Till now – revision of prob + stats (with a sim focus) - random number gen
- variance redn

Rest of course – Discrete event simulation (systems with state)

– Generating 'complex' random processes (MCMC)

ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

Unit 9: Intro to Markov Chains

Sid Banerjee

School of ORIE, Cornell University

random process = $U[0,1]$ source + memory (state) (all random proc are Markov)

random process

index

indexed collection of rvs $X_t \in \mathcal{S}$, one for each $t \in \mathcal{T}$

– $\mathcal{S}$: state space, $\mathcal{T}$: index set

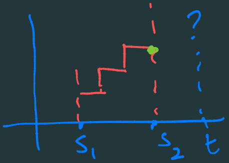
Markov property

random process X_t has the Markov property if the probability of moving to a future state only depends on the present state and not on past states

- $T = \{0, 1, 2, \ldots\}$, $(X_0, X_1, X_2, \ldots, X_t)$

$$\mathbb{P}[X_{t+1} = x \mid X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t] = \mathbb{P}[X_{t+1} = x \mid X_t = x_t]$$

- $T = [0, \infty)$,



$$\mathbb{P}[X_t = x \mid X_{t'}, t' \in [s_1, s_2]] = \mathbb{P}[X_t = x \mid X_{s_2}]$$

most recent known state

- Basic prob model = indep r.v.
- Markov prop covers all other processes! You can define 'state'

random process

random process

indexed collection of rvs $X_t \in \mathcal{S}$, one for each $t \in T$

– $\mathcal{S}$: state space, $\mathcal{T}$: index set

Markov property

random process X_t has the Markov property if the probability of moving to a future state only depends on the present state and not on past states

four types

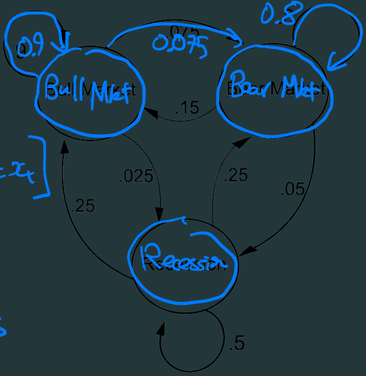
- $\mathcal{S}$ discrete, $\mathcal{T}$ discrete: discrete-time Markov chain (DTMC)
  - random walk
- $\mathcal{S}$ ~~continuous~~ discrete, $\mathcal{T}$ ~~discrete~~ continuous: continuous-time Markov chain (CTMC)
  - Poisson process
- $\mathcal{S}$ ~~discrete~~ continuous, $\mathcal{T}$ ~~continuous~~ discrete: discrete-time Markov process
- $\mathcal{S}$ continuous, $\mathcal{T}$ continuous: Markov process
  - Brownian motion

S discrete = chain

S continuous = Markov process

Markov chains: basic definition (Discrete Time - DTMC)

- $T = \{0,1,2,\dots\}$, S = some discrete set $\{0,1,2,\dots\}$

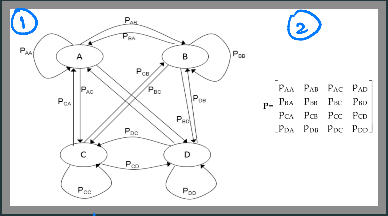


- Collection $(X_0, X_1, \dots)$ of rvs s.t.

$$\mathbb{P}[X_{t+1}=x \mid X_0=x_0,\dots,X_t=x_t] = \mathbb{P}[X_{t+1}=x \mid X_t=x_t]$$

- Can represent each state as a node in a graph, edges represent transition prob

Markov chains: transition-diagram and transition matrix



← $S = \{A, B, C, D\}$

$T = \{0, 1, 2, \ldots\}$

← 2 representations of MC

Idea – MC completely defined by trans prob $P_{xy} = \mathbb{P}[X_{t+1} = y \mid X_t = x]$

time homogeneous

(2) $P = \begin{pmatrix} P_{AA} & P_{AB} & P_{AC} & P_{AD} \\ & \ddots & & \\ & & \ddots & \\ & & & P_{DD} \end{pmatrix}$

