a sub-martingale, then
\[
\forall a > 0, a P(\min_{k \leq n} X_k \leq -a) \leq E X_n - E[X_n; \min_{k \leq n} X_k \leq -a] - E X_0 \leq E X_n^+ - E X_0.
\]
Note that if \( n \geq 1 \), then \( X_0 \) may be changed to \( X_1 \).

[Proof of sub-martingale inequality] (1) Let \( (X_n) \) be a sub-martingale and \( \forall a > 0 \). We divide an event \( A = \{\max_{k \leq n} X_k \geq a \} \) by using first times such that \( X_k \geq a \), i.e., let
\[
A_0 = \{X_0 \geq a\}, \quad A_k = \{X_k \geq a, \forall j \leq k-1, X_j < a\},
\]
then \( A = \bigcup_{k \leq n} A_k \) is a disjoint union and \( A_k \in \mathcal{F}_k \). Hence, we have
\[
E[X_n; A] = \sum_{k \leq n} E[X_n; A_k] \geq \sum_{k \leq n} E[X_k; A] \geq a \sum_{k \leq n} P(A_k) = a P(A).
\]

(2) On the later half, we use
\[
p \int_0^{\infty} a^{p-1} 1_{\{a \leq Y\}} \, da = p \int_0^Y a^{p-1} \, da = Y^p.
\]
Let \( Y = \max_{k \leq n} |X_k| \). Then by the above and martingale ineq., we have
\[
E Y^p = p \int_0^{\infty} a^{p-1} P(Y \geq a) \, da \leq p \int_0^{\infty} a^{p-1} \frac{1}{a} E[|X_n|; Y \geq a] \, da
\]
\[
= p E \left[ |X_n| \int_0^Y a^{p-2} \, da \right] = \frac{p}{p-1} E \left[ |X_n| Y^{p-1} \right].
\]
Moreover, by using Hölder for the last term, noting that \( 1/p = 1 - 1/p = (p-1)/p \),
\[
(\text{The last term}) \leq \frac{p}{p-1} \|X_n\|_p (E Y^p)^{1/q}.
\]
The desired inequality is obtained.

[Proof of Corollary 3.3] Let \( \tau \) be a hitting time to \((\infty, a)\), i.e.,
\[
\tau = \min\{k \leq n; X_k \leq -a\} (= \infty \text{ if } \{\} = \emptyset).
\]
Moreover, let \( \sigma = \tau \wedge n \). Then these are ST’s and \( \sigma \leq n \) is bounded. \( B = \{\min_{k \leq n} X_k \leq -a\} = \{\tau \leq n\} = \bigcup_{k \leq n} B_k \), where \( B_k = \{\tau = k\} \) is an event of that \( X_k \) is equal to or lower than \( -a \) at first at time \( k \). Furthermore, if \( k < n \), then \( B_k = \{\sigma = k\} \) and \( B_n \subset \{\sigma = n\} \) (which contains \( \{\tau = \infty\} \)). Thus, \( X_\sigma = X_k \leq -a \) on \( B_k \). Hence, by OST, \( E[X_\sigma| \mathcal{F}_n] \geq X_\sigma \) a.s. and
\[
E X_0 \leq E X_\sigma = E[X_\sigma; B] + E[X_\sigma; B^c] = \sum_{k \leq n} E[X_\sigma; B_k] + E[X_\sigma; B^c] \leq -a P(B) + E[X_\sigma; B^c].
\]
Therefore, noting that \( B^c \cap \{\sigma = k\} = \emptyset \in \mathcal{F}_k \), that is, \( B^c \in \mathcal{F}_\sigma \), we have
\[
a P(B) \leq E[X_\sigma; B^c] - E X_0 \leq E[X_n; B^c] - E X_0 \leq E[X_n^+] - E X_0.
\]

Convergence Theorem of Martingales

**Theorem 3.7** (Sub-martingale convergence theorem) If a sub-martingale \((X_n, \mathcal{F}_n)\) satisfies \(\sup_n E[X_n^+] < \infty\), then \(X_n\) converges a.s., i.e., \(P(\lim X_n) = 1\).

The condition for a sub-martingale: \(\sup_n E[X_n^+] < \infty\) is equivalent to \(\sup_n E[X_n] < \infty\). (It is obvious from \(E[X_n^+] - E[X_n^-] \geq EX_1\), i.e., \(E[X_n^-] \leq E[X_n^+] - EX_1\).)

In order to show this result, we need the following result of “cross number of a sub-martingale”

**Theorem 3.8** (Cross Number Theorem) Let \(H_N\) be a cross number of a sample path of a sub-martingale \((X_n, \mathcal{F}_n)\) for an interval \((a, b)\) from left to right until \(n \leq N\). Then it holds \((b - a)EH_N \leq E[(X_N - a)^+]\).

[Proof of convergence theorem of martingales]

\[
\{\liminf X_n < \limsup X_n\} \subset \bigcup_{a < b} \{\liminf X_n < a < b < \limsup X_n\}.
\]

Let the right hand be \(A_{a,b}\) and it is enough to show \(P(A_{a,b}) = 0\). On this event, \((X_n)\) crosses infinitely many times for the interval \((a, b)\) (from left to right). Let \(H\) be a cross number of \((X_n)\) from \(a\) to \(b\). Then it is enough to show \(P(H = \infty) = 0\). Let \(H_N\) be a cross number of \((X_n)\) from \(a\) to \(b\). Then by Cross Number Theorem, we have

\[
EH_N \leq \frac{E[(X_N - a)^+]}{(b - a)} \leq \sup_{N \geq 1} \frac{EX_N^+ + |a|}{(b - a)} < \infty.
\]

and by monotonicity and by convergence theorem, \(0 \leq EH_N \uparrow EH\). Thus, \(EH < \infty\). This means \(H < \infty\) a.s., i.e., 

\(P(H = \infty) = 0\).

