Probability Theory & Mathematical Statistics
Random Processes
Random Processes

(Stochastic processes)

- A **random process** is a family \( \{ X_t : t \in \mathcal{T} \} \) of random variables

- \( \mathcal{T} \) is a set of indices, where each \( t \in \mathcal{T} \) is usually interpreted as **time**
  - **discrete-time**: countable \( \mathcal{T} \), usually \( \mathcal{T} = \{0,1,2,...\} \) or \( \mathcal{T} = \{1,2,...\} \)
  - **continuous-time**: uncountable \( \mathcal{T} \), usually \( \mathcal{T} = [0,\infty) \)

- \( X_t \) takes values in a **state space** \( \mathcal{S} \)
  - **discrete-space**: countable \( \mathcal{S} \), e.g. \( \mathcal{S} = \mathbb{Z} \)
  - **continuous-space**: uncountable \( \mathcal{S} \), e.g. \( \mathcal{S} = \mathbb{R} \)
Random Processes

*(Stochastic processes)*

- **Bernoulli process**: i.i.d. Bernoulli trials $X_0, X_1, X_2, \ldots \in \{0,1\}$

- **Branching (Galton-Watson) process**: $X_0 = 1$ and $X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)}$

  where $\{\xi_j^{(n)} : n,j \geq 0\}$ are i.i.d. non-negative integer-valued random variables

- **Poisson process**: continuous-time counting process $\{N(t) \mid t \geq 0\}$ such that

  $N(t) = \max \{n \mid X_1 + \cdots + X_n \leq t\}$ for any $t \geq 0$

  where $\{X_i\}$ are i.i.d. exponential random variables with parameter $\lambda > 0$
Martingales
Martingale (鞅)

• A sequence \( \{Y_n : n \geq 0\} \) of random variables is a **martingale** with respect to another sequence \( \{X_n : n \geq 0\} \) if, for all \( n \geq 0 \),
  
  \[ \mathbb{E} \left[ |Y_n| \right] < \infty \]

  \[ \mathbb{E} \left[ Y_{n+1} \mid X_0, X_1, \ldots, X_n \right] = Y_n \quad \text{(martingale property)} \]

• By definition: \( Y_n \) is a function of \( X_0, X_1, \ldots, X_n \)

• Current capital \( Y_n \) in a **fair gambling game** with outcomes \( X_0, X_1, \ldots, X_n \)
  
  • **Super-martingale** (上鞅): \( \mathbb{E} \left[ Y_{n+1} \mid X_0, X_1, \ldots, X_n \right] \leq Y_n \)
  
  • **Sub-martingale** (下鞅): \( \mathbb{E} \left[ Y_{n+1} \mid X_0, X_1, \ldots, X_n \right] \geq Y_n \)
Martingale (鞅)

- A sequence \( \{ Y_n : n \geq 0 \} \) of random variables is a **martingale** with respect to another sequence \( \{ X_n : n \geq 0 \} \) if, for all \( n \geq 0 \),
  - \( \mathbb{E} [ | Y_n | ] < \infty \)
  - \( \mathbb{E} \left[ Y_{n+1} \mid X_0, X_1, \ldots, X_n \right] = Y_n \quad \text{(martingale property)} \)
- \( \{ X_n : n \geq 0 \} \) are defined on the probability space \((\Omega, \Sigma, \text{Pr})\)
  - \( (X_0, X_1, \ldots, X_n) \) defines a sub-\( \sigma \)-field \( \Sigma_n \subseteq \Sigma \) (the smallest \( \sigma \)-field s.t. \( (X_0, \ldots, X_n) \) is \( \Sigma_n \)-measurable)
  - \( \{ \Sigma_n : n \geq 0 \} \) is a filtration of \( \Sigma \), i.e. \( \Sigma_0 \subseteq \Sigma_1 \subseteq \cdots \subseteq \Sigma \)
  - The martingale property is expressed as \( \mathbb{E} \left[ Y_{n+1} \mid \Sigma_n \right] = Y_n \)
Examples of Martingale