- Need $P_{xy} \geq 0 \ \forall x, y \in S$
- Need $\sum_{x \in \{A,B,C,D\}} P_{Ax} = 1$ (every row sums to 1)

(stochastic matrix)

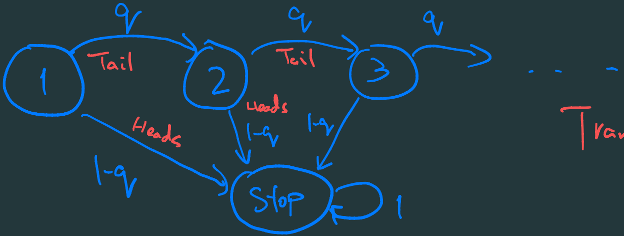
(1) directed graph with edge wts

- All edge wts > 0 ($P_{xy} = 0 \Rightarrow$ no edge betn x, y)
- 'Out degree' $\sum_y$ wt of $(x,y) = 1 \ \forall x$
- Can have self loops

example: coin tosses and geometric rv.

recall the Geometric rv $p(k) = q^{k-1}(1-q)\,\forall\, k \in \{1, 2, \ldots\}$ (# of coin tosses till heads)

we can view this as a Markov chain as follows:



Transition diagram

$$P = \begin{array}{c|ccccc} & \text{Stop} & 1 & 2 & 2 & \cdots \\ \hline \text{Stop} & 1 & 0 & 0 & \cdots & \\ 1 & 1-q & 0 & q & 0 & \cdots \\ 2 & 1-q & 0 & 0 & q & \cdots \\ \vdots & & & & & \ddots \end{array}$$

Transition matrix

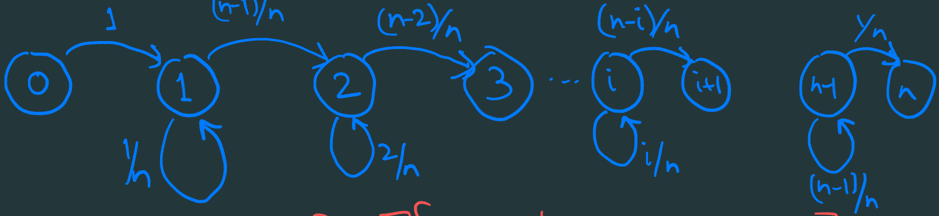
(Note: can be infinite!)

a brand of cereal always distributes a baseball card in every cereal box, chosen randomly from a set of n distinct cards

Markov chain model for number of cards owned by a collector:

$S \equiv$ # of unique cards owned by collector $= \{0, 1, 2, \ldots, n\}$

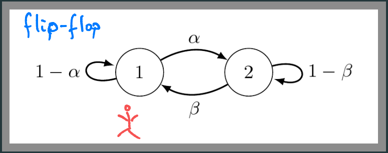
$T \equiv \{0, 1, 2, \ldots\} \equiv$ buy new cereal box



Q: $\mathbb{E}[\text{Time to collect all coupons}] = ?$

$(= H_n = \sum_{i=1}^{n} 1/i \approx n \ln n)$

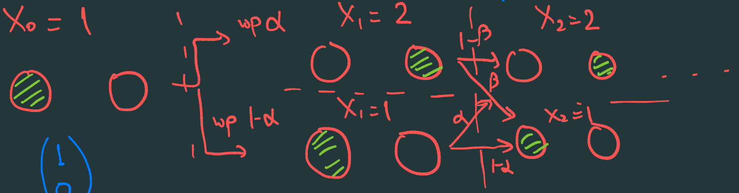
Markov chains: two viewpoints



1) Random particle view

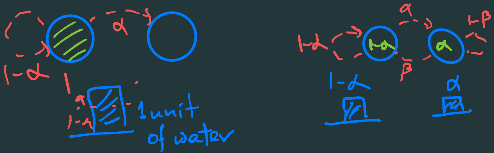
2) Flow view | Distribution view

RP view .