In general, let \(H_N(\omega) = H_N(\omega; a, b)\) be a cross number of a sample path \(\{X_n(\omega)\}_{1 \leq n \leq N}\) of a stochastic process \(\{X_1, \ldots, X_N\}\) for an interval \((a, b)\) from left to right. This can be expressed by using hitting times; For \(n > N\), let \(X_n \equiv X_N\) and let \(\tau = \min\{n \geq 1; X_n \in (\infty, a]\}, \tau_2 = \min\{n \geq \tau; X_n \in [b, \infty)\}\), and similarly we define \(\tau_1, \tau_3, \ldots\) (if \(\{\} = \emptyset\) then they are \(\infty\)). These are ST’s (\(\rightarrow\) the next question). If we set \(m = \max\{n \leq N; \tau < \infty\}\), then it can be defined as \(H_N = H_N(a, b) := [m/2] \geq 0\) (\([\cdot]\) is a Gaussian notion, i.e., an integer part). Corresponding to \(2k - 1 \leq m, 2k < m\), it holds \(X_{\tau_2k-1} \leq a, X_{\tau_2k} \geq b\).

For simplicity, denote \(H = H_N, H = [m/2]\), i.e., \(m = 2H\) or \(2H + 1\).

(i) In case of \(m = 2H + 1\), then \(X_N < b\) (if not, then \(\tau_{2H+2}\) is finite) and

\[
\sum_{k=1}^{H} (X_{\tau_{2k+1}} - X_{\tau_{2k}}) \leq (a - b)H = -(b - a)H.
\]

(ii) In case of \(m = 2H\), then \(X_N > a\), and

\[
\sum_{k=1}^{H-1} (X_{\tau_{2k+1}} - X_{\tau_{2k}}) + (X_N - X_{\tau_{2H}}) = \sum_{k=1}^{H-1} (X_{\tau_{2k+1}} - X_{\tau_{2k}}) + (a - X_{\tau_{2H}}) + (X_N - a) \\
\leq (a - b)H + (X_N - a).
\]

Hence, let \(Y_k = X_{\tau_k}\) if \(k \leq m\), \(X_N\) if \(k > m\). By \(H \leq N\), the above inequality is

\[
\sum_{k=1}^{N} (Y_{2k+1} - Y_{2k}) = \sum_{k=1}^{H} (Y_{2k+1} - Y_{2k}) \leq -(b - a)H + (X_N - a)^+.
\]

**Question 3.7** Show the above \(\tau_k\) is a ST.
\[ \tau_1 = n = \{ X_n \leq -a, X_1, \ldots X_n-1 > a \}, \quad \tau_2 = n = \{ X_n \geq b, \tau_1 < n, X_{\tau_1+1}, \ldots X_{n-1} < b \} \] are both in \( \mathcal{F}_n \).

**Proof of Cross Number Theorem** For each \( k, \tau_k \wedge N \) is a bounded \( (\mathcal{F}_n) \)-ST. Let \( G_k = \mathcal{F}_{\tau_k} \) and \( Y_k = X_{\tau_k \wedge N} \). Then by OST, \((Y_k, G_k)\) is also a sub-martingale. Hence, 0 \( \leq \sum_{k=1}^{N} E[Y_{2k+1} - Y_{2k}] \leq - (b-a) E H_N + E[(X_N - a)^+] \). \]

**Theorem 3.9** For a sub-martingale \((X_n, \mathcal{F}_n)\), the following are equivalent:

1. \( \{X_n\} \) is UI.
2. \( \{X_n\} \) converges in \( L^1 \), i.e., \( X_n \rightarrow \mathbb{E} X \) in \( L^1 \).
3. \( \{X_n\} \) converges a.s. and set \( X = \lim X_n \), then \( X \in L^1 \), \( EX_n \rightarrow EX \), \( E[X | \mathcal{F}_n] \geq X_n \) a.s.

**Proof.** Recall that under the conditions \( X_n \in L^1, X_n \rightarrow X \) a.s., the following are equivalent: \( \{X_n\} \):

