Foundations of Data Science

Random Processes

尹一通、刘明谋 Nanjing University, 2024 Fall

Doob Sequence

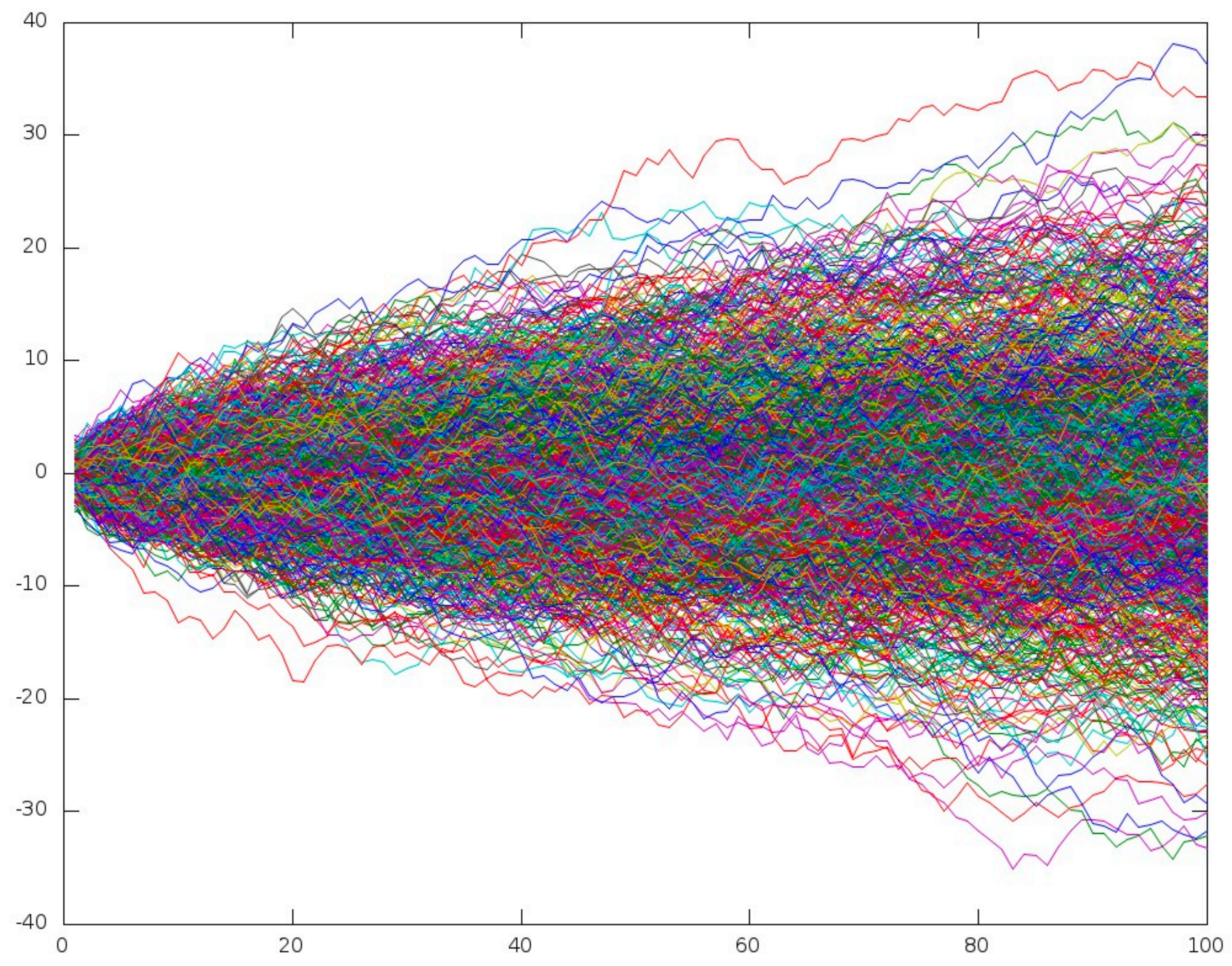
- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

$$Y_0 = \mathbb{E}\left[f(X_1, \ldots, X_n)\right] \dashrightarrow f(X_1, \ldots, X_n) = Y_n$$

no information

full information

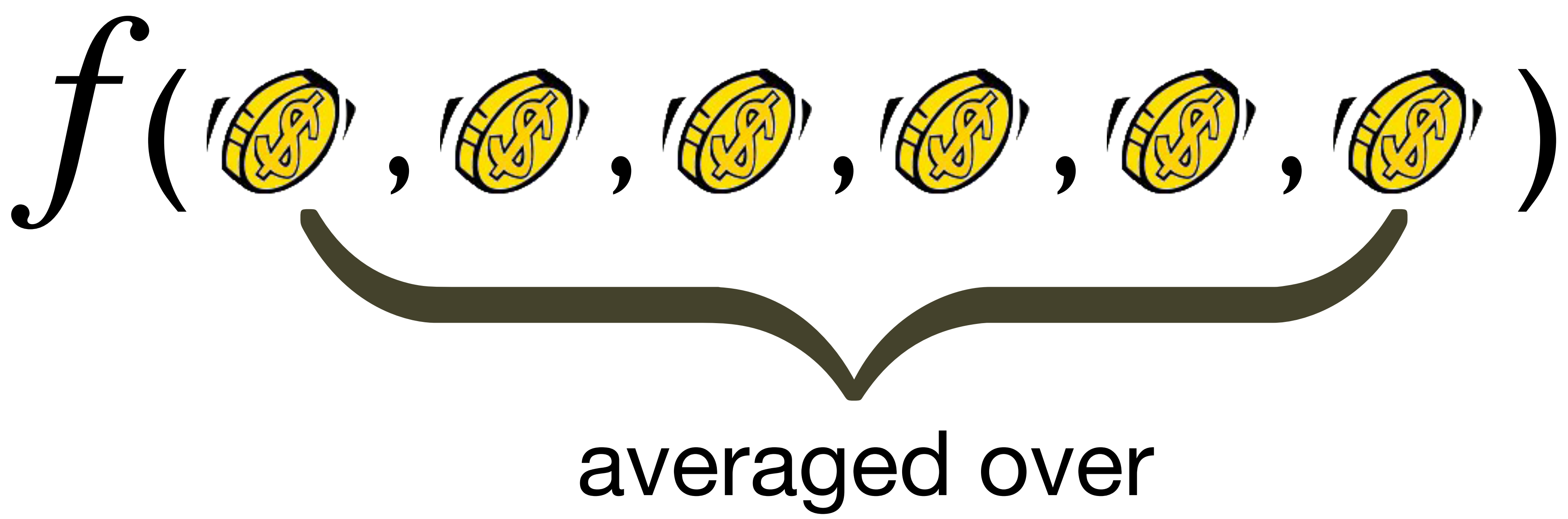


$$\Pr\left(\left|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]\right| < t\right)$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \leq i \leq n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

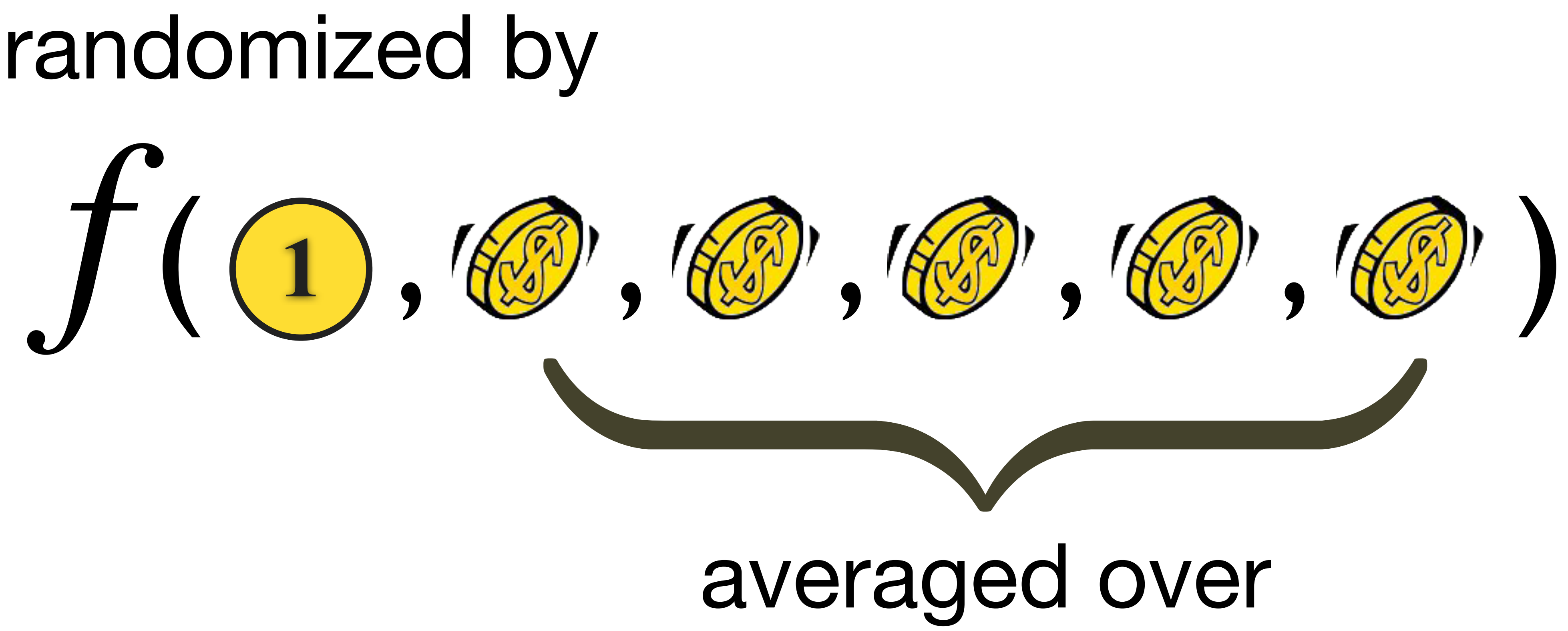


$$\mathbb{E}[f] = Y_0$$

Doob Sequence

- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \leq i \leq n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$