$\begin{pmatrix}1\\0\end{pmatrix}$

$\pi_0 = \begin{pmatrix}\mathbb{P}[X_0=1]\\ \mathbb{P}[X_0=2]\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}$ $\qquad \pi_1 = \begin{pmatrix}1-\alpha\\ \alpha\end{pmatrix}$ $\qquad \boxed{\pi_t^T = \pi_{t-1}^T P}$



$(1-\alpha)^2 + \alpha\beta$ $\qquad \alpha(1-\beta) + \alpha(1-\alpha)$

$\begin{pmatrix}\pi_2(1) & \pi_2(2)\end{pmatrix} = \begin{pmatrix}1-\alpha & \alpha\end{pmatrix}\begin{pmatrix}1-\alpha & \alpha\\ \beta & 1-\beta\end{pmatrix}$

Markov chains: long-term behavior

Possible long-term behavior

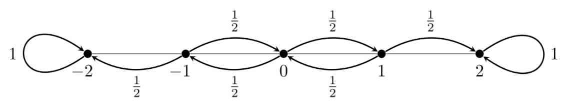
- $\boxed{\text{absorption}}$ ≡ particle gets stuck
- transient ≡ particle 'escapes' off to ∞ (or oscillates)
- recurrent ≡ particle 'keeps coming back' to state 0

undesirable

$\boxed{\text{positive recurrent}}$

null recurrent

Markov chains: absorbing chain

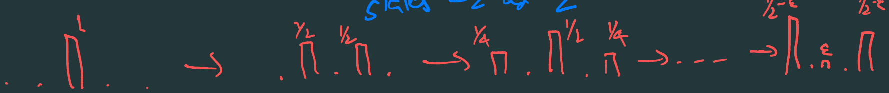


(gambler's ruin)

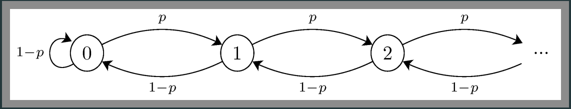
States $\{-2, 2\}$ are absorbing

— Particle view - particle eventually gets stuck (either at -2 or 2) with prob 1

— Flow view - All 1 unit of flow end up in states -2 and 2



Markov chains: transient/recurrent chain



birth-death

- Rough idea – Particle visits 0 'infinitely often'

  ⇒ Chain is recurrent

  Particle never visits 0 after some t

  ⇒ chain is transient (Eg, $p=1$)

- <u>Flow view</u> – the amount of liquid in each state 'settles down' (converges) – positive recurrent

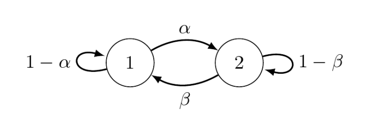
Markov chains: steady-state behavior

- $\pi_{t+1}^T = \pi_t^T P$ - MC flow update eqn
- In steady-state - $\boxed{\pi_t \xrightarrow{t\uparrow\infty} \pi \quad \text{s.t. } \pi^T = \pi^T P}$

  (if MC is positive recurrent)

  'flow-balance'

example:



If $\alpha > 0, \beta > 0$

Steady-state - $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)^T$

Check - $\pi^T P = \begin{pmatrix}\frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta}\end{pmatrix}\begin{pmatrix}1-\alpha & \alpha \\ \beta & 1-\beta\end{pmatrix} = \pi^T$

a little diversion (to mark the day. . .)

Bertrand's ballot problem

two candidates A and B contest an election with n votes, out of whom a majority a people vote for A and $b = n - a < a$ vote for B

if votes are counted in random order, what is the chance A is always in the lead?

