ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

Unit 7: Generating Random Processes

Sid Banerjee

School of ORIE, Cornell University

clicker question: quick feedback

how much of the content we have covered till now makes sense to you?

(a) $\leq 25\%$

(b) between 25% and 50%

(c) between 50% and 75%

(d) It all makes sense!

clicker question: quick feedback

how is the pace of the class?

(a) yawn

(b) decent pace with some over-speeding

(d) perfect

(c) generally fast with occasional slow bits

(b) struggling to keep up

review and roadmap

generating random variables

we have seen how to:

- generate pseudorandom $U[0,1]$ samples
- transform $U[0,1]$ samples to another rv using (any 1-d rv with cdf F)
  - inversion
  - acceptance-rejection

two special cases

- generating random vectors $X = (X_1, X_2, \dots, X_d)$
  - multivariate Normal rvs and correlated vectors (copulas)
- generating time-indexed random processes $\{X_t\} = \{X_1, X_2, \dots\}$
  - Exponential rvs and the Poisson process

random processes

random process

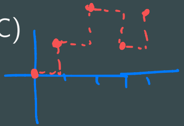
random process

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

- $\mathcal{S}$: state space (Eg - $X_t \in \{1,2,\dots\}$ or $X_t \in \{-2,-1,0,1,2,\dots\}$ or $X_t \in \mathbb{R}$) etc
- $\mathcal{T}$: index set (Eg - $\mathcal{T} = \{1,2,\dots\}$ or $\{-2,-1,0,1,2\,\dots\}$ or $\mathbb{R}_+$ or $\mathbb{R}^2$ etc

four types

- $\mathcal{S}$ discrete, $\mathcal{T}$ discrete: discrete-time Markov chain (DTMC)
  - random walk $T = \{0,1,2,\dots\}$, $X_k \in \mathbb{Z}$



- $\mathcal{S}$ ~~discrete~~ continuous, $\mathcal{T}$ ~~continuous~~ discrete: discrete-time Markov process

  $T = \mathbb{Z}_+$, $X_k \in \mathbb{R}$ Eg - monthly rate of returns

- $\mathcal{S}$ ~~continuous~~ discrete, $\mathcal{T}$ ~~discrete~~ continuous: continuous-time Markov chain (CTMC)
  - Poisson process $T \in \mathbb{R}_+$, $X_k \in \mathbb{Z}$ Eg - # of people in a queue at time t
- $\mathcal{S}$ continuous, $\mathcal{T}$ continuous: Markov process - Advantage - ordering of events is 'easy'
  - Brownian motion $T \in \mathbb{R}_+$, $X_k \in \mathbb{R}$

discrete-time Markov chain

the random walk

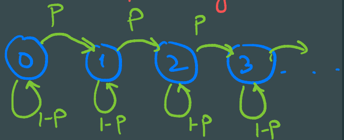
$X_0 = 0$, $Y_k \sim Ber(p)$ iid, and

$$X_t = \sum_{i=0}^{t} Y_k$$

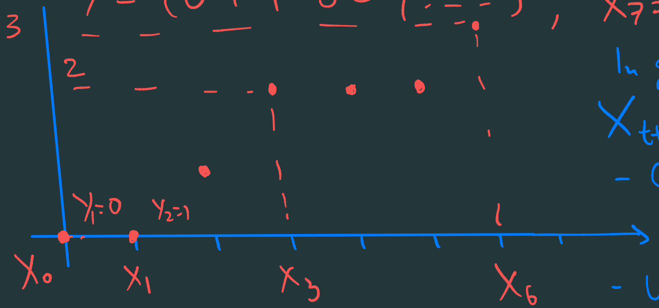
In general $Y_k \in \mathbb{Z}$

(given X_t, X_{t+1} takes discrete values)

transition diagram



$Y = (0\ 1\ 1\ 0\ 0\ 1 \ldots)$, $X_7 = X_6 + Y_7$



In general

$X_{t+1} = X_t + Y_{t+1}$

- Generate $Y_{t+1} \sim Ber(p)$
- Update X_{t+1}

Next - want τ to be $\mathbb{R}_+$

counting process

counting process

non-negative integer-valued stochastic process $[N(t) : t \geq 0]$ "number"

- $N(0) = 0$
- $N(t) = \#$ of arrivals during time interval $(0, t]$

- $[N(t) : t \geq 0]$ increases by jumps
- T_n = time of the n-th arrival, $T_0 = 0$
- A_n = be the interarrival time for the n-th arrival.