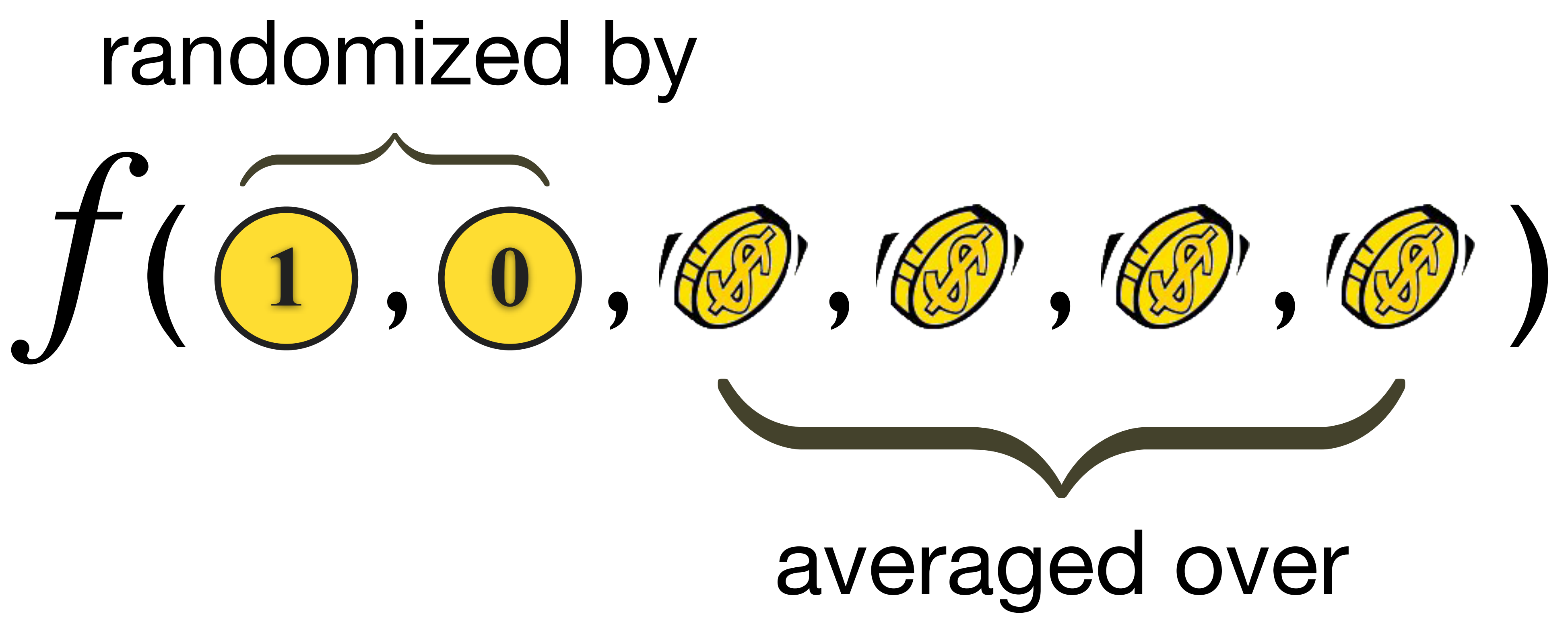


$$\mathbb{E}[f] = Y_0 \to Y_1$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \leq i \leq n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$

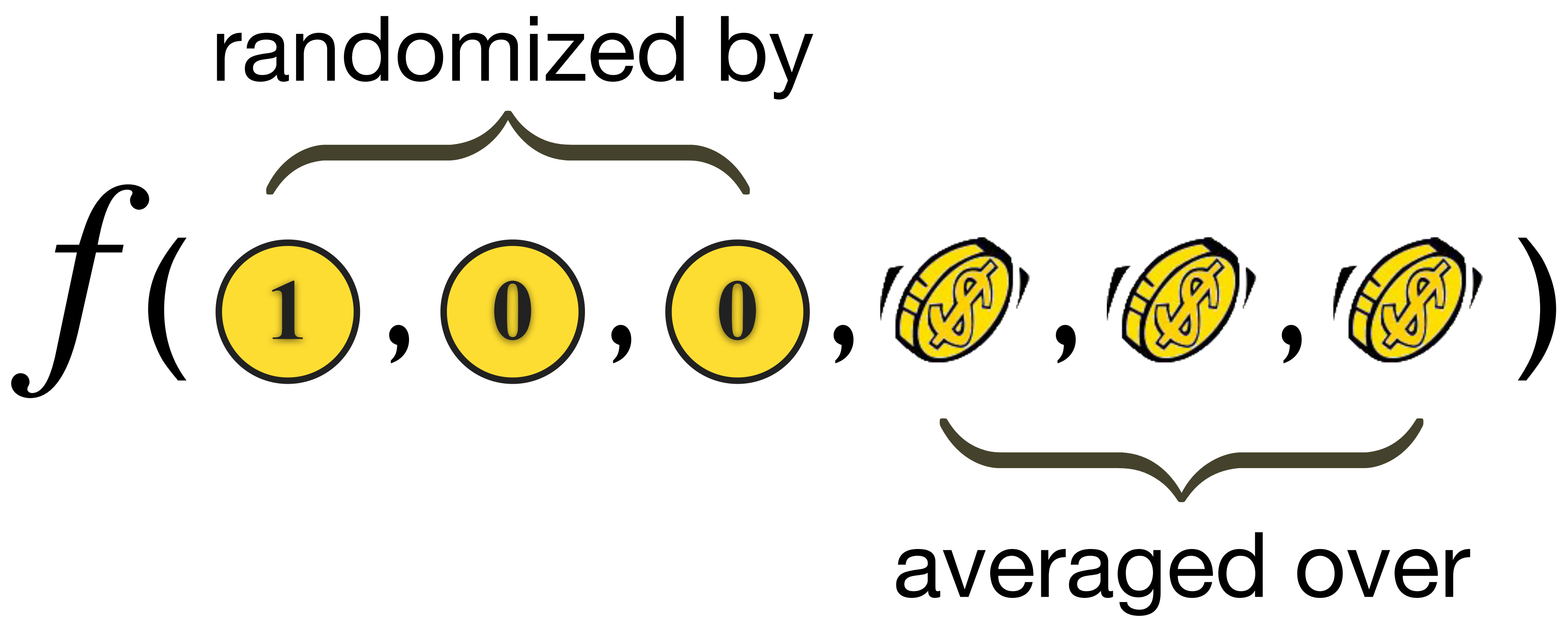


$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \leq i \leq n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i\right]$$

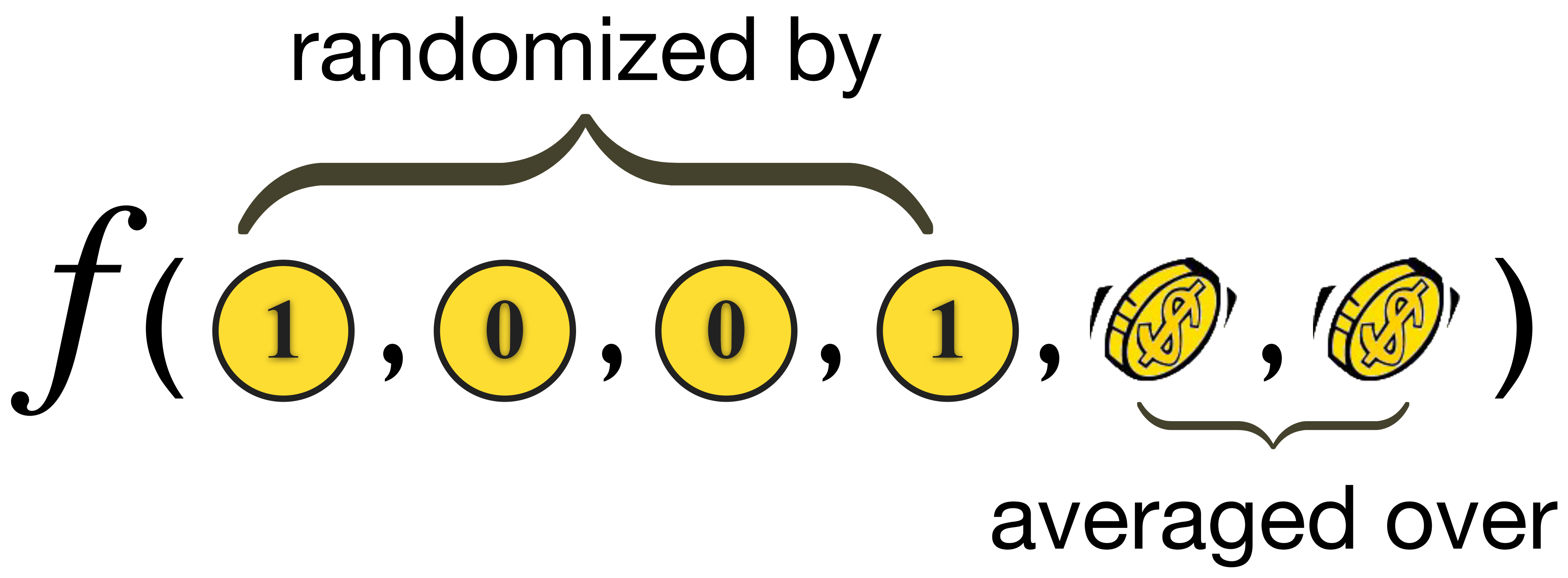


$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3$$

Doob Sequence

- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \leq i \leq n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$

