Stochastic Process

- Till now we've modelled only single experiments.
  - Or, repeating it with new random bits.

- Now, we study a sequence of rnd experiments, that may depend on previous outcomes.

Defn: • A stochastic process is a sequence $\{X_t\}_{t \in I}$ of random variables taking values in a state space S. Index set I can be thought of as time.

- S could be discrete (eg. coin toss every day)
- S " " continuous (eg. stock price every day)

▷ $P(X_1=x_1 \wedge \cdots \wedge X_k=x_k) = P(X_1=x_1) \cdot P(X_2=x_2 \mid X_1=x_1) \cdots P(X_k=x_k \mid X_1=x_1 \wedge \cdots \wedge X_{k-1}=x_{k-1})$.

the order is fixed.

▷ Process is independent iff $P(X_1=x_1 \wedge \cdots \wedge X_k=x_k) = \prod_{1\le i\le k} P(X_i=x_i)$.

Defn: Markov Chain is a stochastic process where X_k is independent of $X_{k-2}, X_{k-3}, \ldots, X_1$.

(may depend on X_{k-1})

I.e. $P(X_k = x_k \mid X_{k-1} = x_{k-1} \wedge \cdots \wedge X_1 = x_1) = P(X_k = x_k \mid \underline{X_{k-1} = x_{k-1}})$,

for all $k \geq 2$ & $x_1, \ldots, x_k \in S$.

▷ Trivially, independent rnd. variables $\{X_i\}_i$ is a Markov Chain.

- Other egs: Toss a coin many times. Let $X_i := \#(\text{Heads till } i\text{-th toss})$, $\forall i \geq 1$. Then,

$$P(X_i = x_i \mid X_{i-1} = x_{i-1} \wedge \cdots \wedge X_1 = x_1) = P(X_i = x_i \mid X_{i-1} = x_{i-1})$$

(Why?)

$$= \begin{cases} \frac{1}{2}, & \text{if } x_i - x_{i-1} \leq 1. \\ 0, & \text{else} \end{cases}$$

$\Rightarrow \{X_i\}_{i\geq 1}$ is a Markov Chain.

$\hookrightarrow$ It's also homogeneous.

Defn: $\{X_i\}_{i\in I}$ is homogeneous Markov Chain if the conditional probability does not depend on the time. I.e. $P(X_k = i \mid X_{k-1} = j) = P(X_1 = i \mid X_0 = j)$.

Theorem: (i) If $\{X_i\}_{i\geq 0}$ are independent rnd. variables then $\{Y_i := \sum_{j\leq i} X_j\}_{i\geq 0}$ is a Markov Chain.

(ii) If $\{X_i\}_{i\geq 0}$ are independent identically distributed (i.i.d.) then $\{Y_i\}_{i\geq 0}$ is a homogeneous Markov Chain.

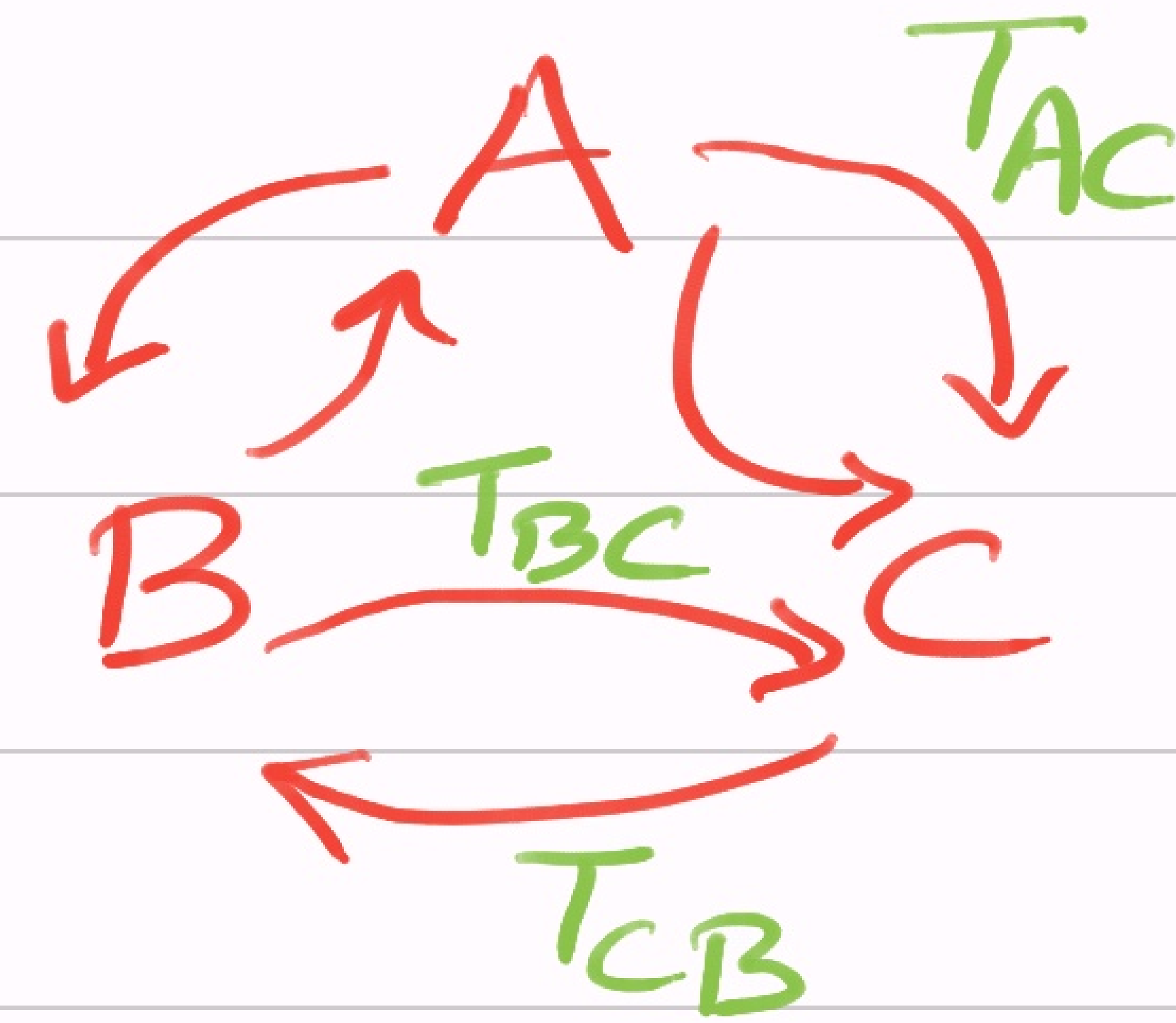
Pf: (i) $P(Y_i = y_i \mid Y_{i-1} = y_{i-1} \wedge \cdots \wedge Y_1 = y_1) = P(X_i = y_i - y_{i-1})$