1. UI, \( E[|X_n - X|] \rightarrow 0 \), \( E[X_n] \rightarrow E[X] \).
2. (1) \( \Rightarrow \) (2): UI implies (UI) \( \sup E[|X_n|] < \infty \), Hence by sub-martingale convergence theorem we have \( X_n \rightarrow \mathbb{E} \mathbb{X} \) a.s. and in \( L^1 \) by UI.
3. (2) \( \Rightarrow \) (3): \( |E[X_n] - E[X]| \leq E[|X_n - X|] \rightarrow 0 \) and \( \sup E[|X_n|] < \infty \), because a convergence sequence is bounded. By sub-martingale convergence theorem, we have \( X_n \rightarrow \mathbb{E} \mathbb{X} \) a.s. On the other hand \( L^1 \) convergence implies \( X_n \rightarrow X \) a.s. for a suitable sub-sequence. Thus, \( X = X \) a.s. It remains to show that \( E[X | \mathcal{F}_n] \geq X_n \) a.s. By sub-martingale property, for \( n \in \mathbb{N}, \forall A \in \mathcal{F}_n, E[X_{n+k+1} | A] \geq E[X_{n+k} | A] \) (\( \forall k \geq 1 \)). Hence, letting \( k \rightarrow \infty \), we have \( E[X | A] \geq E[X_n | A] \). The result i obtained.
4. (3) \( \Rightarrow \) (1): By convergence a.s. it is enough to show \( E[X_n | A] \rightarrow E[X] \), because we have \( (X_n) \): UI. By the assumption, for \( n \in \mathbb{N}, \forall A \in \mathcal{F}_n, E[X_{n+k+1} | A] \geq E[X_{n+k} | A] \). Let \( A = \{ X_n \geq 0 \} \), then \( EX_n^+ \leq EX^+ \), i.e, \( \sup EX_n^+ \leq EX^+ < \infty \). On the other hand, by \( X_n^+ \rightarrow X^+ \) a.s. and by Fatou, we have \( \liminf EX_n^+ \geq EX^+ \). Hence, \( \lim EX_n^+ = EX^+ \). Similarly, we have \( \lim EX_n^- = EX^- \). Therefore, \( E[X_n | A] = E[X] \).
4 Continuous-time Markov Chains

Let \( t \geq 0 \) be a continuous-time parameter. Let \( S \) be a countable set. An \( S \)-valued RVs (a stochastic process) \( (X_t)_{t \geq 0} \) is a continuous-time Markov chain if it has the following Markov property:

For \( s, t \geq 0, i, j, k, u, \ell \in S, 0 \leq u \ell < s (\ell \leq \ell_0), \)
\[
P(X_{t+s} = j \mid X_s = i, X_{u\ell} = k) = P(X_{t+s} = j \mid X_s = i).
\]

Moreover, we also assume the time-homogeneity:
\[
P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i).
\]
We define this as a transition probability; \( q_t(i, j) = P(X_t = j \mid X_0 = i) \).

4.1 Exponential times

In order to construct a continuous-time Markov chain from a discrete-time Markov chain, we can consider changing jump intervals to random. Hence, we introduce an exponential time (= a random time which has an exponential distribution).

**Definition 4.1** For a constant \( \alpha > 0 \), a random variable \( \tau = \tau(\omega) \) is distributed by an exponential distribution with parameter \( \alpha \), i.e,
\[
P(\tau > t) = \int_t^\infty e^{-\alpha s} ds = e^{-\alpha t}
\]
In other word, \( \tau \) has a distribution with a density function \( f(s) = \alpha e^{-\alpha s} \). In this text, we call \( \tau \) as an exponential time or simply, an exponential time.

Its mean and variance are the following:
\[
E[\tau] = \int_0^\infty \alpha e^{-\alpha s} ds = \frac{1}{\alpha}, \quad V(\tau) = E[\tau^2] - (E[\tau])^2 = \frac{1}{\alpha^2}.
\]

**Question 4.1** Make sure the above calculation of variance.

**Proposition 4.1** If \( \tau \) is an exponential time, then it has the following memoryless property. For \( t, s \geq 0, \)
\[
P(\tau > t + s \mid \tau > s) = P(\tau > t).
\]

**Proof.**
\[
P(\tau > t + s \mid \tau > s) = \frac{P(\tau > t + s)}{P(\tau > s)} = \frac{e^{-(t+s)}}{e^{-s}} = e^{-t} = P(\tau > t).
\]

**Proposition 4.2** If \( \tau_1, \tau_2, \ldots, \tau_n \) are independent \( \alpha_1, \alpha_2, \ldots, \alpha_n \)-exponential times, respectively, then \( \min\{\tau_1, \tau_2, \ldots, \tau_n\} \) is \((\alpha_1 + \alpha_2 + \cdots + \alpha_n)\)-exponential time. Moreover,
\[
P(\min\{\tau_1, \tau_2, \ldots, \tau_n\} = \tau_k) = \frac{\alpha_k}{\alpha_1 + \alpha_2 + \cdots + \alpha_n}.
\]

**Proof.** For simplicity, we only show the case of \( n = 2, k = 1 \).
\[
P(\tau_1 < \tau_2 > t) = P(\tau_1 > t, \tau_2 > t) = P(\tau_1 > t)P(\tau_2 > t) = e^{-(\alpha_1 + \alpha_2)t}.
\]
Moreover, since the joint distribution of $\tau_1, \tau_2$ is the product of each ones by their independence, we have

$$P(\min\{\tau_1, \tau_2\} = \tau_1) = P(\tau_1 < \tau_2)$$

$$= \int_0^\infty d\alpha_1 e^{-\alpha_1 s} P(s < \tau_2)$$

$$= \int_0^\infty d\alpha_1 e^{-\alpha_1 s} e^{-\alpha_2 s}$$

$$= \frac{\alpha_1}{\alpha_1 + \alpha_2}.$$

The other cases are the same.

**Example 4.1** There is a system of two devices $A$ and $B$. The time to failure of $A$ is an 1-exp. time and the time to failure of $B$ is an 2-exp. time. These are failure independent and the system is failure if at least one is failure. Find the mean time to failure of the system.

By the previous proposition, the time to failure of the system is a 3-exp. time, and hence, the mean is $1/3$.

### 4.2 Poisson processes

We describe a Poisson process as a simple example of a continuous-time Markov chain.