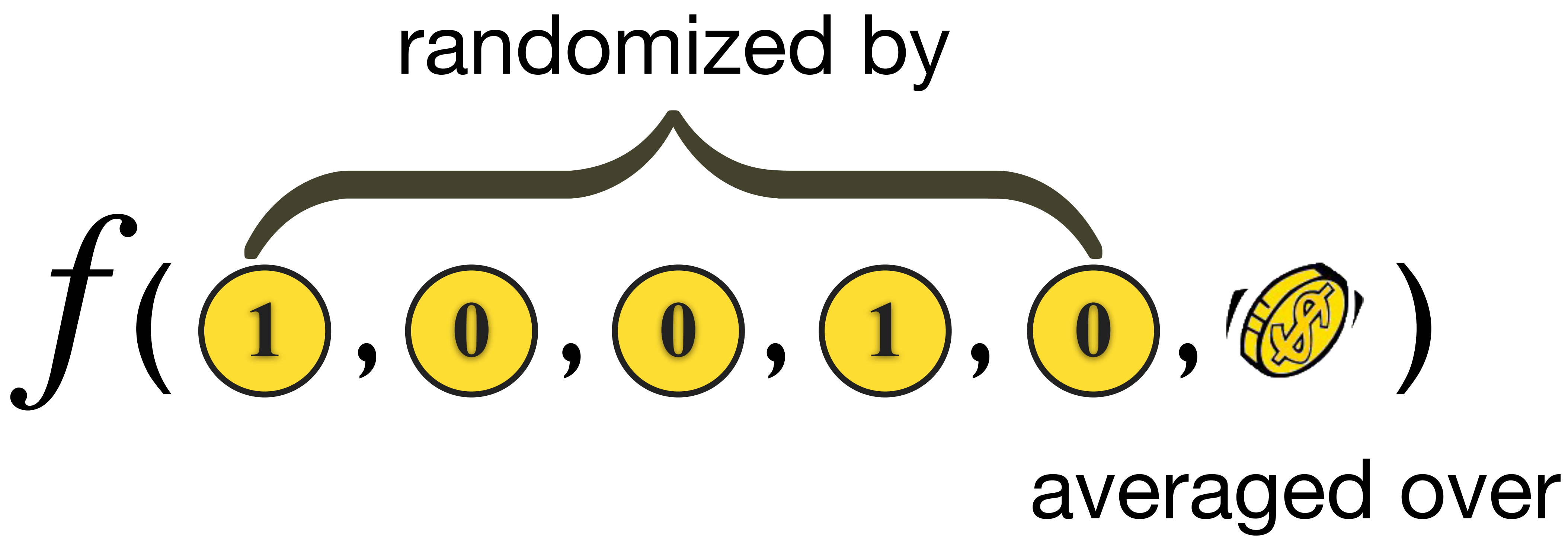


$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3 \to Y_4$$

Doob Sequence

- The **Doob sequence** $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$

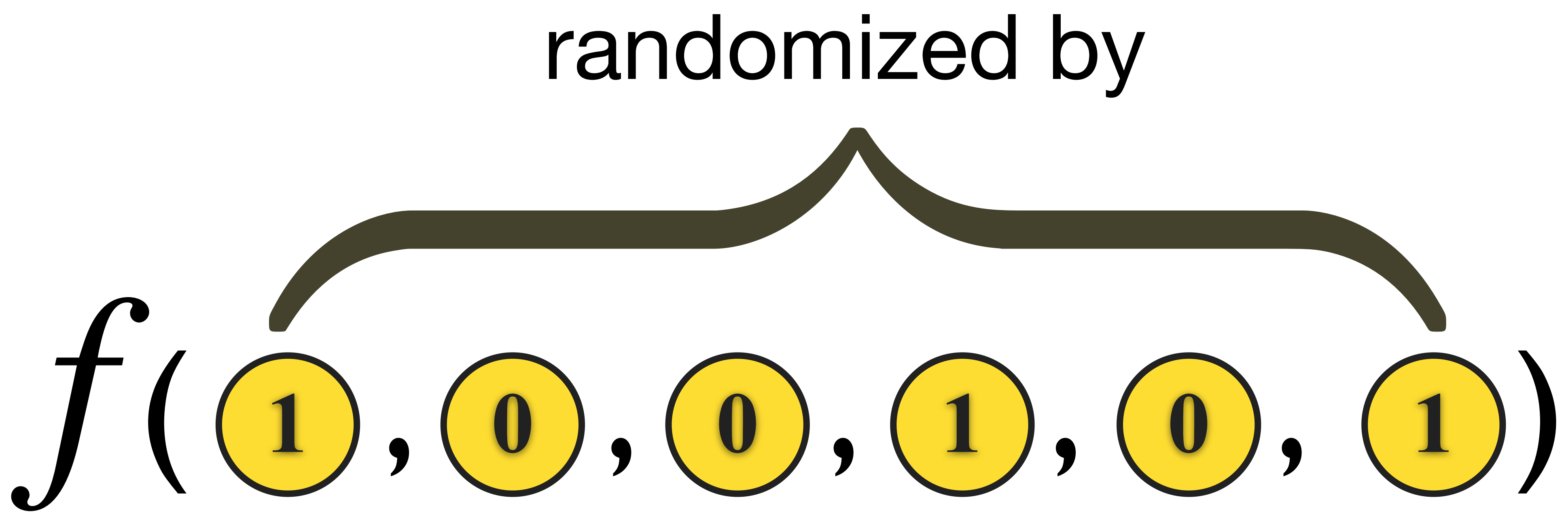


$$\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3 \to Y_4 \to Y_5$$

Doob Sequence

- The <u>**Doob sequence**</u> $Y_0, Y_1, \ldots, Y_n$ of n-variate function $f : \mathbb{R}^n \to \mathbb{R}$ on random variables $X_1, \ldots, X_n$, is given by

$$\forall 0 \le i \le n: \qquad Y_i = \mathbb{E}\left[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_i \right]$$



no information $\mathbb{E}[f] = Y_0 \to Y_1 \to Y_2 \to Y_3 \to Y_4 \to Y_5 \to Y_6 = f$ full information

Poisson Point Process

(Stochastic counting process with exponential interarrival)

- The **Poisson process** $\{N(t) \mid t \geq 0\}$ with rate $\lambda > 0$ is a continuous time process defined as follows —— imagine we have such a clock:
  - $N(t)$ counts the number of times the clock rings up to time t, initially $N(0) = 0$;
  - The time elapse (**interarrival time**) between any two consecutive ringings (including the time elapse before 1st ringing) is independent exponential with parameter λ
- Due to **memoryless** and **minimum:** The process defined by k independent clocks with the same rate λ is equivalent to the 1-clock process with rate $k\lambda$
- (Poisson distribution) For any $t, s \geq 0$ and any integer $n \geq 0$,

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

Random Processes
(*Stochastic* processes)

- A **random process** is a family $\{X_t : t \in \mathscr{T}\}$ of random variables
- $\mathscr{T}$ is a set of indices, where each $t \in \mathscr{T}$ is usually interpreted as **time**
  - **discrete-time**: countable $\mathscr{T}$, usually $\mathscr{T} = \{0,1,2,\ldots\}$ or $\mathscr{T} = \{1,2,\ldots\}$
  - **continuous-time**: uncountable $\mathscr{T}$, usually $\mathscr{T} = [0,\infty)$
- X_t takes values in a **state space** $\mathscr{S}$
  - **discrete-space**: countable $\mathscr{S}$, e.g. $\mathscr{S} = \mathbb{Z}$
  - **continuous-space**: uncountable $\mathscr{S}$, e.g. $\mathscr{S} = \mathbb{R}$

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

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

  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 <u>martingale</u> with respect to another sequence $\{X_n : n \geq 0\}$ if, for all $n \geq 0$,
  - $\mathbb{E}\left[|Y_n|\right] < \infty$
  - $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] = Y_n$ (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 $X_0, X_1, \ldots, X_n$
  - <u>Super-martingale</u> (上鞅): $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] \leq Y_n$
  - <u>Sub-martingale</u> (下鞅): $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] \geq Y_n$

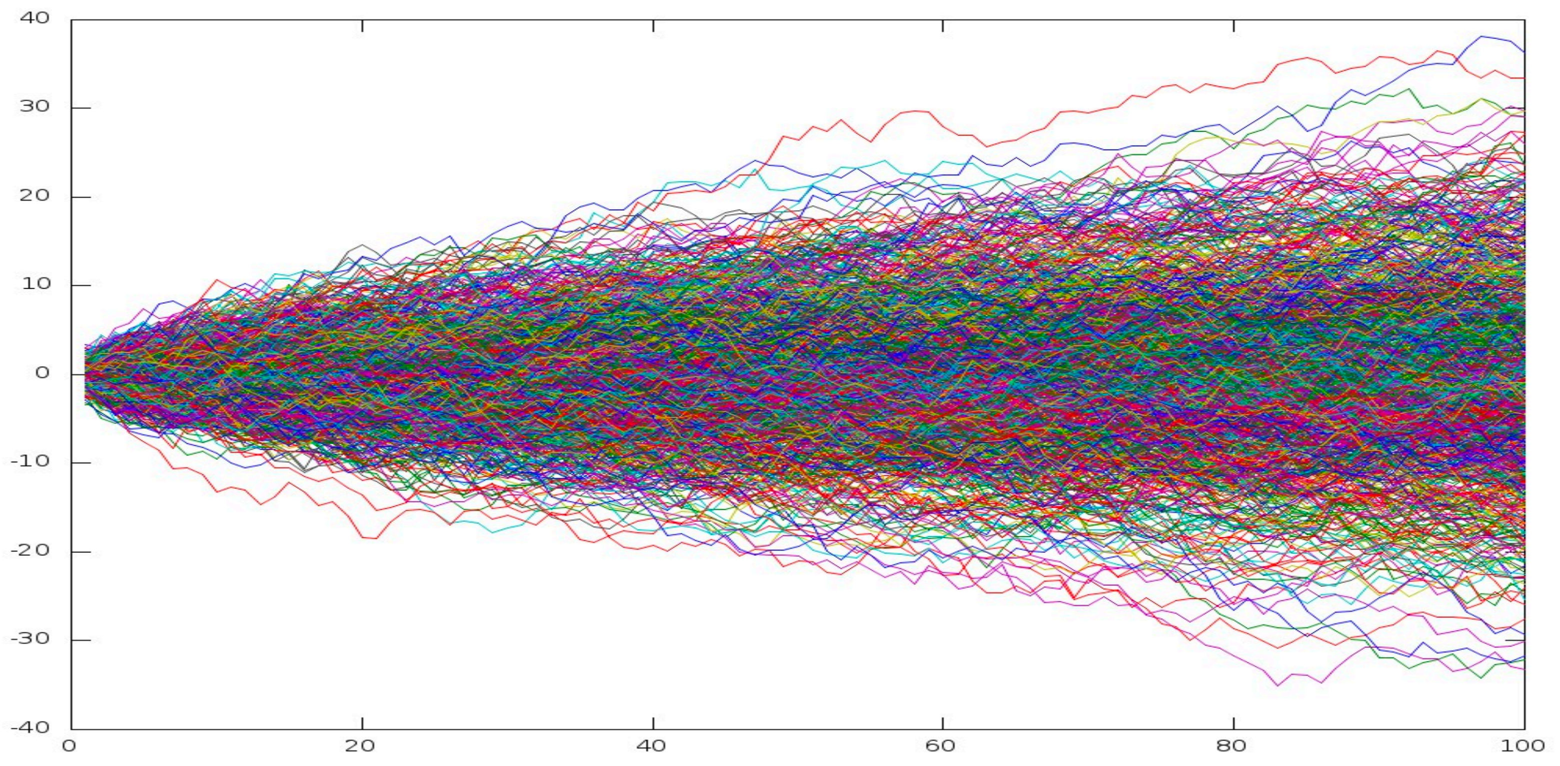
Martingale (鞅)

- A sequence $\{Y_n : n \geq 0\}$ of random variables is a <u>martingale</u> with respect to another sequence $\{X_n : n \geq 0\}$ if, for all $n \geq 0$,
  - $\mathbb{E}\left[|Y_n|\right] < \infty$
  - $\mathbb{E}\left[Y_{n+1} \mid X_0, X_1, \ldots, X_n\right] = Y_n$ (martingale property)
- $\{X_n : n \geq 0\}$ are defined on the probability space $(\Omega, \Sigma, \Pr)$
  - $(X_0, X_1, \ldots, X_n)$ defines a sub-σ-field $\Sigma_n \subseteq \Sigma$ (the smallest σ-field s.t. $(X_0, \ldots, X_n)$ is Σ_n-measurable)
  - $\{\Sigma_n : n \geq 0\}$ is a <u>**filtration**</u> of Σ, 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:** $Y_n = (p/(1-p))^{X_n}$, 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 $\{X_n : n \geq 0\}$:

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

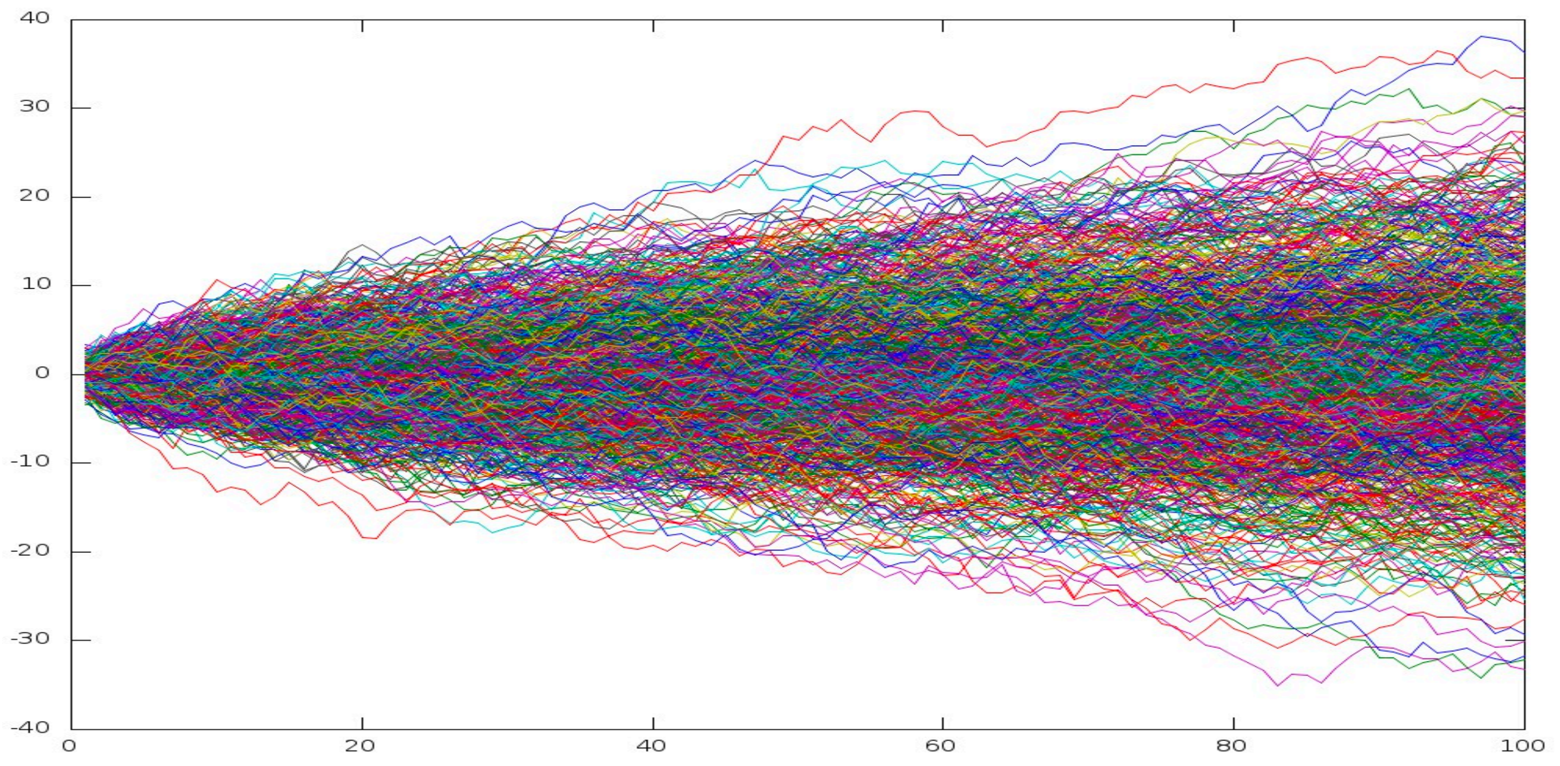
- **Concentration of measure** (tail inequality): Azuma's inequality

$$\Pr\left(\left|Y_n - Y_0\right| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

- **Optional stopping theorem** (**OST**): good quitting strategy (i.e. **stopping time** τ)

$$\mathbb{E}[Y_\tau] > \mathbb{E}[Y_0]\ ?$$

Fair Gambling Game



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

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

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

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

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

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 σ-fields $\Sigma_0 \subseteq \Sigma_1 \subseteq \cdots$ such that $(X_0, X_1, \ldots, X_n)$ is Σ_n-measurable (and Σ_n is the smallest such σ-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 $X_0, X_1, \ldots, X_n$

Stopped Martingale

- Consider a martingale $\{Y_n : n \geq 0\}$ and a stopping time T, both with respect to $\{X_n : n \geq 0\}$. The *stopped martingale* $\{Y_n^T : n \geq 0\}$ is defined as

$$Y_n^T \triangleq \begin{cases} Y_n & \text{if } n \leq T \\ Y_T & \text{if } n > T \end{cases}$$

- Stopped martingales are martingale.

Proof: Note event $T \geq i$ is determined by evaluation of $X_0, \ldots, X_{i-1}$ only. Also note $Y_i^T = Y_{i-1}^T + \mathbf{1}_{T\geq i} \cdot (Y_i - Y_{i-1})$. Let's calculate $\mathbb{E}\left[Y_{i+1}^T \,|\, X_0 \ldots X_i\right]$:

$$\mathbb{E}\left[Y_i^T + \mathbf{1}_{T>i} \cdot (Y_{i+1} - Y_i) \,|\, X_0 \ldots X_i\right] = \mathbb{E}\left[Y_i^T \,|\, X_0 \ldots X_i\right] + \mathbb{E}\left[\mathbf{1}_{T>i} \cdot (Y_{i+1} - Y_i) \,|\, X_0 \ldots X_i\right]$$

$$= Y_i^T + \mathbf{1}_{T>i}(\mathbb{E}[Y_{i+1} \,|\, \ldots X_i] - Y_i)$$

($\mathbf{1}_{T>i}$, Y_i determined by $X_{\leq i}$)

(=Y_i)

Optional Stopping Theorem (OST)

(Martingale Stopping Theorem)

- <u>Optional Stopping Theorem</u> **(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

$$\mathbb{E}\left[Y_T\right] = \mathbb{E}\left[Y_0\right]$$

  if any one of the following conditions holds:

  - (bounded time) there is a finite N such that $T < N$.
  - (bounded range) $T < \infty$ a.s., and there is a finite c s.t. $|Y_t| < c$ for all 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)

- <u>Optional Stopping Theorem</u> **(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

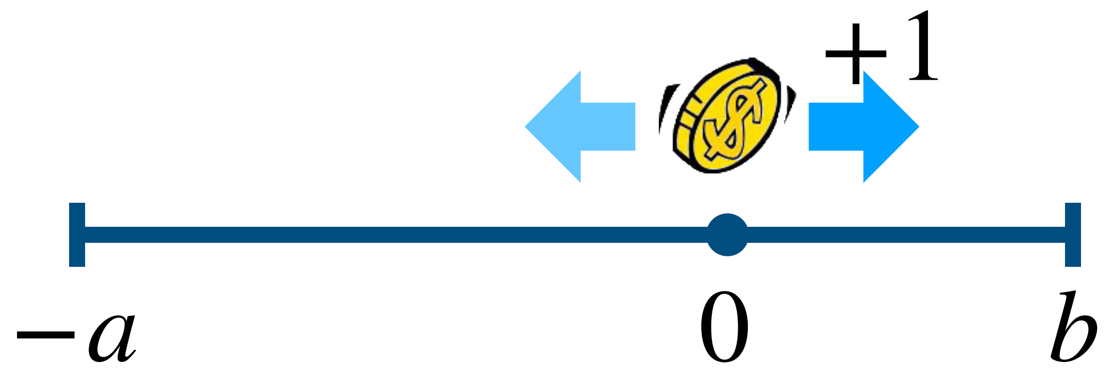
$$\mathbb{E}\left[Y_T\right] = \mathbb{E}\left[Y_0\right]$$

(general condition) if all the following conditions hold:

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

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**) R.V.s
- 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 $|Y_t^T| \leq \max\{a, b\}$ for all $0 \leq t$ and $T < \infty$ a.s.

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

$$\mathbb{E}[Y_T] = b \cdot \Pr(Y_T = b) - a \cdot \Pr(Y_T \neq b) \quad \implies \quad \Pr(Y_T = b) = \frac{a}{a+b}$$

Wald's Equation
(Linearity of expectation with randomly many random variables)

- <u>Wald's equation</u>: Let $X_1, X_2, \ldots$ be i.i.d. R.V. with $\mu = \mathbb{E}[X_i] < \infty$. Let T be 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 $Y_t = \Sigma_{i=1}^{t}(X_i - \mu)$, which is a martingale. Observe that:

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

By **OST**: $\mathbb{E}[Y_T] = \mathbb{E}[Y_1] = 0$. Note that $\mathbb{E}[Y_T] = \mathbb{E}\left[\Sigma_{i=1}^{T} X_i\right] - \mathbb{E}[T] \cdot \mu$

Optional Stopping Theorem (OST)
(Martingale Stopping Theorem)

- <u>Optional Stopping Theorem</u> **(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
$$\mathbb{E}\left[Y_T\right] = \mathbb{E}\left[Y_0\right]$$
if any one of the following conditions holds:
  - (bounded time) there is a finite N such that $T < N$.
  - (bounded range) $T < \infty$ a.s., and there is a finite c s.t. $|Y_t| < c$ for all 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 $n \leq T \leq m$ be a **stopping time**, both with respect to $\{X_t : t \geq 0\}$. Then

$$\mathbb{E}\left[Y_T \,|\, X_0, \ldots, X_{n-1}\right] = Y_n$$

- **Proof**:

$$\mathbb{E}[Y_T \,|\, X_{<n}] = \mathbb{E}\left[\mathbb{E}[Y_T \,|\, X_{<m}] \,|\, X_{<n}\right] = \mathbb{E}\left[\Sigma_{k\in[n,m]}\mathbb{E}[Y_k \cdot I(T=k) \,|\, X_{<m}] \middle| X_{<n}\right]$$

$$= \mathbb{E}\left[\Sigma_{k\in[n,m)}Y_k \cdot I(T=k) \middle| X_{<n}\right] + \mathbb{E}\left[\mathbb{E}[Y_m \cdot I(T=m) \,|\, X_{<m}] \middle| X_{<n}\right]$$

$$\mathbb{E}\left[\mathbb{E}[Y_m \cdot I(T=m) \,|\, X_{<m}] \middle| X_{<n}\right] = \mathbb{E}\left[I(T=m) \cdot \mathbb{E}[Y_m \,|\, X_{<m}] \middle| X_{<n}\right]$$

$$= \mathbb{E}\left[I(T=m) \cdot Y_{m-1} \mid X_{<n}\right]$$

- Let $\{Y_t : t \geq 0\}$ be a **martingale** and $n \leq T \leq m$ be a **stopping time**, both with respect to $\{X_t : t \geq 0\}$. Then

$$\mathbb{E}\left[Y_T \,|\, X_0, \ldots, X_{n-1}\right] = Y_n$$

- **Proof** (count.):

$$\mathbb{E}[Y_T \,|\, X_{<n}] = \mathbb{E}\left[\sum_{k \in [n,m)} Y_k \cdot I(T = k) \,\middle|\, X_{<n}\right] + \mathbb{E}\left[I(T = m) \cdot Y_{m-1} \mid X_{<n}\right]$$

$$= \mathbb{E}\left[Y_{\min\{T, m-1\}} \mid X_{<n}\right]$$

$$\cdots$$

$$= \mathbb{E}\left[Y_{\min\{T, n\}} \mid X_{<n}\right] = \mathbb{E}[Y_n \,|\, X_{<n}] = Y_n$$

Optional Stopping Theorem (OST)
(Martingale Stopping Theorem)

- Let $\{Y_t : t \geq 0\}$ be a **martingale** and T be a **stopping time**, both with respect to $\{X_t : t \geq 0\}$. If $\Pr(T < \infty) = 1$, $\mathbb{E}\left[\max_t \left|Y_t\right|\right] < \infty$ for all $t \leq T$, then
$$\mathbb{E}\left[Y_T\right] = Y_0$$
- **Proof**: $\lim_{n\to\infty} \left|\mathbb{E}\left[Y_{\min\{T,n\}}\right] - \mathbb{E}[Y_T]\right| = 0 \implies \mathbb{E}[Y_T] = \lim_{n\to\infty} \mathbb{E}\left[Y_{\min\{T,n\}}\right]$

Let $T' = \min\{T, n\}$, then $T' \in [0,n]$, so $\mathbb{E}[Y_{T'}] = Y_0$ by *bounded time* case.