$a > n/2$

Want to model as a Markov chain

- $T = \{0, 1, 2, \ldots, n\}$
- $X_t = \#$ of votes A leads by after t votes are counted
- $S = \{-n, -n+1, \ldots -1, 0, 1, \ldots, n\} \times T$

$\{-b, -b+1, \ldots, -1, 0, 1, \ldots, a\} \times T$

$a_t + b_t = t$
$a_t - b_t = i$

$a_t = (t+i)/2$
$b_t = (t-i)/2$

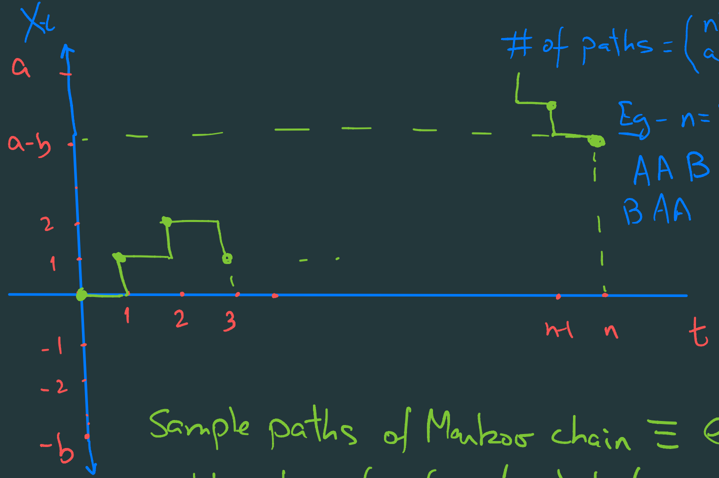
x_0, $t=0$; x_1, $t=1$

$$\begin{tikzcd}
 & (1,1) \\
(0,0) \arrow[ur, "a/n"] \arrow[dr, "b/n"'] & \\
 & (-1,1)
\end{tikzcd}$$

$$\begin{tikzcd}
(i-1,t+1) & (i,t) \arrow[l, "\frac{b-b_t}{n-t}"'] \arrow[r, "\frac{a-a_t}{n-t}"] & (i+1,t+1)
\end{tikzcd}$$

Random walk view

(Markov chain Monte Carlo)



$\#$ of paths $= \binom{n}{a} \approx n^a$

Eg - $n = 3$, $a = 2$

AAB, ABA, BAA

Sample paths of Markov chain $\equiv$ every possible step fn from $(0,0)$ to $(n, a-b)$ such that at each time you go up or down by 1 step

and a little magic! (Recall - X_t = A's lead after t votes are counted)

Bertrand's ballot theorem

two candidates A and B contest an election with n votes, out of whom a majority a people vote for A and $b = n - a < a$ vote for B

if votes are counted in random order, then $\mathbb{P}[A \text{ is always in the lead}] = 2\left(\frac{a - n/2}{n}\right) = \frac{a-b}{a+b}$

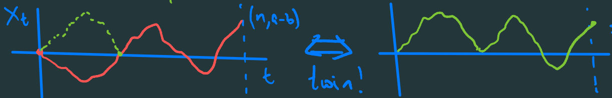
Reflection method

$$\mathbb{P}[\text{A always in lead}] = 1 - \mathbb{P}[\overbrace{\text{There is a tie at some time}}^{\text{ie, } X_t = 0 \text{ for some } t}]$$

$$= 1 - \underbrace{\mathbb{P}[X_1 = 1 \text{ and } \exists t \text{ st } X_t = 0]}_{M} - \underbrace{\mathbb{P}[X_1 = -1 \text{ and } \exists t \text{ st } X_t = 0]}_{N} \quad \leftarrow \text{but } M = N!$$

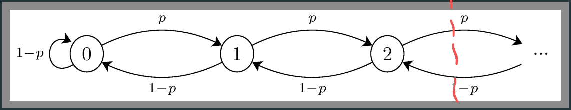
$$= 1 - 2 \cdot \mathbb{P}[X_1 = -1] = 1 - \frac{2b}{a+b} = \frac{a-b}{a+b}$$

Why is $M = N$? Because every sample path in $\{X_1 = -1, \exists t \text{ s.t. } X_t = 0\}$ has a unique twin sample path in $\{X_1 = 1, \exists t \text{ s.t. } X_t = 0\}$



$\Rightarrow \mathbb{P}[X_1 = -1, \exists t \text{ s.t. } X_t = 0]$
$= \mathbb{P}[X_1 = 1, \exists t \text{ st } X_t = 0]$
$= b/a+b$!