**Definition 4.2** For $\lambda > 0$, a stochastic process $(X_t)_{t \geq 0}$ is a Poisson process with a parameter $\lambda$ (it is simply called a $\lambda$-Poisson process) if the following holds:

1. $X_0 = 0$,
2. For $0 \leq s < t$, $X_t - X_s$ has a Poisson distribution with a parameter $\lambda(t - s)$, i.e.,

$$P(X_t - X_s = k) = e^{-\lambda(t-s)} \frac{\lambda^k (t-s)^k}{k!} \quad (k = 0, 1, 2, \ldots).$$
3. $X_t$ has independent increments. That is, for $0 < t_1 < t_2 < \cdots < t_n$, $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

**Theorem 4.1** A Poisson process is a continuous time Markov chain.

It is easy to see by the above independent increments.

**Question 4.2** Let $S$ be a countable set. Show in general, if an $S$-valued continuous-time stochastic process starting from 0 has independent increments, then it is a continuous-time Markov chain.

**Ans.** Let $X_t$ be the process satisfies the assumption. For $0 \leq t_1 < t_2 < \cdots < t_n < t_{n+1}$. By using the independence of $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_{n+1}} - X_{t_n}$, and by a similar way to the discrete-time case, we can show the independence of $X_{t_{n+1}} - X_{t_n}, (X_{t_1}, \ldots, X_{t_n})$, and of $X_{t_{n+1}} - X_{t_n}, X_{t_n}$. From these we have Markov property:

$$P(X_{t_{n+1}} = j_{n+1} | X_{t_n} = j_k, 0 \leq k \leq n) = P(X_{t_{n+1}} - X_{t_n} = j_{n+1} - j_k | X_{t_k} = j_k, 0 \leq k \leq n)$$

$$= P(X_{t_{n+1}} - X_{t_n} = j_{n+1} - j_k)$$

$$= P(X_{t_{n+1}} = j_{n+1} - j_k | X_{t_n} = j_k)$$

$$= P(X_{t_{n+1}} = j_{n+1} | X_{t_n} = j_k).$$
Theorem 4.2 (Construction of a Poisson process) Let $\sigma_1, \sigma_2, \ldots$ be independent $\lambda$-exponential times. Let $\tau_n = \sum_{k=1}^{n} \sigma_k$ and $\tau_0 = 0$. Define

$$X_t = n \iff \tau_n \leq t < \tau_{n+1},$$

that is,

$$X_t := \sum_{n=0}^{\infty} n 1_{[\tau_n, \tau_{n+1})}(t) = \max\{n; \tau_n \leq t\}.$$

Then, $(X_t)$ is a $\lambda$-Poisson process.

Note The inverse of the above result holds, that is, if $(X_t)_{t \geq 0}$ is a $\lambda$-Poisson process and let $\tau_1, \tau_2, \ldots$ be jump times of it, then $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots$ are i.i.d. and each of them is a $\lambda$-exponential time.

In order to show the above result, we use the following.

Proposition 4.3 The sum of independent $n$-number of $\lambda$-exponential times $\sigma_k$: $\tau = \sum_{k=1}^{n} \sigma_k$ is distributed by the gamma distribution $\Gamma(n, \lambda)$, i.e.,

$$P(\tau < t) = \int_{0}^{t} \frac{1}{(n-1)!} \lambda^n s^{n-1} e^{-\lambda s} ds.$$

Proof. By the independence of $(\sigma_n)$,

$$P(\sigma_1 + \cdots + \sigma_n < t) = \int_{s_1 + \cdots + s_n < t} \lambda^n e^{-\lambda (s_1 + \cdots + s_n)} ds_1 \cdots ds_n.$$

By the change of variables such that $u_k = s_1 + \cdots + s_k$ ($k = 1, \ldots, n$) and $s = u_n$,

$$\int_{s_1 + \cdots + s_n < t} \lambda^n e^{-\lambda (s_1 + \cdots + s_n)} ds_1 \cdots ds_n = \int_{0}^{t} du_n \int_{0}^{u_n} du_{n-1} \cdots \int_{0}^{u_2} du_1 \lambda^n e^{-\lambda u_n}$$

$$= \int_{0}^{t} du_n \int_{0}^{u_n} du_{n-1} \cdots \int_{0}^{u_2} du_2 \lambda^n e^{-\lambda u_n}$$

$$= \int_{0}^{t} du_n \frac{1}{(n-1)!} u_n^{n-1} \lambda^n e^{-\lambda u_n}$$

$$= \int_{0}^{t} ds \frac{1}{(n-1)!} \lambda^n s^{n-1} e^{-\lambda s}.$$

[Proof of Theorem 4.2]. Since $\tau_n$ is independent of $\sigma_{n+1}$ and distributed by $\Gamma(n, \lambda)$, we have

$$P(X_t = n) = P(\tau_n \leq t < \tau_{n+1} = \tau_n + \sigma_{n+1})$$

$$= \int_{0}^{t} ds \frac{1}{(n-1)!} \lambda^n s^{n-1} e^{-\lambda s} P(t < s + \sigma_{n+1})$$

$$= \int_{0}^{t} ds \frac{1}{(n-1)!} \lambda^n s^{n-1} e^{-\lambda s} e^{-(t-s)\lambda}$$

$$= e^{-\lambda t} \frac{\lambda^n}{(n-1)!} \int_{0}^{t} s^{n-1} ds = e^{-\lambda t} \frac{\lambda^n t^n}{n!}.$$