Therefore, $\mathbb{E}[Y_T] = \lim_{n\to\infty} \mathbb{E}\left[Y_{\min\{T,n\}}\right] = Y_0$

Optional Stopping Theorem (OST)
(Martingale Stopping Theorem)

- Let $\{Y_t : t \geq 0\}$ be a **martingale** and T be a **stopping time**, both with respect to $\{X_t : t \geq 0\}$. If $\Pr(T < \infty) = 1$, $\mathbb{E}\left[\max_t \left|Y_t\right|\right] < \infty$ for all $t \leq T$, then
$$\mathbb{E}\left[Y_T\right] = Y_0$$

- **Proof** (cont.): Let $W \triangleq \max_t \left|Y_{\min\{T,t\}}\right|$. By assumption, $\mathbb{E}[\,|Y_T|\,] \leq \mathbb{E}[W] < \infty$.

$$\left|\mathbb{E}\left[Y_{\min\{T,n\}}\right] - \mathbb{E}[Y_T]\right| \leq \mathbb{E}\left[\left|Y_{\min\{T,n\}} - Y_T\right| I(T \geq n)\right] \leq 2\mathbb{E}[W \cdot I(T \geq n)]$$

Since $\Pr(T < \infty) = 1$ and $\mathbb{E}[W] < \infty$, $\lim_{n \to \infty} 2\mathbb{E}[W \cdot I(T \geq n)] = 0$

Optional Stopping Theorem (OST)

(Martingale Stopping Theorem)

- Let $\{Y_t : t \geq 0\}$ be a **martingale** and T be a **stopping time**, both with respect to $\{X_t : t \geq 0\}$. If $\Pr(T < \infty) = 1, \mathbb{E}[T] < \infty$, and $\mathbb{E}[\,|Y_{t+1} - Y_t|\, X_{\leq t}] \leq c$ for all t, then

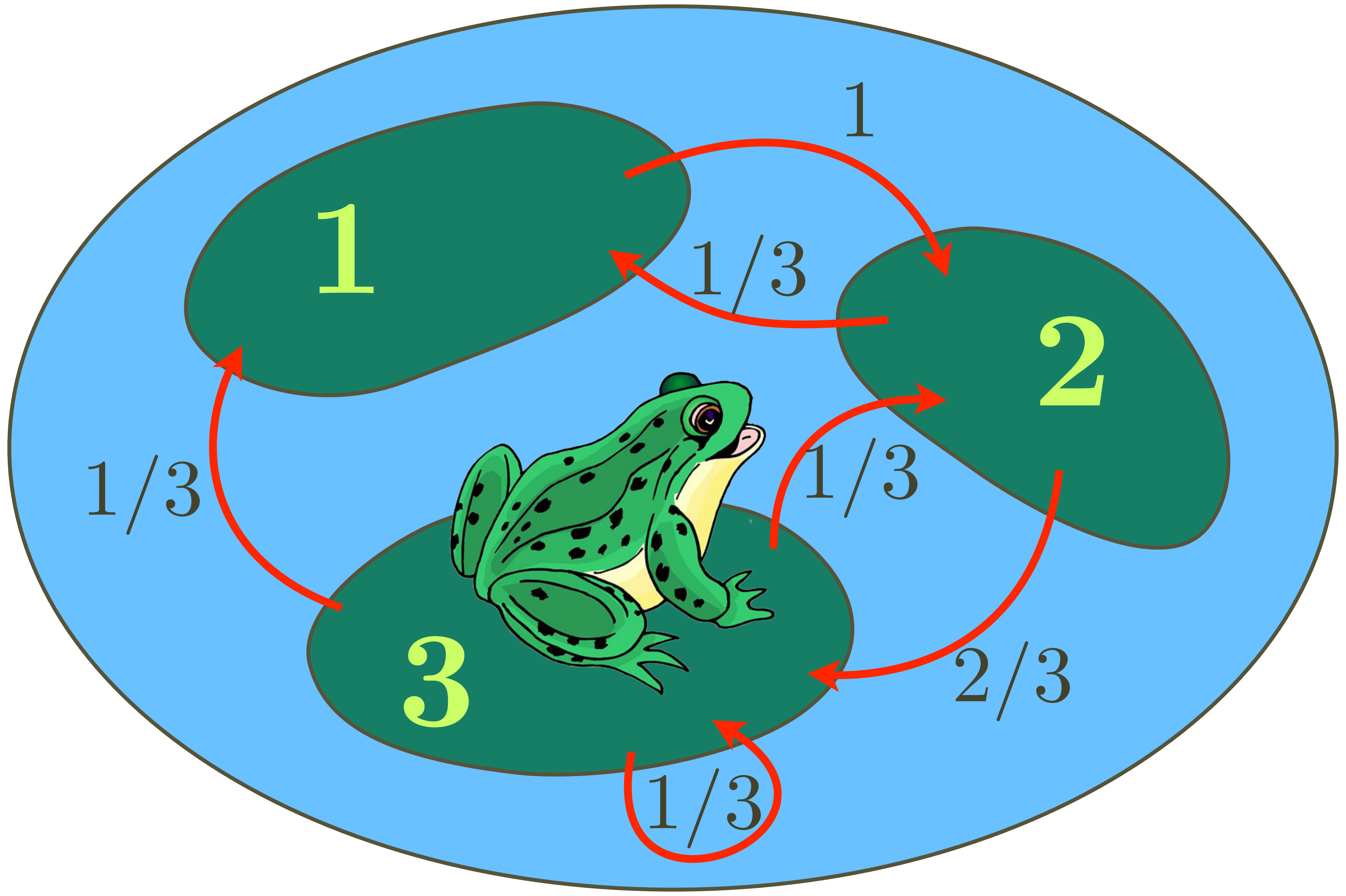
$$\mathbb{E}\left[Y_T\right] = Y_0$$

- **Proof**: Let $Z_n \triangleq |Y_n - Y_{n-1}|, Z_0 \triangleq |Y_0|, W \triangleq Z_0 + \ldots Z_T$. Clearly $W \geq |Y_T|$.

$$\mathbb{E}[W] = \Sigma_{k\geq 0}\mathbb{E}[Z_k \cdot I(T \geq k)] = \Sigma_{k\geq 0}\mathbb{E}\left[\mathbb{E}[Z_k \cdot I(T \geq k)\,|\, X_{<k}]\right]$$
$$= \Sigma_{k\geq 0}\mathbb{E}\left[I(T \geq k) \cdot \mathbb{E}[\,|Y_k - Y_{k-1}|\,\Big|\, X_{<k}]\right] \leq \Sigma_{k\geq 0} c \cdot \Pr(T \geq k)$$

$$\mathbb{E}[W] \leq \Sigma_{k\geq 0} c \cdot \Pr(T \geq k) \leq c \cdot (1 + \mathbb{E}[T]) < \infty$$

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

$$X_0 \to X_1 \to \cdots \to X_{t-1} \to X_t \to X_{t+1}$$

Transition Matrix (转移矩阵)

- A discrete-time random process $X_0, X_1, X_2, \ldots$ is a <u>**Markov chain**</u> 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) $$= P(x_t, x_{t+1}) \quad = P^{(t)}(x_t, x_{t+1})$$

- P is called the <u>**transition matrix**</u>: (assuming discrete-space)

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

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

- P is a (**row/right-)stochastic** matrix: $P \geq 0$ and $P\mathbf{1} = \mathbf{1}$

Transition Matrix (转移矩阵)

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

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

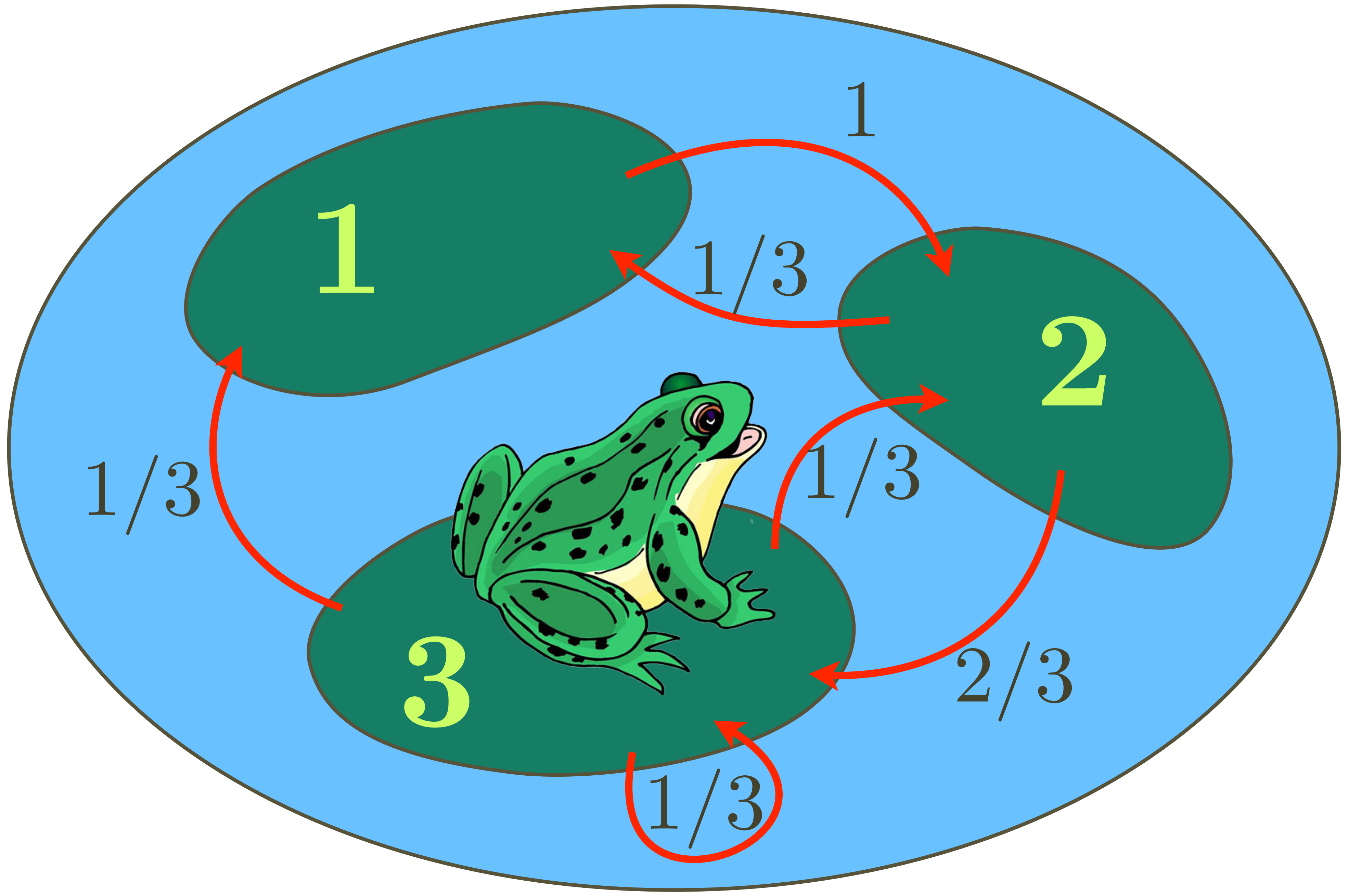
where $P \in \mathbb{R}_{\geq 0}^{\mathcal{S} \times \mathcal{S}}$ is the **transition matrix**, which is a (**row/right-**)**stochastic** matrix