Markov chains: steady-state for infinite chains



flow balance

$p < 1-p$
(ie, $\underline{p < 1/2}$)

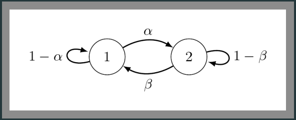
<u>Claim</u> $$\pi(i) = \left(\frac{p}{1-p}\right)^i \left(1 - \frac{p}{1-p}\right)$$ >0

<u>How to check?</u> - $\underbrace{\pi^T P = \pi^T}_{\text{infinite \# of eqns}}$, $\underbrace{\sum_{i=0}^{\infty} \pi(i) = 1}_{\text{easy to check}}$

Flow balance for cut $\{0, 1, \dots i\}$ $\{i+1, i+2, \dots \infty\}$

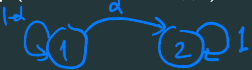
To check: $\pi(i) P_{i \to i+1} = \pi(i+1) P_{i+1 \to i} \Rightarrow \left(\frac{p}{1-p}\right)^i \cdot p = \left(\frac{p}{1-p}\right)^{i+1} (1-p)$ ✓

clicker question: long-term behavior in the flip-flop



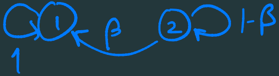
in the flip-flop Markov chain, which of the following outcomes is possible for $\pi_t(1) = \mathbb{P}[X_t = 1]$ (for different α, β)?

(a) $\lim_{t\to\infty} \pi_t(1) = 0$ $\quad \beta = 0,\ \alpha \neq 0$



9

(b) $\lim_{t\to\infty} \pi_t(1) = 1$ $\quad \alpha = 0,\ \beta \neq 0$



19

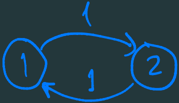
(c) $\lim_{t\to\infty} \pi_t(1)$ settles down to a constant $\pi(1) \in (0, 1)$

39

$\alpha \in (0,1),\ \beta \in [0,1)$

$\Rightarrow \pi = \begin{pmatrix} \beta/\alpha+\beta \\ \alpha/\alpha+\beta \end{pmatrix}$

(d) $\lim_{t\to\infty} \pi_t(1)$ oscillates



10

(e) all of these ✓

22

DTMC: applications and problems

Applications

1) Model complex phenomena over time
(eg. Ballot problem, coupon collector)

2) Generate complex random variables (MCMC)

Problem

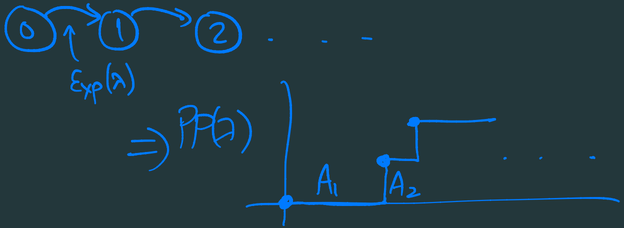
- Specifying DTMCs may be complex
  - multiple 'events' between 2 times
  - order of events matter

from discrete to continuous time

Markov property

random process X_t has the Markov property if the probability of moving to a future state only depends on the present state and not on past states