By a similar way,

$$P(\tau_{n+1} > t + s, X_t = n) = P(\tau_{n+1} > t + s, \tau_n \leq t < \tau_{n+1})$$

$$= P(\tau_n + \sigma_{n+1} > t + s, \tau_n \leq t)$$

$$= \int_{0}^{t} du \frac{1}{(n-1)!} \lambda^n u^{n-1} e^{-\lambda u} P(u + \sigma_{n+1} > t + s)$$

$$= \int_{0}^{t} du \frac{1}{(n-1)!} \lambda^n u^{n-1} e^{-\lambda u} e^{-\lambda(t+s-u)} = e^{-\lambda(t+s)} \frac{\lambda^n t^n}{n!}.$$
Hence,

\[(4.1) \quad P(\tau_{n+1} > t + s \mid X_t = n) = e^{-\lambda s} = P(\sigma_1 > s).\]

Moreover,

\[(4.2) \quad \text{under the condition } X_t = n, \tau_{n+1} - t, \sigma_{n+2}, \ldots, \sigma_{n+m} \text{ has the same distribution as } \sigma_1, \sigma_2, \ldots, \sigma_m.\]

In fact,

\[
P(\tau_{n+1} - t > s_1, \sigma_{n+2} > s_2, \ldots, \sigma_{n+m} > s_m \mid X_t = n)\]
\[
= P(\tau_n \leq t < \tau_{n+1}, \tau_{n+1} - t > s_1, \sigma_{n+2} > s_2, \ldots, \sigma_{n+m} > s_m) / P(X_t = n)\]
\[
= P(\tau_n \leq t, \tau_{n+1} - t > s_1) P(\sigma_{n+2} > s_2, \ldots, \sigma_{n+m} > s_m) / P(X_t = n)\]
\[
= P(\tau_n \leq t, \tau_{n+1} - t > s_1, X_t = n) P(\sigma_2 > s_2, \ldots, \sigma_m > s_m)\]
\[
= P(\sigma_1 > s) P(\sigma_2 > s_2, \ldots, \sigma_m > s_m)\]
\[
= P(\sigma_1 > s, \sigma_2 > s_2, \ldots, \sigma_m > s_m).\]

By this and noting that \(\tau_{n+m} - t = (\tau_{n+1} - t) + \sigma_{n+2} + \cdots + \tau_{n+m},\) we have in general, for \(m \geq 1,\) we can get

\[
P(\tau_{n+m} > t + s \mid X_t = n) = P(\tau_m > s).\]

By subtracting the above from the above with \(m + 1\) instead of \(m,\) we have

\[
P(\tau_{n+m} \leq t + s < \tau_{n+m+1} \mid X_t = n) = P(\tau_m \leq s < \tau_{m+1}) = P(X_s = m).\]

By using this, for \(n \geq 0, m \geq 1,\)

\[
P(X_t = n, X_{t+s} - X_t = m) = P(X_t = n, X_{t+s} = n + m) = P(X_t = n) P(X_{t+s} = n + m \mid X_t = n)\]
\[
= P(X_t = n) P(\tau_{n+m} \leq t + s < \tau_{n+m+1} \mid X_t = n) = P(X_t = n) P(X_s = m).\]

By summing on \(n \geq 0,\)

\[
P(X_{t+s} - X_t = m) = P(X_s = m) = e^{-\lambda \sum_{k=1}^{m} s_k} / m!.\]

In case of \(m = 0,\) it can be seen \(P(X_{t+s} - X_t = m) = e^{-\lambda s},\) and this is included in the above. In fact, by

\[
P(\tau_n > t + s \mid X_t = n) = P(\tau_n > t + s \mid \tau_n \leq t < \tau_{n+1}) = 0,\]

if we subtract this from (4.1), then

\[
P(X_{t+s} = n \mid X_t = n) = P(\tau_n \leq t + s < \tau_{n+1} \mid X_t = n) = e^{-\lambda s}.\]

Thus,

\[
P(X_t = n, X_{t+s} - X_t = 0) = P(X_t = n, X_{t+s} = n) = P(\tau_n = t + s < \tau_{n+1} \mid X_t = n) = P(X_t = n) e^{-\lambda s}.\]

Hence, by summing on \(n \geq 0,\) we have \(P(X_{t+s} - X_t = 0) = e^{-\lambda s}.\)

Finally on the independence of increments, by using (4.2), we have for \(0 \leq t_1 < \cdots < t_k,\)

\[
P(X_{t_0} = n_0, X_{t_1} - X_{t_0} = n_1, \ldots, X_{t_k} - X_{t_{k-1}} = n_k)\]
\[
= P(X_{t_0} = n_0, X_{t_1} = n_0 + n_1, \ldots, X_{t_k} = n_0 + \cdots + n_k)\]
\[
= P(X_{t_0} = n_0) P(X_{t_1 - t_0} = n_1, \ldots, X_{t_k - t_0} = n_1 + \cdots + n_k).\]
Therefore, by repeating this, we have the following independent increments:

\[ P(X_{t_k} = n_0, X_{t_1} - X_{t_0} = n_1, \ldots, X_{t_k} - X_{t_{k-1}} = n_k) \]
\[ = P(X_{t_0} = n_0)P(X_{t_1} - t_0 = n_1) \cdots P(X_{t_k} - t_{k-1} = n_k) \]
\[ = P(X_{t_0} = n_0)P(X_{t_1} - X_{t_0} = n_1) \cdots P(X_{t_k} - X_{t_{k-1}} = n_k). \]

**Example 4.2** The number of times of calling to a fire station is a 20-times per hour Poisson process, i.e., a 20-Poisson process. In that only 20 percent is urgent. Then, is the number of times of those requiring urgent a Poisson process? If it is so, then what is the parameter?