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

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

$$\pi^{(0)} \xrightarrow{P} \pi^{(1)} \xrightarrow{P} \cdots \xrightarrow{P} \pi^{(t)} \xrightarrow{P} \pi^{(t+1)} \xrightarrow{P} \cdots$$

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)

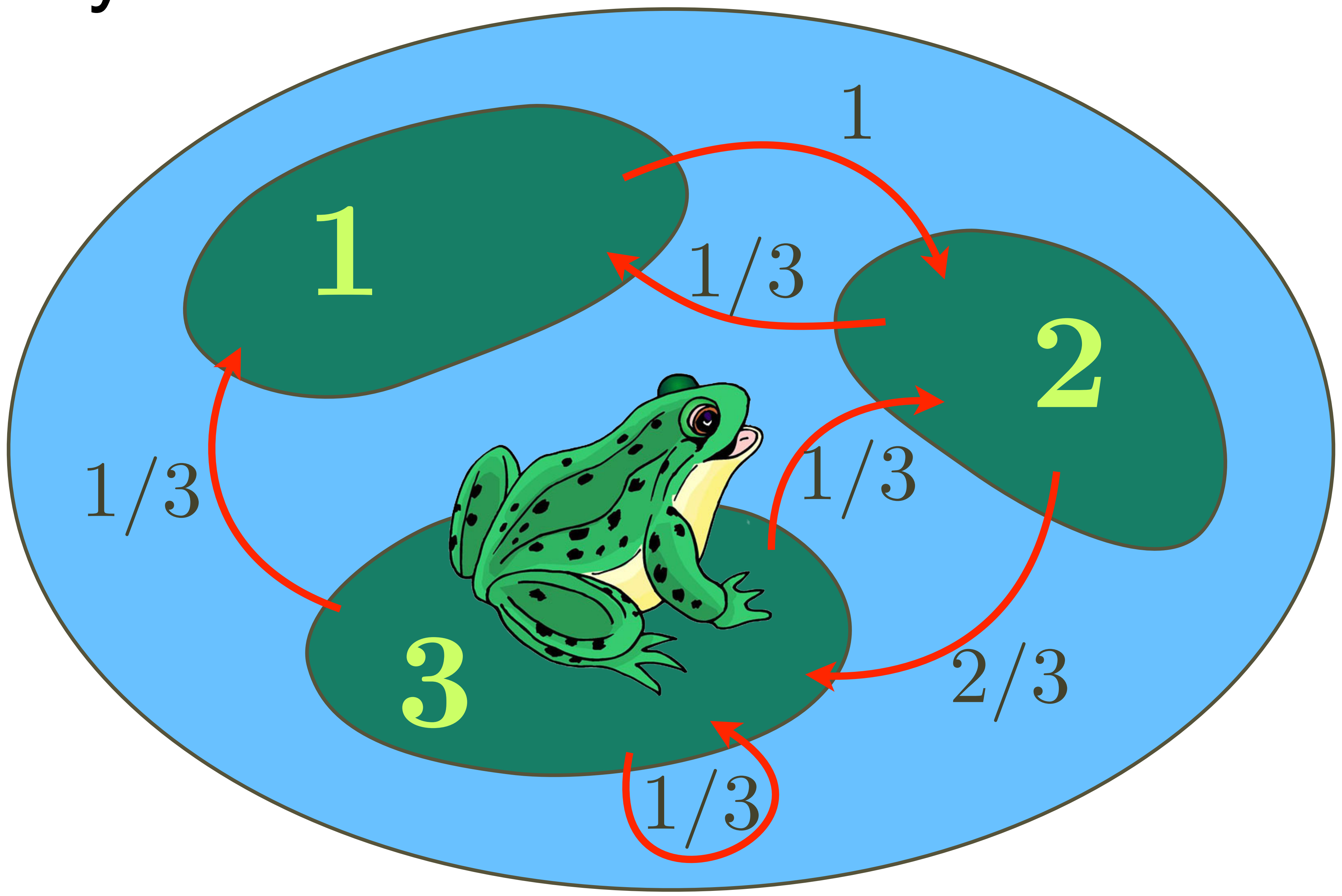
- A distribution (*pmf*) π on state space $\mathcal{S}$ is called a **stationary distribution** of the Markov chain P if

$$\pi P = \pi$$

- π is a **fixpoint (equilibrium)** of the linear dynamic system

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1/3 & 0 & 2/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \quad \pi = \left(\frac{1}{4}, \frac{3}{8}, \frac{3}{8}\right)$$

$$P^{20} \approx \begin{bmatrix} 0.2500 & 0.3750 & 0.3750 \\ 0.2500 & 0.3750 & 0.3750 \\ 0.2500 & 0.3750 & 0.3750 \end{bmatrix}$$

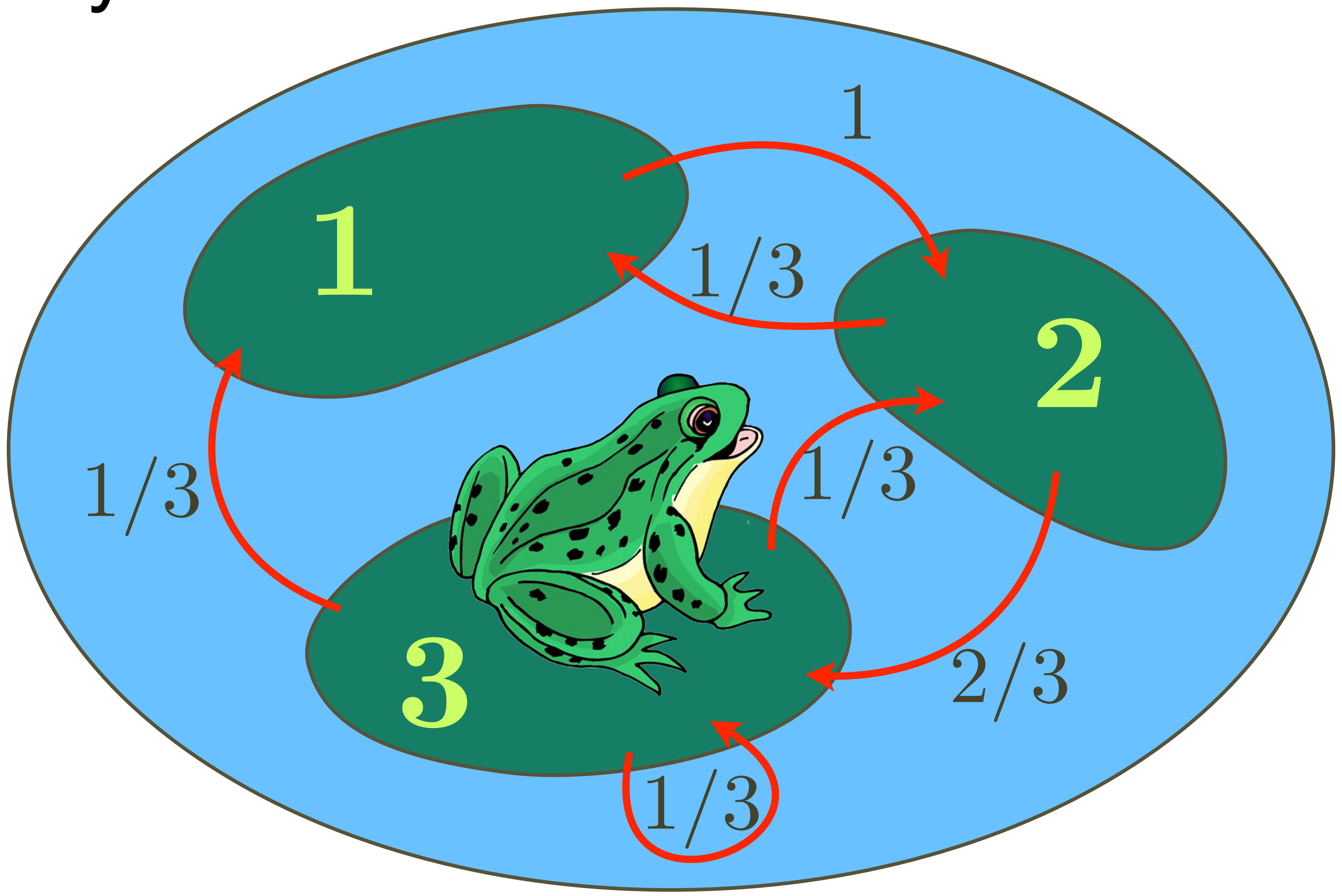


Stationary Distribution (稳态分布)