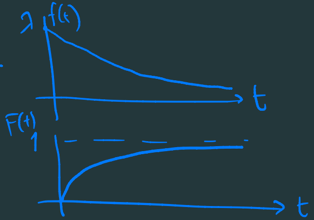
- $\mathcal{S}$ discrete, $\mathcal{T}$ discrete: discrete-time Markov chain (DTMC)
  - random walk
- $\mathcal{S}$ ~~continuous~~ discrete, $\mathcal{T}$ ~~discrete~~ continuous: continuous-time Markov chain (CTMC)
  - Poisson process – PP is the 'generator' for all CTMC



exponential distribution cheat sheet $\lambda \equiv 1/\text{sec} \Rightarrow \lambda \equiv$ 'rate'

can extend DTMCs to continuous time using special properties of the exponential distribution/poisson process

suppose $T \sim$exponential(λ), then: $\mathbb{E}[T] = 1/\lambda$



- pdf: $f_T(t) = \lambda e^{-\lambda t} \quad \forall\, t \geq 0$
- cdf: $F_T(t) = \mathbb{P}[T \leq t] = 1 - e^{-\lambda t}$
- (memorylessness): cdf of T knowing it is bigger than t?

$$\mathbb{P}[T \leq t + x | T > t] = \frac{\mathbb{P}[t < T \leq t+x]}{\mathbb{P}[T > t]} = \frac{F(t+x) - F(t)}{1 - F(t)} \quad \text{(any r.v.)}$$

$$= \frac{e^{-\lambda t} - e^{-(t+x)\lambda}}{e^{-t\lambda}} = 1 - e^{-x\lambda} = F(x)$$

$= \mathbb{P}[T \leq x]$ iff $T \sim \mathcal{E}xp(\lambda)$

(or for discrete, $T \sim$ Geom(p))

understanding exponential distributions

suppose $T_1, T_2, \dots, T_n$ are all independent, exponentially distributed, with $T_i \sim$exponential(λ_i).

- (minimum of exponentials): let $T_{\min} = \min\{T_i | i \in \{1, 2, \dots, n\}\}$; distribution of $T_{\min}$?

  $T_{\min} \sim$ $\text{Exp}\left(\sum_{i=1}^{n} \lambda_i\right)$

- (first arrival): let $T_{\min} = \arg\min_i \cancel{\min}\{T_i | i \in \{1, 2, \dots, n\}\}$; distribution of $T_{\min}$?

  $T_{\min} \sim$ i wp $\dfrac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}$

clicker question: spreading a rumor

$N(t)$  $n=3$

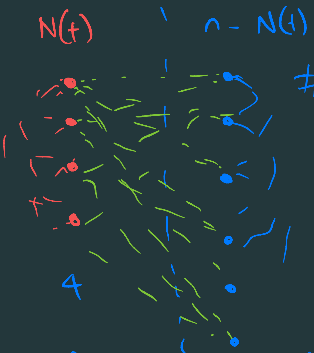
we model rumor spreading among n people using a Markovian model:

– each pair of people (i, j) independently meet after Exponential$(1/\tau)$ time

– when a person in the know meets someone who is unaware, then the rumor spreads

suppose at time t, there are $N(t)$ people who know the rumor

what is the distribution of the time T after which the number of people in the know increases to $N(t) + 1$?



# of green edges = $N(t)\left(n-N(t)\right)$

~~(a)~~ Exponential$(N(t)/\tau)$

~~(b)~~ Exponential$(N(t)\tau)$

~~(c)~~ Poisson$(N(t)^2/\tau)$ $\mathrm{Exp}\left(\frac{1}{\tau}N(t)(n-N(t))\right)$

~~(d)~~ Exponential$(N(t)^2/\tau)$

~~(e)~~ Exponential$(N(t)(N(t) + 1)/2\tau)$

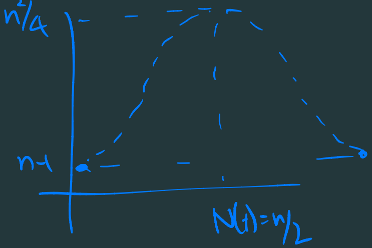
min of green clocks $\sim$ $\mathrm{Exp}\left(\frac{1}{\tau}N(t)(n-N(t))\right)$

$$1 \xrightarrow{Exp(1.(n-1))} 2 \xrightarrow{Exp(2.(n-2))} 3 \xrightarrow{Exp(3,(n-3))} 4 \quad \cdots$$