The answer is a 4-times per hour Poisson process and it is easily obtained from the following proposition:

**Proposition 4.4** Let \( X_t \) be a \( \lambda \)-Poisson process. The jumps of \( X_t \) has two kinds of type I and type II. They are independent and each appears with probability \( p \) or \( 1-p \), respectively. Let the process of jumps of type I only be \( Y_t \), and type II only be \( Z_t \). Then, they are independent and each is \( \lambda p \) or \( \lambda(1-p) \)-Poisson process, respectively.

**Proof.** By \( X_t = Y_t + Z_t \), under \( X_t = n + k \), \( Y_t = k \) means \( k \)-times of jumps are chosen from \( n+k \)-times with probability \( p \).

\[ P(Y_t = k, Z_t = n) = P(Y_t = k| X_t = n + k) = \binom{n+k}{k} p^k(1-p)^n. \]

Hence,

\[ P(Y_t = k, Z_t = n) = P(Y_t = k, Z_t = n| X_t = n + k)P(X_t = n + k) \]
\[ = \binom{n+k}{k} p^k(1-p)^n e^{-\lambda t} \frac{(\lambda t)^{n+k}}{(n+k)!} \]
\[ = e^{-\lambda pt} \frac{(\lambda pt)^k}{k!} \]
\[ e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^n}{n!}. \]

**Question 4.3** Check the last equation.

By summing in \( n \geq 0 \) on both sides, we have

\[ P(Y_t = k) = e^{-\lambda pt} \frac{(\lambda pt)^k}{k!}. \]

Similarly, by summing in \( k \geq 0 \),

\[ P(Z_t = n) = e^{-\lambda(1-p)t} \frac{(\lambda(1-p)t)^n}{n!}. \]

Moreover, by the above three equations, we have

\[ P(Y_t = k, Z_t = n) = P(Y_t = k)P(Z_t = n). \]

Therefore, \( Y_t, Z_t \) are independent and each has \( \lambda p \) or \( \lambda(1-p) \)-Poisson distribution. For \( Y_t - Y_s \), by a similar way to the above, we have

\[ P(Y_{t+s} - Y_s = k) = \sum_{n \geq 0} P(Y_{t+s} - Y_s = k| X_{t+s} - X_s = n + k)P(X_{t+s} - X_s = n + k) \]
\[ = \sum_{n \geq 0} \binom{n+k}{k} p^k(1-p)^n e^{-\lambda t} \frac{(\lambda t)^{n+k}}{(n+k)!} \]
\[ = e^{-\lambda pt} \frac{(\lambda pt)^k}{k!}. \]
Thus, \( P(Y_{t+s} - Y_s = k) = P(Y_t = k) \). Furthermore, by the same way we can show
\[
(4.3) \quad P(Y_s = k_1, Y_{t+s} - Y_s = k_2, Z_s = n_1, Z_{t+s} - Z_s = n_2)
= P(Y_s = k_1)P(Y_{t+s} - Y_s = k_2)P(Z_s = n_1)P(Z_{t+s} - Z_s = n_2).
\]

(Make condition by \( \{X_s = k_1 + n_1, X_{t+s} - X_s = k_2 + n_2\} \) and use the independence of them.) Hence we have the independent increments of \((Y_t), (Z_t)\), respectively, and \((Y_t)\) is a \( \lambda p \)-Poisson process and \((Z_t)\) is a \( \lambda (1 - p) \)-Poisson process. Moreover, it is easy to see that
\[
(4.4) \quad P(Y_s = k_1, Y_{t+s} = k_1 + k_2, Z_s = n_1, Z_{t+s} = n_1 + n_2)
= P(Y_s = k_1, Y_{t+s} = k_1 + k_2)P(Z_s = n_1, Z_{t+s} = n_1 + n_2)
\]
and hence, \( \{Y_s, Y_{s+t}\} \) and \( \{Z_s, Z_{s+t}\} \) are independent and more general, we have for \( 0 \leq t_1 < t_2 < \cdots < t_m \), \( \{Y_{t_1}, \ldots, Y_{t_m}\} \) and \( \{Z_{t_1}, \ldots, Z_{t_m}\} \) are independent. This means independence of \((Y_t), (Z_t)\) as processes.

**Question 4.4** Show (4.3) in the above proof, and show (4.4) from this.

### 4.3 Continuous-time random walks

Let \( S \) be a countable linear space. Let \( (p_j)_{j \in S} \) be a distribution. A stochastic process \((X_t)_{t \geq 0}\) which jumps from \( i \in S \) to \( i + j \) with probability \( p_j \) in each independent 1-exponential times is called a **continuous-time random walk**.

This is constructed as \( X_t := Y_{S_t} \) by using a discrete-time RW \((Y_n)_{n \geq 0}\) with one-step distribution \((p_j)\) and a 1-Poisson process \((S_t)\) and it is independent of \((Y_n)\).

Since \((Y_n)\) and \((S_t)\) have independent increments, \((X_t)\) also has independent increments. Hence, by Question 4.2 it is a continuous-time Markov chain.