- Doob martingale: \( Y_i = \mathbb{E} \left[ f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right] \)
  - vertex/edge exposure martingale for random graph
- Capital in a fair gambling game (arbitrary betting strategy)
- Unbiased 1D random walk: \( Y_n = \sum_{i=1}^{n} X_i \) with i.i.d. uniform \( X_i \in \{ -1, 1 \} \)
- de Moivre’s martingale: \( Y_n = \left( \frac{p}{1 - p} \right)^{X_n} \), where \( X_n = \sum_{i=1}^{n} X_i \) and \( X_i \in \{ -1, 1 \} \) are independent with \( \Pr(X_i = 1) = p \)
- Polya’s urn: The urn contains marbles with different colors. At each turn, a marble is selected u.a.r., and replaced with \( k \) marbles of that same color.
Studies of Martingale

• For martingale \( \{Y_n : n \geq 0\} \) with respect to \( \{X_n : n \geq 0\} \):

\[
\mathbb{E} \left[ Y_{n+1} \mid X_0, X_1, \ldots, X_n \right] = Y_n
\]

• Concentration of measure (tail inequality): under what condition

\[
\Pr \left( |Y_n - Y_0| \geq t \right) \leq ?
\]

• Optional stopping theorem (OST): under what condition for a stopping time \( \tau \)

\[
\mathbb{E}[Y_\tau] = \mathbb{E}[Y_0]
\]
Fair Gambling Game

- If $\{Y_n : n \geq 0\}$ is a martingale with respect to $\{X_n : n \geq 0\}$, then $\forall n \geq 0$,

$$\mathbb{E}[Y_n] = \mathbb{E}[Y_0]$$

Proof: By total expectation $\mathbb{E}[Y_n] = \mathbb{E}\left[\mathbb{E}\left[ Y_n \mid X_0, X_1, \ldots, X_{n-1} \right] \right]$

As a martingale, $\mathbb{E}\left[ Y_n \mid X_0, X_1, \ldots, X_{n-1} \right] = Y_{n-1}$

$$\Rightarrow \mathbb{E}[Y_n] = \mathbb{E}\left[\mathbb{E}\left[ Y_n \mid X_0, X_1, \ldots, X_{n-1} \right] \right] = \mathbb{E}[Y_{n-1}]$$
Stopping Time

- A nonnegative integer-valued random variable $T$ is a **stopping time** with respect to the sequence $\{X_t : t = 0,1,2,\ldots\}$ if for any $n \geq 0$ the occurrence of the event $T = n$ is determined by the evaluation of $X_0, X_1, \ldots, X_n$

- Formally, $\{X_t : t = 0,1,2,\ldots\}$ defines a filtration of $\sigma$-fields $\Sigma_0 \subseteq \Sigma_1 \subseteq \ldots$ such that $(X_0, X_1, \ldots, X_n)$ is $\Sigma_n$-measurable (and $\Sigma_n$ is the smallest such $\sigma$-field). Then $T$ is a stopping time if $\{T = n\} \in \Sigma_n$ for any $n \geq 0$.

- Intuitively, $T$ is a stopping time, if whether stopping at time $n$ is determined by the outcomes of $X_0, X_1, \ldots, X_n$
Optional Stopping Theorem (OST)
(Martingale Stopping Theorem)

- Optional Stopping Theorem (OST): Let \( \{ Y_t : t \geq 0 \} \) be a martingale and \( T \) be a stopping time, both with respect to \( \{ X_t : t \geq 0 \} \). Then

\[
\mathbb{E} [ Y_T ] = \mathbb{E} [ Y_0 ]
\]

if any one of the following conditions holds:

- (bounded time) there is a finite \( n \) such that \( T < n \) a.s.
- (bounded range) \( T < \infty \) a.s. and there is a finite \( c \) s.t. \( |Y_t| < c \) for all \( t \leq T \)
- (bounded differences) \( \mathbb{E} [ T ] < \infty \) and there is a finite \( c \) such that

\[
\mathbb{E} [ | Y_{t+1} - Y_t | \mid X_0, X_1, \ldots, X_t ] < c \text{ for all } t \geq 0
\]
Optional Stopping Theorem (OST) (Martingale Stopping Theorem)

- **Optional Stopping Theorem (OST):** Let \( \{Y_t : t \geq 0\} \) be a martingale and \( T \) be a stopping time, both with respect to \( \{X_t : t \geq 0\} \). Then

\[
\mathbb{E}[Y_T] = \mathbb{E}[Y_0]
\]

(general condition) if all the following conditions hold:

- \( \Pr(T < \infty) = 1 \)
- \( \mathbb{E}[|Y_T|] < \infty \)
- \( \lim_{n \to \infty} \mathbb{E}[Y_n \cdot I[T > n]] = 0 \)

- The proof of this general OST utilizes *Doob’s optional sampling* argument
Gambler’s Ruin
(Symmetric Random Walk in One-Dimension)

Let $Y_t = \sum_{i=1}^{t} X_i$ where $X_i \in \{-1, +1\}$ are i.i.d. uniform (Rademacher) RVs

Let $T$ be the first time $t$ that $Y_t = -a$ or $Y_t = b$

$\{Y_t : t \geq 0\}$ is a martingale and $T$ is a stopping time (both w.r.t. $\{X_i : i \geq 1\}$)

satisfying that $|Y_t| \leq \max\{a, b\}$ for all $0 \leq t \leq T$ and $T < \infty$ a.s.

(OST) $\implies \mathbb{E}[Y_T] = \mathbb{E}[Y_0] = 0$

$\mathbb{E}[Y_T] = b \cdot \Pr(Y_T = b) - a \cdot \Pr(Y_T \neq b) \implies \Pr(Y_T = b) = \frac{a}{a + b}$
Wald’s Equation
(Linearity of expectation with randomly many random variables)

- **Wald’s equation**: Let $X_1, X_2, \ldots$ be i.i.d. non-negative with $\mu = \mathbb{E}[X_i] < \infty$. Let $T$ if a stopping time with respect to $X_1, X_2, \ldots$. If $\mathbb{E}[T] < \infty$, then

$$\mathbb{E} \left[ \sum_{i=1}^{T} X_i \right] = \mathbb{E}[T] \cdot \mu$$

- **Proof**: For $t \geq 1$, let $Y_t = \sum_{i=1}^{t} (X_i - \mu)$, which is a martingale. Observe that:

$$\mathbb{E}[T] < \infty \quad \text{and} \quad \mathbb{E}[|Y_{t+1} - Y_t| | X_1, \ldots, X_t] = \mathbb{E}[|X_{t+1} - \mu|] \leq 2\mu$$

By OST: $\mathbb{E}[Y_T] = \mathbb{E}[Y_1] = 0$. Note that $\mathbb{E}[Y_T] = \mathbb{E}\left[\sum_{i=1}^{T} X_i\right] - \mathbb{E}[T] \cdot \mu$
Markov Chain
Markov Chain (马尔可夫链)

• A discrete-time random process $X_0, X_1, X_2, \ldots$ is a Markov chain if

\[ \Pr(X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0) = \Pr(X_{t+1} = x_{t+1} \mid X_t = x_t) \]

• The Markov property (memoryless property):

  • The next state $X_{t+1}$ depends on the current state $X_t$ but is independent of the history $X_0, X_1, \ldots, X_{t-1}$ of how the process arrived at state $X_t$

  • $X_{t+1}$ is conditionally independent of $X_0, X_1, \ldots, X_{t-1}$ given $X_t$

\[ X_0 \rightarrow X_1 \rightarrow \cdots \rightarrow X_{t-1} \rightarrow X_t \rightarrow X_{t+1} \]
Transition Matrix (转移矩阵)

• A discrete-time random process \(X_0, X_1, X_2, \ldots\) is a Markov chain if

\[
\Pr(X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0) = \Pr(X_{t+1} = x_{t+1} \mid X_t = x_t)
\]

(time-homogeneous)

\[
P = P(x_t, x_{t+1}) = P^{(t)}(x_t, x_{t+1})
\]

• \(P\) is called the transition matrix: (assuming discrete-space)

\[
P(x, y) = \Pr(X_{t+1} = y \mid X_t = x)\text{ for any } x, y \in \mathcal{S}
\]

where \(\mathcal{S}\) is the discrete state space on which \(X_0, X_1, X_2, \ldots\) take values.