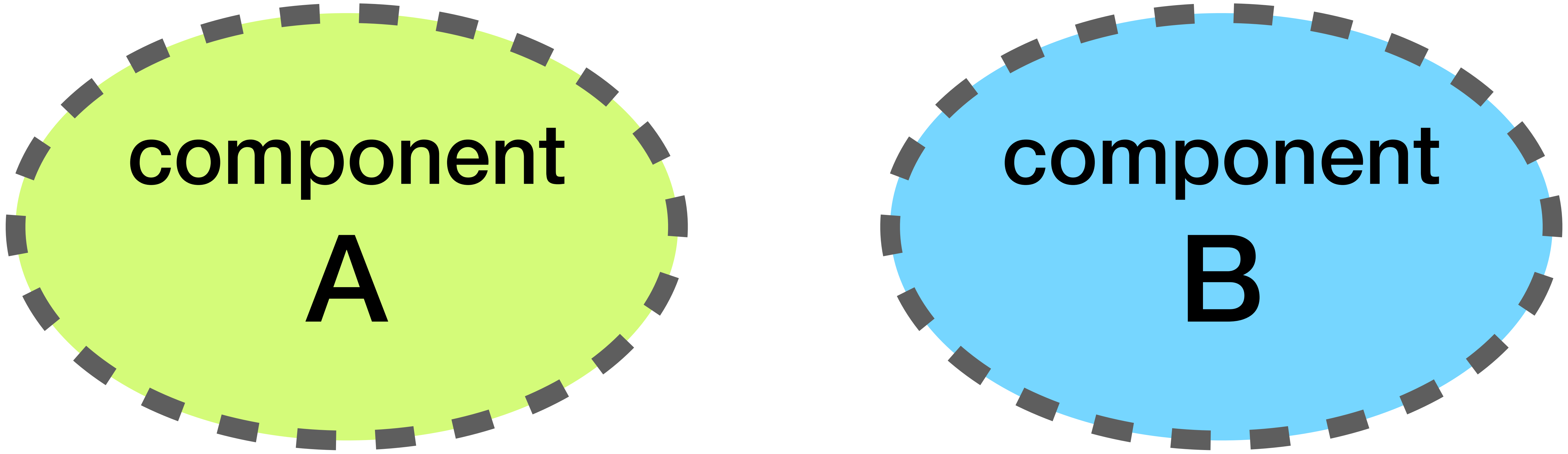
- A distribution (*pmf*) π on state space $\mathcal{S}$ is called a <u>**stationary distribution**</u> of the Markov chain P if

$$\pi P = \pi$$

- π is a **fixpoint (equilibrium)** of the linear dynamic system
- **Perron-Frobenius Theorem**:
    - stochastic matrix P: $P\mathbf{1} = \mathbf{1}$
    - 1 is also a **left eigenvalue** of P
    - **left eigenvector** $\pi P = \pi$ is nonnegative
- stationary distribution always exists

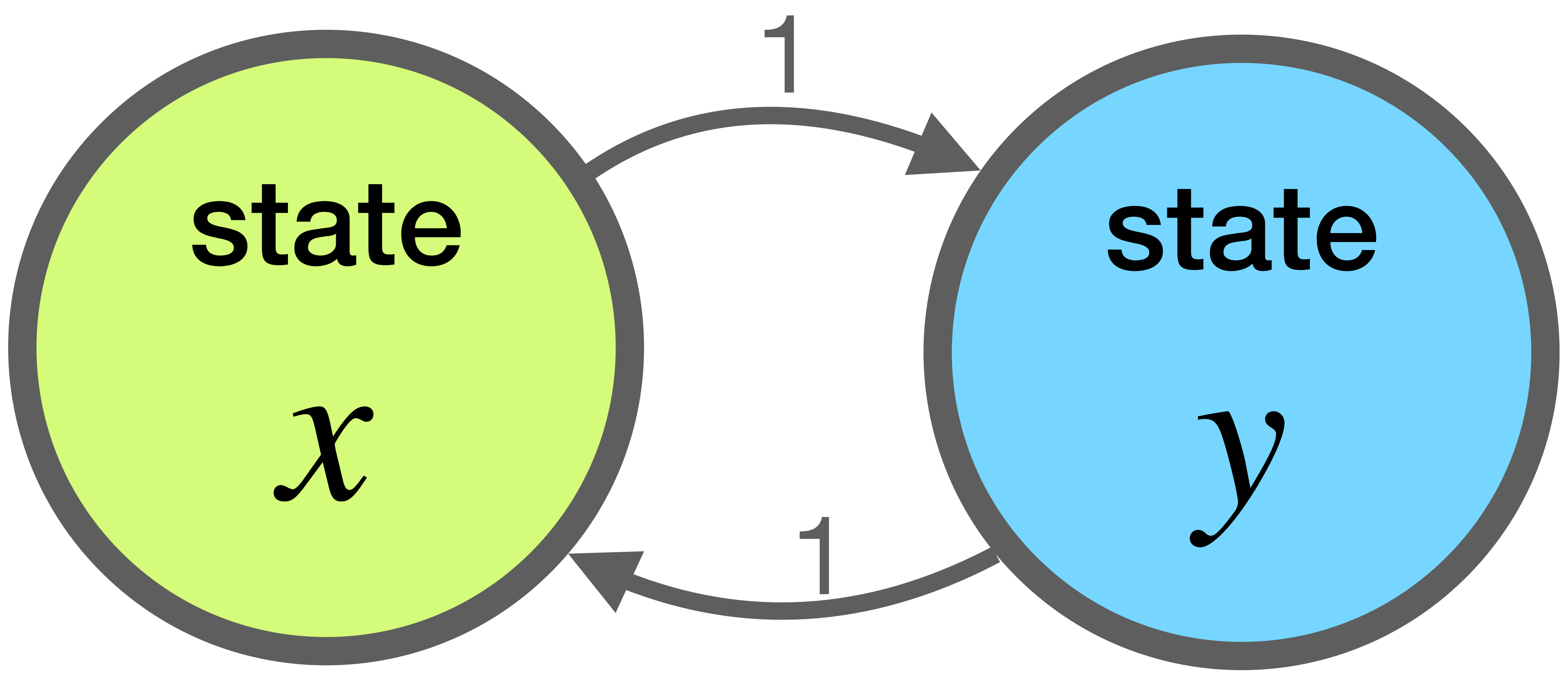


Examples



$$P = \begin{bmatrix} P_A & 0 \\ 0 & P_B \end{bmatrix}$$

stationary distributions: $\pi = \lambda \pi_A + (1 - \lambda)\pi_B$



doesn't always converge: $(a, b) \to (b, a) \to (a, b) \ldots$

Convergence Theorem

- **Irreducibility:** the chain is **irreducible** if P is an irreducible matrix (不可约矩阵)

  $\iff$ the state space $\mathcal{S}$ is **strongly connected** under P

- **Ergodicity:** the chain is **ergodic (遍历)** if all states are *aperiodic* (无周期)

  and *positive recurrent* (正常返)

Convergence Theorem

- <u>Markov chain convergence theorem</u> (Fundamental Theorem of MC):

  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 π 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 **<u>irreducible</u>** if P is an irreducible matrix (不可约矩阵)

  $\iff$ the state space $\mathcal{S}$ is **strongly connected** under P

- **Ergodicity:** the chain is **<u>ergodic</u> (遍历)** if all states are ***aperiodic*** (无周期)

  and ***positive recurrent*** (正常返)

Ergodicity

- Let $X_0, X_1, X_2 \dots$ 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 $d(x) = \text{gcd}\{t \geq 1 \mid P^t(x,x) > 0\}$
  - 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 $\mathbb{E}\left[\min\{t \geq 1 : X_t = x\} \mid X_0 = x\right] < \infty$

- *Kakutani Shizuo* (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

- <u>Markov chain convergence theorem</u> (Fundamental Theorem of MC):

  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 π on $\mathcal{S}$ such that

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

finiteness → existence

irreducibility → uniqueness

ergodicity → convergence

existence and uniqueness: } Perron-Frobenius

Convergence Theorem

- <u>Markov chain convergence theorem</u> (Fundamental Theorem of MC):

  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 π on $\mathcal{S}$ such that

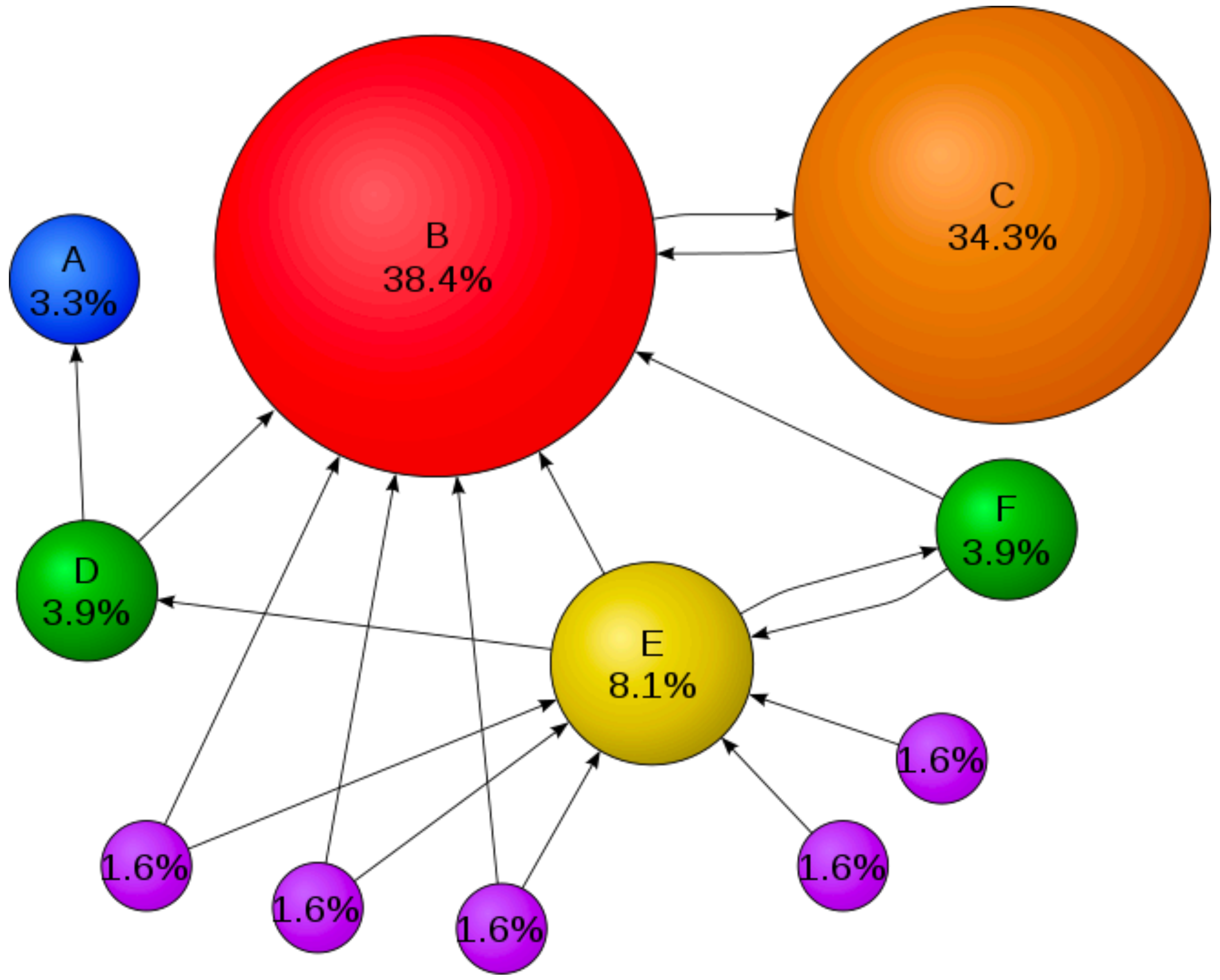
$$\pi(x) = \lim_{t \to \infty} \Pr(X_t = x \mid X_0 = x_0) \text{ for any } x_0 \in \mathcal{S}$$

- **Proof**: (By coupling)

$$\begin{array}{ccccccccccccc}
x_0 = & X_0 & \to & X_1 & \to & \cdots & \to & X_n & & & & & \\
 & \nparallel & & \nparallel & & \nparallel & & \parallel & \searrow & & & & \\
\pi \sim & Y_0 & \to & Y_1 & \to & \cdots & \to & Y_n & \to & Y_{n+1} & \to \cdots \to & Y_t \sim \pi &
\end{array}$$

irreducibility + ergodicity $\Longrightarrow$ occurs a.s.