### 4.4 Continuous-time Galton-Watson processes

Let \( \lambda > 0 \). There are several particles and each divides independently into \( k \geq 0 \) particles if \( k = 0 \), then it exterminates with probability \( p_k \) after \( \lambda \)-exponential time. Each divided particles divides or exterminates independently under the same law. Then, at the time \( t \), we denote the total number of particles as \( Z_t \) and it is called a **continuous-time Galton-Watson process**.

This is constructed as follows: Let \( \{X_n\} \) be a discrete-time GW-process with offspring probability \( (p_k) \), and \((S_t)\) be an independent \( \lambda \)-Poisson process, and set \( Z_t := X_{S_t} \).

The mean of offspring number is
\[
m := \sum_{k \geq 1} kp_k.
\]

**Theorem 4.3** Let \( 0 < p_0 + p_1 < 1 \).
\[
P(\forall t \geq 0, Z_t \geq 1) > 0 \iff m > 1.
\]

Moreover, for \( t \geq 0 \), \( E[Z_t \mid Z_0 = 1] = e^{\lambda(m-1) t} \).

**Proof.** By the result of the discrete-time case and by the above construction, the first half is easily obtained. We consider the expectation. Let \( E_1[\ast] := E[\ast] \mid Z_0 = 1 \). By \( Z_t = X_{S_t} \) and \( E_1[X_n] = m^n \),
\[
E_1[Z_t] = \sum_{n=0}^{\infty} E_1[Z_t \mid S_t = n]P(S_t = n) = \sum_{n=0}^{\infty} E_1[X_n \mid S_t = n]P(S_t = n)
= \sum_{n=0}^{\infty} E_1[X_n]P(S_t = n) = \sum_{n=0}^{\infty} m^n e^{-\lambda t} \frac{(\lambda t)^n}{n!} = e^{\lambda(m-1) t}
\]

\[\Box\]
4.5 Continuous-time Markov chains & transition probabilities

Let $S$ be a countable set. An $S$-valued continuous Markov chain $(X_t)_{t\geq 0}$ can be defined by the same way as a continuous-time RW. That is, $X_t := Y_{S_t}$, where $(Y_n)_{n\geq 0}$ is a discrete time Markov chain with transition probability $p(i, j)$ and $(S_t)_{t\geq 0}$ is a 1-Poisson process and independent of $(Y_n)$.

Moreover, we have the transition probability $P(X_t = j \mid X_s = i, X_{u \leq t} = k_{u \leq t} (\ell \leq t_0)) = P(X_t = j \mid X_0 = i) =: q_t(i, j)$.

Moreover, $q_t(i, j)$ can be obtained by using an $n$-th transition probability $p_n(i, j)$ as follows;

$$q_t(i, j) = \sum_{n \geq 0} e^{-t/n} p_n(i, j).$$

[Proof of that $X_t := Y_{S_t}$ is a Markov chain] For simplicity, let the RHS of the above be $\tilde{q}_t(i, j)$ and we show the Markov property in case of $t_0 = 1$ only. Let $u < s, t \leq n$. We first show

$$P(X_{t+s} = j \mid X_s = i, X_u = k, S_s = n', S_u = \ell) = P(X_{t+s} = j \mid X_s = i, S_s = n') = \tilde{q}_t(i, j).$$

The independence of $(Y_n)$, $(S_t)$, the Markov property of $(Y_n)$ and independent increments of $(S_t)$ imply

$$P(X_{t+s} = j \mid X_s = i, S_s = n', S_u = \ell) = \sum_{m \geq 0} P(X_{t+s} = j, S_{t+s} = n + m \mid X_s = i, S_s = n', S_u = \ell) \sum_{m \geq 0} P(Y_{n+m} = j, Y_u = k, S_s = n', S_u = \ell) P(S_{t+s} - S_s = m) P(S_s = n, S_u = \ell) \sum_{m \geq 0} P(Y_{n+m} = j, Y_u = k) P(S_{t+s} - S_s = m) P(S_s = n, S_u = \ell) = \sum_{m \geq 0} P(Y_{n+m} = j, Y_u = k) P(S_{t+s} - S_s = m) P(S_s = n, S_u = \ell) = \sum_{m \geq 0} P(Y_{n+m} = j) P(S_t = m) = \tilde{q}_t(i, j).$$

On the other hand, by a similar way we have

$$P(X_{t+s} = j \mid X_s = i, S_s = n) = \sum_{m \geq 0} P(X_{t+s} = j, S_{t+s} = n + m \mid X_s = i, S_s = n) = \sum_{m \geq 0} P(Y_{n+m} = j, Y_u = k) P(S_{t+s} - S_s = m) P(S_s = n) \sum_{m \geq 0} P(Y_{n+m} = j) P(S_t = m) = \tilde{q}_t(i, j).$$

Hence we have (4.5). The last term $\tilde{q}_t(i, j)$ is independent of $\ell \leq n, k \in S$, $u < s$, and events of the condition are mutually disjoint in $\ell \leq n$. Thus, by summing on $\ell \leq n$, we have the same result (see the next question). Therefore we have the time-homogeneous Markov property:

$$P(X_{t+s} = j \mid X_s = i, X_u = k) = P(X_{t+s} = j \mid X_s = i) = q_t(i, j).$$

Moreover, we have the transition probability

$$q_t(i, j) = P(X_t = j \mid X_0 = i) = \tilde{q}_t(i, j).$$
Question 4.5  Events \( \{B_n\} \) are mutually disjoint and there exist a event \( A \) and 0 ≤ q ≤ 1 such that \( P(A| B_k) = q \) \((^n \geq 1)\). Then show \( P(A| \bigcup B_n) = q \).