• \(P\) is a (row-)stochastic matrix: \(P \geq 0\) and \(P1 = 1\)
Transition Matrix (转移矩阵)

- For a Markov chain $X_0, X_1, X_2, \ldots$ with discrete state space $\mathcal{S}$

  \[
  \Pr(X_{t+1} = y \mid X_t = x) = P(x, y)
  \]

  where $P \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}_{\geq 0}$ is the transition matrix, which is a (row-)stochastic matrix

- Let $\pi^{(t)}(x) = \Pr(X_t = x)$ be the mass function (pmf) of $X_t$. By total probability:

  \[
  \pi^{(t+1)}(y) = \Pr(X_{t+1} = y) = \sum_{x \in \mathcal{S}} \Pr(X_{t+1} = y \mid X_t = x) \Pr(X_t = x) = \pi^{(t)} P
  \]

  $\pi^{(0)} \xrightarrow{P} \pi^{(1)} \xrightarrow{P} \cdots \xrightarrow{P} \pi^{(t)} \xrightarrow{P} \pi^{(t+1)} \xrightarrow{P} \cdots$
Random Walk (随机游走)

• WLOG: a Markov chain is a **random walk** on state space $\mathcal{S}$

• Each state $x \in \mathcal{S}$ corresponds to a vertex

• Given the current state $x \in \mathcal{S}$, the probability of next state being $y \in \mathcal{S}$ is:

$$P(x, y) = \Pr(X_{t+1} = y \mid X_t = x)$$

• Initially, $\pi^{(0)}(x) = \Pr(X_0 = x)$, for $t \geq 0$:

$$\pi^{(t+1)} = \pi^{(t)}P$$
Stationary Distribution (稳态分布)

• A distribution (pmf) \( \pi \) on state space \( \mathcal{S} \) is called a stationary distribution of the Markov chain \( P \) if

\[
\pi P = \pi
\]

• \( \pi \) is a fixpoint (equilibrium) of the linear dynamic system

\[
P = \begin{bmatrix}
0 & 1 & 0 \\
1/3 & 0 & 2/3 \\
1/3 & 1/3 & 1/3
\end{bmatrix}
\]

\[
\pi = \left( \frac{1}{4}, \frac{3}{8}, \frac{3}{8} \right)
\]

\[
P^{20} \approx \begin{bmatrix}
0.2500 & 0.3750 & 0.3750 \\
0.2500 & 0.3750 & 0.3750 \\
0.2500 & 0.3750 & 0.3750
\end{bmatrix}
\]
Convergence Theorem

- **Markov chain convergence theorem:**
  
  If a Markov chain $X_0, X_1, X_2 \ldots$ on state space $\mathcal{S}$ is *irreducible* and *ergodic*, then there is a unique stationary distribution $\pi$ on $\mathcal{S}$ such that
  
  $$
  \pi(x) = \lim_{t \to \infty} \Pr(X_t = x \mid X_0 = x_0) \text{ for any } x_0 \in \mathcal{S}
  $$

- **Irreducibility:** the chain is *irreducible* if $P$ is an irreducible matrix (不可约矩阵)
  
  $$
  \iff \text{ the state space } \mathcal{S} \text{ is strongly connected under } P
  $$

- **Ergodicity:** the chain is *ergodic* if all states are *aperiodic* (无周期)
  
  and *positive recurrent* (正常返)
**Ergodicity**

- Let $X_0, X_1, X_2\ldots$ be a Markov chain on state space $\mathcal{S}$ with transition matrix $P$.
  
- The **period** $d(x)$ of a state $x \in \mathcal{S}$ is $d(x) = \gcd\{t \geq 1 \mid P^t(x, x) > 0\}$
  
- A state $x \in \mathcal{S}$ is called **aperiodic** if $d(x) = 1$
  
- $P(x, x) > 0 \implies x$ is aperiodic
  
- A state $x \in \mathcal{S}$ is called **recurrent** if $\Pr(\exists t \geq 1, X_t = x \mid X_0 = x) = 1$
  
  and further called **positive recurrent** if $\mathbb{E}[\min\{t \geq 1 : X_t = x\} \mid X_0 = x] < \infty$
  
- **Shizuo Kakutani** (角谷静夫): random walk is recurrent on $\mathbb{Z}^2$ but non-recurrent on $\mathbb{Z}^3$
  
  “A drunk man will find his way home, but a drunk bird may get lost forever.”
  