$\longrightarrow$

agent $i \to$ agent j

rate of spread

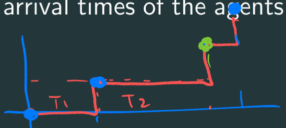


Poisson process cheat sheet

given Poisson processes $X(t) \sim PP(\lambda)$

- (inter-arrival times): let $\{A_1, A_2, A_3 \ldots\}$ be the arrival times of the agents; then

  $T_i = A_i - A_{i-1} \sim \mathrm{Exp}(\lambda)$



- (splitting): suppose we probabilistically split arrivals from $X(t)$ to $Y(t)$ with probability p, else to $Z(t) = X(t) - Y(t)$

  $Y(t) \sim PP(p\lambda)$ and $Y(t) \perp\!\!\!\perp Z(t)$

  $Z(t) \sim PP((1-p)\lambda)$

- (time-varying rate): a time-varying rate of $\lambda(t) \in [0, \lambda^\star]$ is equivalent to a $PP(\lambda^\star)$ for which arrivals at time t are thinned with probability $p(t) = \lambda(t)/\lambda^\star$

Poisson process cheat sheet (contnd)

given independent Poisson processes
$X_1(t) \sim PP(\lambda_1), X_2(t) \sim PP(\lambda_2), , X_3(t) \sim PP(\lambda_3)$

- (superposition): suppose $S(t) = X_1(t) + X_2(t) + X_3(t)$

  $S(t) \sim \quad PP(\lambda_1 + \lambda_2 + \lambda_3)$

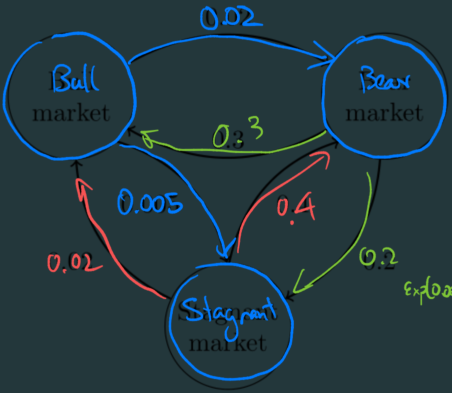
- (first arrival): let $I_{\min}$ = the identity (i.e., $\{1, 2, 3\}$) of the first arrival among the three processes; distribution of $I_{\min}$?