$= P(Y_i = y_i \mid Y_{i-1} = y_{i-1})$.

(ii) $= P(X_1 = y_i - y_{i-1})$ [$\because P(X_i = \varepsilon) = P(X_1 = \varepsilon)$.]

$= P(Y_1 = y_i \mid Y_0 = y_{i-1}) \Rightarrow$ homogeneity. $\square$

- Representing in a figure:



$S = \{A, B, C\}$ ← state-space

Transition probabilities: $T_{AC}, \ldots$

- Equivalently, we use matrices:

Defn: • T is an $|S| \times |S|$ matrix; entries in $[0,1]$.

• (i,j)-th entry is $T_{ij} := P(X_1 = j \mid X_0 = i)$.

• T is the transition matrix of a (homog.) Markov chain.

• Also, specify the initial probability distribution, to start the process. Call it column-vector $\mu \in [0,1]^{|S|}$, with $\mu_i := P(X_0 = i)$.

▷ $|\mu| = \sum_i \mu_i = 1$. [Pf: $\sum \mu_i = \sum_i P(X_0 = i) = 1$ due to partition.]

▷ Each row ~~(or column)~~ of T sums to 1.

Such matrices are called stochastic (not doubly-stochastic.)

Pf: • Consider i-th row: $|T_{i*}| = \sum_{j \le |S|} T_{ij} = \sum_j P(X_1 = j \mid X_0 = i)$
$= 1$, by partition.

~~• Similarly, column-sum $|T_{*j}| = 1$.~~ X Why? □

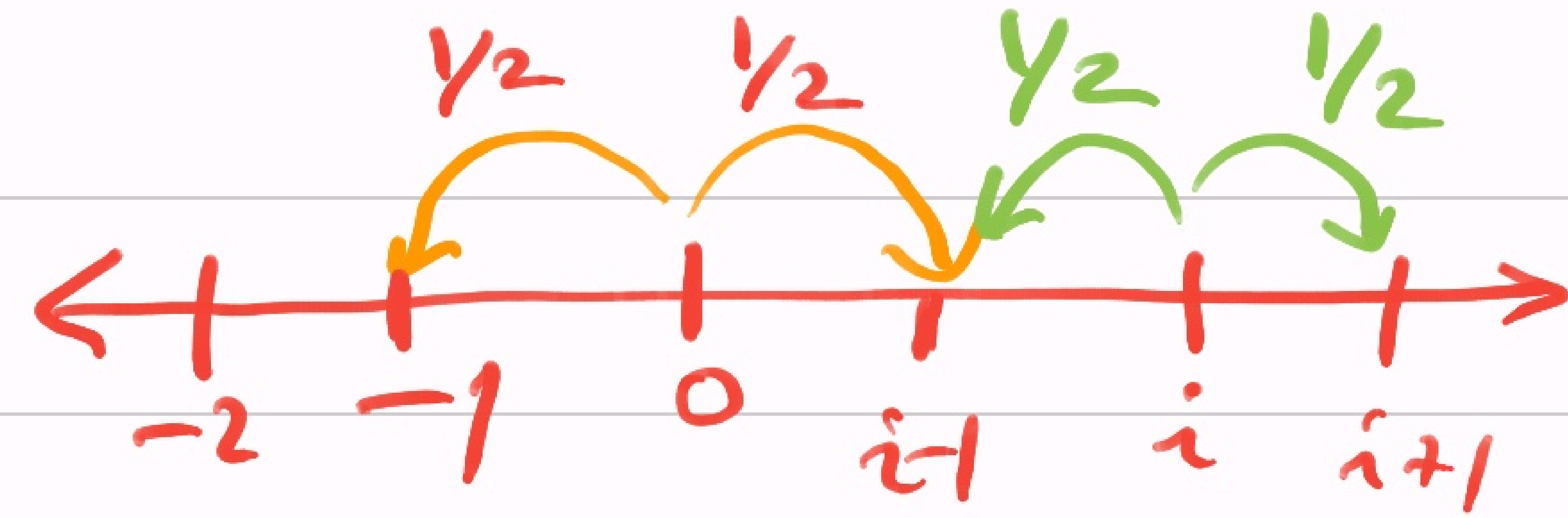
▷ Any $\mu \in [0,1]^{|S|}$ & $T \in [0,1]^{|S| \times |S|}$ s.t. $|\mu| = 1$ & T is stochastic, define a (homog.) Markov chain over state-space of size $|S|$.