- On finite state space $\mathcal{S}$: irreducible $\implies$ all states are positive recurrent
Convergence Theorem

• Markov chain convergence theorem:

If a Markov chain $X_0, X_1, X_2 \ldots$ on state space $\mathcal{S}$ is **irreducible** and **ergodic**, then there is a unique stationary distribution $\pi$ on $\mathcal{S}$ such that

$$
\pi(x) = \lim_{t \to \infty} \Pr(X_t = x \mid X_0 = x_0) \text{ for any } x_0 \in \mathcal{S}
$$

• Finite Markov chain (with finite state space $\mathcal{S}$):

lazy (i.e. $P(x, x) > 0$) and **strongly connected** $P$

$\implies$ always converge to the unique stationary distribution $\pi = \pi P$
Convergence Theorem

- **Markov chain convergence theorem:**

  If a Markov chain $X_0, X_1, X_2 \ldots$ on state space $\mathcal{S}$ is **irreducible** and **ergodic**, then there is a unique stationary distribution $\pi$ on $\mathcal{S}$ such that

  \[
  \pi(x) = \lim_{t \to \infty} \Pr(X_t = x \mid X_0 = x_0) \text{ for any } x_0 \in \mathcal{S}
  \]

- **Proof:** (By coupling)

  \[
  x_0 = X_0 \to X_1 \to \ldots \to X_n \parallel \begin{array}{c}
  Y_0 \to Y_1 \to \ldots \to Y_n \\
  \parallel
  \end{array} \downarrow
  \]

  $\pi \sim Y_0 \to Y_1 \to \ldots \to Y_t \sim \pi$

  **irreducibility + ergodicity $\implies$ occurs a.s.**
PageRank

- Each webpage \( x \in S \) is assigned a rank \( r(x) \):
  - A page has high rank if pointed by many high-rank pages.
  - High-rank pages have greater influence.
  - Pages pointing to few others have greater influence.

- Linear system: \( r(x) = \sum_{y \rightarrow x} \frac{r(y)}{d^+(y)} \) where \( d^+(y) \) is the out-degree of page \( y \)

- Stationary distribution \( rP = r \) for the random walk (tireless internet surfer)

\[
P(x, y) = \begin{cases} 
\frac{1}{d^+(x)} & \text{if } x \rightarrow y \\
0 & \text{o.w.}
\end{cases}
\]
Mixing of Markov Chain

• **Markov chain convergence theorem:**
  
  If a Markov chain $X_0, X_1, X_2 \ldots$ on state space $\mathcal{S}$ is *irreducible* and *ergodic*, then there is a unique stationary distribution $\pi$ on $\mathcal{S}$ such that
  
  $$
  \pi(x) = \lim_{t \to \infty} \Pr(X_t = x \mid X_0 = x_0) \text{ for any } x_0 \in \mathcal{S}
  $$

  • How fast is the convergence rate?

  • **Mixing time:** let $\pi^{(t)}_x(y) = (1_x P^t)_y = \Pr(X_t = y \mid X_0 = x)$

    $$
    \tau(\epsilon) = \max_{x \in \mathcal{S}} \min_{t \geq 1} \left\{ \frac{1}{1} \left\| \pi^{(t)}_x - \pi \right\|_1 \leq 2\epsilon \right\}
    $$
Random Processes
Random Processes

- Stationary processes: \((X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \sim (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h})\)
  - i.i.d. variables, stationary Markov chains, stationary Gaussian process, ...

- Renewal (or counting) processes: \(N(t) = \max \{n \mid X_1 + \cdots + X_n \leq t\}\) where \(\{X_i : i \geq 1\}\) are i.i.d. nonnegative-valued random variables
  - Poisson processes (the only renewal processes that are Markov chains)

- Wiener process (Brownian motion): continuous-time continuous-space \(\{W(t) \in \mathbb{R} : t \geq 0\}\) with time-homogeneity and independent increments
  
  \(W(s_i) - W(t_i)\) are independent whenever the intervals \((s_i, t_i]\) are disjoint
Diffusion Processes
(Stochastic processes with continuous sample paths)

• Let \((\Omega, \Sigma, \Pr)\) be a probability space. A random process \(X : \mathcal{T} \times \Omega \to \mathcal{S}\) with time range \(\mathcal{T}\) and state space \(\mathcal{S}\) is called a diffusion process if there is an \(A \in \Sigma\) with \(\Pr(A) = 1\) such that for all \(\omega \in A\),

\[
X(\omega) : \mathcal{T} \to \mathcal{S}
\]

is a continuous function (between topological spaces).

• The Wiener processes are one-dimensional diffusions.

• Itô (伊藤) calculus may apply!