  $I_{\min} \sim \quad i \text{ wp } \lambda_i / \sum_{i=1}^{3} \lambda_i$

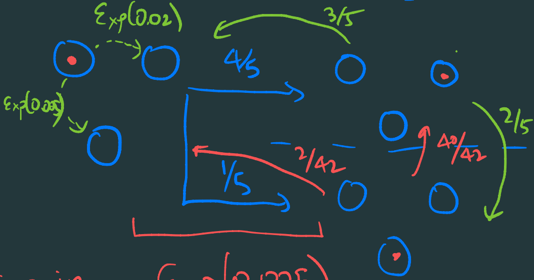
continuous-time Markov chains

CTMC: continuous-time Markov processes on discrete state-space

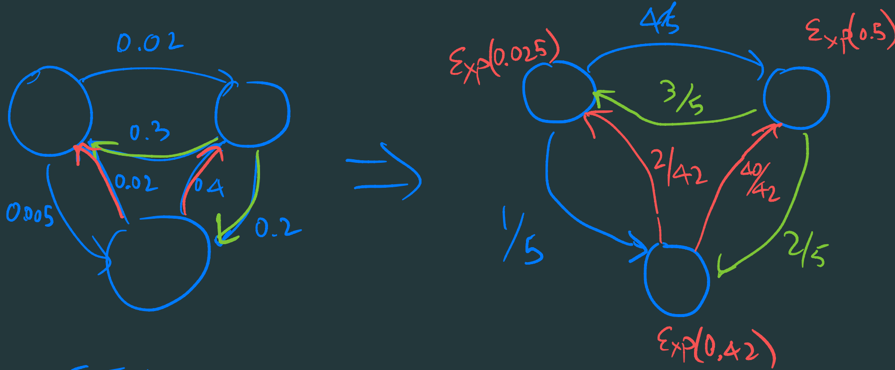
eg. modeling the financial market in continuous time



- directed graph on discrete set of nodes (node = state)
- edge weights ≡ rates of expo timers



holding time in state 'Bull Mkt' ~ Exp(0.025)



CTMC $\Rightarrow$ DTMC

+ holding times

(jump chain)

CTMCs

used as a model for many applications:

– queueing and service systems
– transportation networks
– epidemiology
– communications and computer networks
– agent choice models

advantages

- easy to analyze (in some cases)
- easier to simulate than general discrete-event simulation

problems

- need all inter-event times to be exponentially distributed
- can give spurious insights, hide critical issues

simulating a CTMC

example: queueing

the single-server M/M/1 queue

- number of servers: 1
- capacity: infinite
- service discipline: first-come-first-served (FCFS or FIFO)
- interarrival times: exponential(λ) arrival rate
- service times: exponential(μ) service rate

  (independent interarrival and service times)

- birth death

$$\begin{tikzcd} 0 \arrow[r, bend left, "\lambda"] & 1 \arrow[l, bend left, "\mu"] \arrow[r, bend left, "\lambda"] & 2 \arrow[l, bend left, "\mu"] \arrow[r, bend left, "\lambda"] & 3 \arrow[l, bend left, "\mu"] & \cdots \end{tikzcd}$$

example: simulating an $M/M/1$ queue

- Maintain time T, state $N(T)$, set $N(0) = 0$
- At any state $N(T)$:
  - Advance time to $T + \mathcal{E}xp(\lambda+\mu)$ (if $N(T)>0$)
    (or $T + \mathcal{E}xp(\lambda)$ if $N(T)=0$)
  - Change state to $\begin{cases} N(T)+1 & \text{wp } \lambda/(\lambda+\mu) \\ N(T)-1 & \text{wp } \mu/(\lambda+\mu) \end{cases}$

    (if $N(T) > 0$)

example: epidemics

SIS epidemic (SIR, SEIR, ...)

want to model the spread of the latest influenza strain among the population

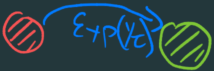
– there is a population of n people

– at any time t, each person i is either susceptible (denoted S or 0) or infected (denoted I or 1)

– infected people get cured on average after time τ, becoming susceptible to future infection. – each pair of people (i, j) independently meet each other with average rate λ

– when a susceptible person meets an infected person, the susceptible person becomes infected

– when two susceptible people or two infected people meet, nothing happens



For each i = I

$I \to S$ trans^n after Exp$(1/\tau)$

For any i = S

Exp$(\lambda I_i(t))$

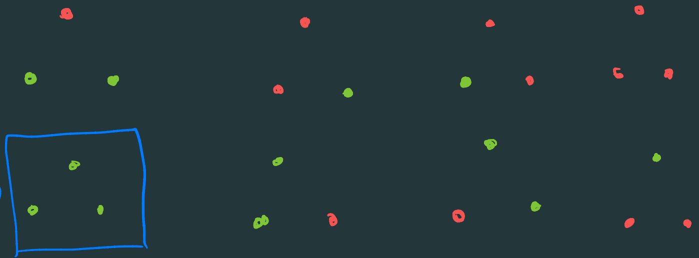
$I_i(t)$ = # of infected neighbors

example: SIS epidemic (contnd)

what assumptions do we need to make this Markovian?

S

absorbing state



$\{S, I\}^n$