Converse is also true!

Pf: [exercise]. □

eg. Drunkard's Walk in one-dimension

- State-space $S = \mathbb{Z}$.



- Transition Probabilities:

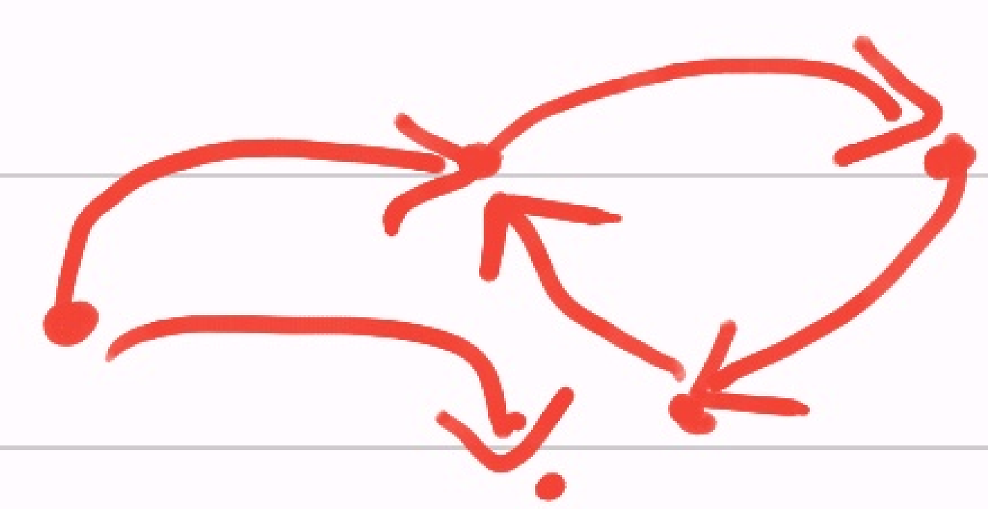
$$T_{ij} := \begin{cases} 1/2, & \text{if } |i-j| = 1. \\ 0, & \text{else} \end{cases}$$

- Initial distribution: $\mu_i := \begin{cases} 1, & \text{if } i = 0 \\ 0, & \text{else} \end{cases}$

— Similarly, one can define it in multi-dimension.

— Random walk on a graph $G = (V, E)$:

use adjacency matrix to model the walk



- Many examples from other fields use Markov chain modeling (i.e. memorylessness!):
e.g. evolution of DNA, Brownian motion, information theory, Internet search, etc.
- This is because of its strong structural properties:

Evolution of a Markov Chain

- Let the process be given by column vector $\underline{\mu}$ & matrix $\underline{M}$. S is its state-space.
- At $t=0$: the prob. distribution on S is given by μ.

Denote $p_0 := \mu$.

- What's the prob. distribution at $t=1$? Call it p_1.

(Think of $p_0, p_1, \dots$ as column vectors.)

(row)

Lemma 1: $p_1^T = p_0^T \cdot M$.

Pf: • j-th entry on RHS $= \sum_{i=1}^{|S|} (p_0)_i \cdot M_{ij}$

$= \sum_i P(X_0 = i) \cdot P(X_1 = j \mid X_0 = i) = \sum_i P(X_0 = i \wedge X_1 = j)$

$= P(X_1 = j) = (p_1)_j =$ j-th entry on LHS. $\square$

Theorem (Evolution): $\forall n \geq 1$, $p_n^T = p_0^T \cdot M^n$.

Pf: • By Lemma 1 & homogeneity: $p_n^T = p_{n-1}^T \cdot M = p_{n-2}^T \cdot M \cdot M = \dots = p_0^T \cdot M^n$. $\square$

⇒ Markov chain evolves via M-multiplication.

- It allows us to use matrix algebra to study p_n!
(real)

- Qn: How does evolution end? What's $\lim_{n\to\infty} p_n = ?$

→ For eg. • does the limit exist?
• does it depend on p_0? (Memoryless?)
• How fast can you compute the limit (if one exists)?