Proposition 4.5 (Chapman-Kolmogorov equation) \( q_{t+s}(i,j) = \sum_{k \in S} q_{t}(i,k) q_{s}(k,j) \).

Proof. \[
\text{[RHS]} = \sum_{k \in S} P(X_t = k | X_0 = i) P(X_{t+s} = j | X_t = k) \\
= \sum_{k \in S} P(X_t = k | X_0 = i) P(X_{t+s} = j | X_t = k, X_0 = i) \\
= \sum_{k \in S} \frac{P(X_{t+s} = j, X_t = k, X_0 = i)}{P(X_0 = i)} \\
= \frac{P(X_{t+s} = j, X_0 = i)}{P(X_0 = i)} = P(X_{t+s} = j | X_0 = i) = \text{[LHS]}
\]

\[\square\]

Proposition 4.6  Let \( Y_n \) be a discrete-time birth and death chain on \( \mathbb{Z}_+ \) with a birth rate \( \lambda_i \), a death rate \( \mu_i \) \((i \in \mathbb{Z}_+)\). The transition probability \( q_h(i,j) \) of a continuous-time birth and death chain \( X_t = Y_{Z_t} \) satisfies the following: (Note that \( \mu_0 = 0, \lambda_i > 0 \), and if \( i \geq 1 \), then \( \mu_i > 0 \).)

\[
q_h(i, i+1) = \lambda_i h + o(h) \\
q_h(i, i-1) = \mu_i h + o(h) \quad (i \geq 1) \\
q_h(i, i) = 1 - (\lambda_i + \mu_i) h + o(h) \\
q_h(i, j) = \delta_{ij}.
\]

In particular, \( \lim_{h \to 0} q_h(0, i) = 1 \). Note that \( q_h(0, -1) = 0 \), \( q_h(0, 0) = 1 \).

Proof. Let \( p_n(i,j) \) be an \( n \)-th transition probability of \( Y_n \). The transition probability \( q_h(i,j) \) satisfies the following as \( h \to 0 \):

\[
q_h(i,j) = \sum_{n \geq 0} e^{-h} \frac{h^n}{n!} p_n(i,j) \\
= e^{-h} \left( \delta_{ij} + hp(i,j) + O(h^2) \right) \\
= \delta_{ij} + hp(i,j) + O(h^2).
\]

Moreover, by noting that

\[
p(i, i+1) = \lambda_i, \quad p(i, i-1) = \mu_i, \quad p(i, i) = 1 - (\lambda_i + \mu_i)
\]

the result is easily obtained. \[\square\]

In general, let \( X_t \) be an \( S \)-valued time-homogeneous Markov chain. For a suitable function \( f : S \to \mathbb{R} \), let

\[
Gf(i) = \lim_{h \to 0} \frac{1}{h} \left( E^i[f(X_h)] - f(i) \right) = \lim_{h \to 0} \frac{1}{h} E^i[f(X_t) - f(X_0)],
\]

where \( E^i[\cdot] = E[\cdot | X_0 = i] \). Then \( G \) is called a generator of \( (X_t) \).

Theorem 4.4 In the above birth and death chain, for a bounded function \( f : \mathbb{Z}_+ \to \mathbb{R} \),

\[
Gf(i) = \lambda_i f(i+1) + \mu_i f(i-1) - (\lambda_i + \mu_i) f(i).
\]

Moreover,

\[
E^i[f(X_t) - f(X_0)] = \int_0^t E^i[Gf(X_s)] ds.
\]
Proof. For sufficiently small $h > 0$,

$$
E^i[f(X_h)] = f(i + 1)q_h(i, i + 1) + f(i - 1)q_h(i, i - 1) + f(i)q_h(i, i) + o(h)
$$

$$
= f(i) + h[\lambda_i f(i + 1) + \mu_i f(i - 1) - (\lambda_i + \mu_i) f(i)] + o(h)
$$

Hence $Gf(i)$ is obtained. Moreover, by Markov property,

$$
E^i[f(X_t) - f(X_0)] = \int_0^t \lim_{h \to 0} \frac{1}{h} E^i[f(X_{s+h}) - f(X_s)] ds
$$

$$
= \int_0^t \lim_{h \to 0} \frac{1}{h} E^i \left[ E^{X_s}[f(X_h)] - f(X_0) \right] ds
$$

$$
= \int_0^t E^i \left[ \lim_{h \to 0} \frac{1}{h} E^{X_s}[f(X_h)] - f(X_0) \right] ds
$$

$$
= \int_0^t E^i [Gf(X_s)] ds.
$$

Note that it is possible to exchange $\lim_{h \to 0}$ and $E^i$ in the above, because by the boundedness of $f$ and $0 < \lambda_i, \mu_i < 1$ we can apply Lebesgue’s convergence theorem.

The above result means $f(i)$ changes to $f(i + 1)$ at rate $\lambda_i$, to $f(i - 1)$ at rate $\mu_i$ and does not change at rate $1 - \lambda_i - \mu_i$. Therefore, if the generator $G$ is known, then the Markov process $(X_t)$ is known. That is, $G$ and $(X_t)$ is one-to-one onto.

On a more general state space $S$ the generator is a very important tool in the theory of Markov processes.

References