PageRank



- Each webpage $x \in \mathcal{S}$ is assigned a rank $r(x)$:
  - High-rank pages have greater influence.
  - A page has high rank if pointed by many high-rank pages.
  - Pages pointing to few others have greater influence.
- Linear system: $r(x) = \sum_{y \to x} \frac{r(y)}{d^+(y)}$ where $d^+(y)$ is the **out-degree** of page y
- **Stationary distribution** $rP = r$ for the random walk (tireless internet surfer)

$$P(x, y) = \begin{cases} \frac{1}{d^+(x)} & \text{if } x \to y \\ 0 & \text{o.w.} \end{cases}$$

Convergence Theorem

- **Markov chain convergence theorem** (Fundamental Theorem of MC):

  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 π 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 $\pi = \pi P$

Time Reversibility

- A Markov chain P is called <u>**time-reversible**</u> or just <u>**reversible**</u> if it satisfies the ***detailed balance equation*** (***DBE***):

$$\pi(x)P(x,y) = \pi(y)P(y,x)$$

for some distribution π over the state space $\mathcal{S}$

- π is a more refined fixpoint: π must be a stationary distribution

$$(\pi P)_y = \sum_x \pi(x)P(x,y) = \sum_x \pi(y)P(y,x) = \pi(y)$$

- **Time-reversible**: assuming $X_0 \sim \pi$

$(X_0, X_1, \ldots, X_n)$ is identically distributed as $(X_n, \ldots, X_1, X_0)$

Convergence Theorem

- <u>Markov chain convergence theorem</u> **(Fundamental Theorem of MC):**

  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 π on $\mathcal{S}$ such that

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

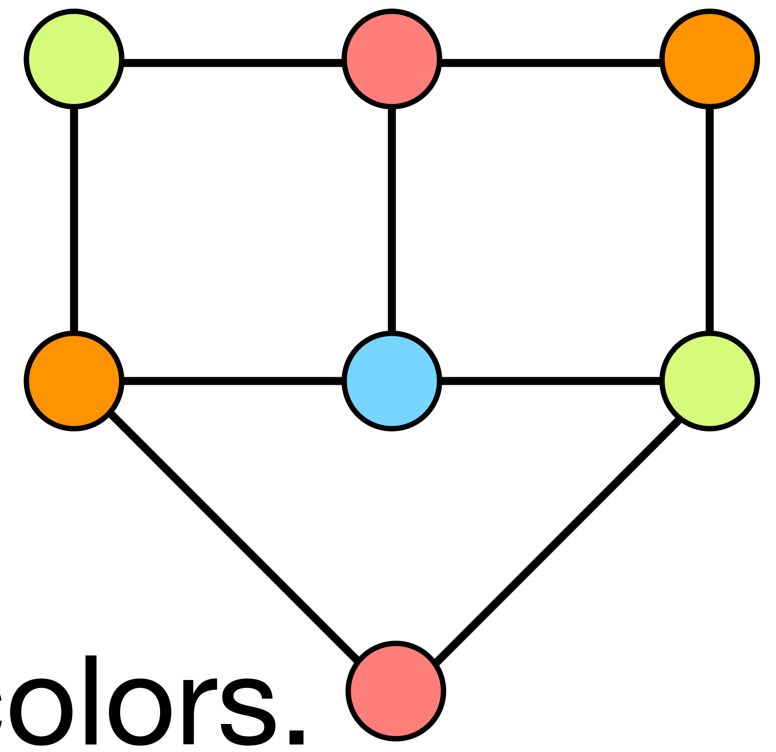
- **Finite Markov chain** (with finite state space $\mathcal{S}$):

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

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

- Detail balance equation: $\pi(x)P(x,y) = \pi(y)P(y,x)$

Markov Chains on Proper Colorings



- Let $G = (V, E)$ be a graph of maximum degree Δ and $[q]$ a set of q colors.
- Let $\Omega = \{\sigma \in [q]^V \mid \forall uv \in E, \sigma_u \neq \sigma_v\}$ be the set of all **proper q-colorings** of G

- **Glauber dynamics**:
  Initially, $X_0 \in \Omega$ is arbitrary. Transition $X_t \to X_{t+1}$:
  - choose a vertex $v \in V$ uniformly at random;
  - $X_{t+1}(u) \leftarrow X_t(u)$ for all $u \neq v$;
  - $X_{t+1}(v) \leftarrow$ uniform random *available* color in $[q] \setminus \{X_t(u) \mid uv \in E\}$;

- $q \geq \Delta + 2 \implies$ the chain is irreducible and ergodic (aperiodic)
- Symmetric $\implies$ time-reversible and the stationary distribution π is uniform over Ω

Counting Constraint Satisfaction Problem

Input: a CSP instance I.

Output: the number of CSP solutions.

Examples:

- Counting independent sets: number of independent sets in a graph.
- Counting matchings: number of matchings in a graph.
- Counting graph colorings: number of proper q-colorings of a graph.
- #SAT: number of satisfying assignments of a CNF.

They are all **#P**-hard!

uniform sampling $\Longrightarrow$ approximate counting

Mixing of Markov Chain

- <u>Markov chain convergence theorem</u>:

  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 π on $\mathcal{S}$ such that

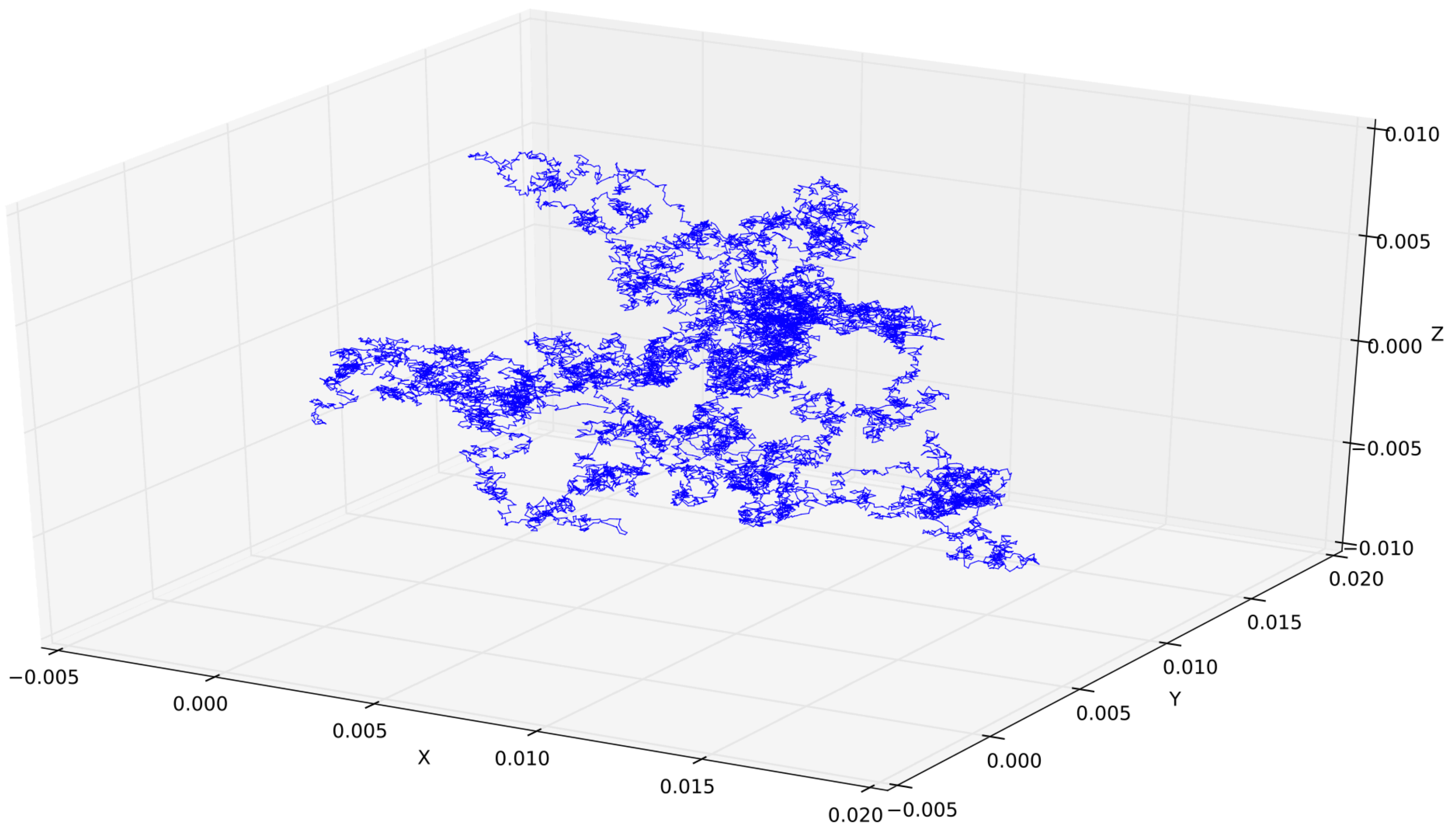
$$\pi(x) = \lim_{t\to\infty} \Pr(X_t = x \mid X_0 = x_0) \text{ for any } x_0 \in \mathcal{S}$$

- How fast is the convergence rate?

- <u>Mixing time</u>: let $\pi_x^{(t)}(y) = (\mathbf{1}_x P^t)_y = \Pr(X_t = y \mid X_0 = x)$

$$\tau(\epsilon) = \max_{x\in S} \min \left\{ t \geq 1 \,\middle|\, \left\| \pi_x^{(t)} - \pi \right\|_1 \leq 2\epsilon \right\}$$

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

  $W(s+u) - W(s) \sim \mathcal{N}(0,u)$

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 <u>**diffusion process**</u> 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!