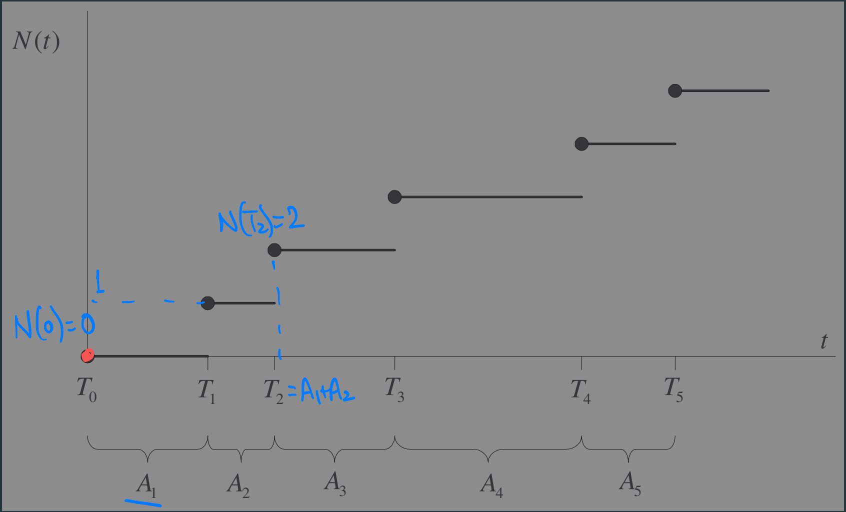
$$A_n = T_n - T_{n-1}$$

$$\Rightarrow T_n = A_1 + A_2 + \dots + A_n$$

$$N(T_n) = n, \quad N(t) = \max\{k \mid T_k \leq t\}$$

- Simulation – 1) Generate A_k
  2) Set $T_k = T_{k-1} + A_k$, $N(T_k) = k$
  3) Go to 1 and repeat

counting process



First 'interarrival' time

desired properties of interarrival times?

- Suppose we know $N(10.1) = 3$, $A_i \sim \text{Unif}[0,5]$, iid
  - When is the next arrival?

$$\mathbb{E}[T_4] = 10.1 + \mathbb{E}[A_4] = 12.6 \; ?$$

What if we learnt $N(4) = 3$ → inconsistent!

  - Need to know T_3 to answer the question
    - even given T_3, $\mathbb{E}[T_4] = 10.1 + \underbrace{\mathbb{E}[A_4 \mid A_4 \geq 10.1 - T_3]}_{\text{complicated}}$

Q: When is it enough to only know $N(t)$

the Exponential distribution

suppose $T \sim Exp(\lambda)$, then:

- pdf: $f_T(t) =$ $\lambda e^{-\lambda t}, \ t \geq 0$
- cdf: $F_T(t) = \mathbb{P}[T \leq t] =$ $1 - e^{-\lambda t}, \ t \geq 0$

$\mathbb{E}[T] = 1/\lambda$

rate $\lambda \geq 0$

why is $Exp(\lambda)$ special?

memorylessness

cdf of T given that T bigger than t?

$\mathbb{P}[T \leq t + x | T > t] =$ $F(x) \ (= 1 - e^{-\lambda x}$ for exponential$)$

$\forall t \geq 0$

Fact – Only 2 dist have this property

- Continuous ≡ Exponential
- Discrete ≡ Geometric

the Exponential distribution: properties

suppose $T_1, T_2, \ldots, T_n$ are all exponentially distributed, with $T_i \sim Exp(\lambda_i)$.

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

  $T_{\min} \sim$ $Exp\left(\sum_{i=1}^{n} \lambda_i\right)$, $T_{2^{nd}\min} = T_{\min} + Exp\left(\sum_{i \neq \min} \lambda_i\right) \ldots$

  'stopwatch'

- (first arrival): let $I_{\min} = \arg\min\{T_i | i \in \{1, 2, \ldots, n\}\}$
  distribution of $I_{\min}$?

  $I_{\min} \sim$ $\left\{ i \quad wp \ \dfrac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \right.$, $i \in \{1, 2, \ldots n\}$

'identity' of the person with minimum T_i

clicker question: parallel simulations

we use 10 GPUs in parallel for simulating a machine learning model

– each GPU finishes one simulation in independent *Exp*(1) time (and then stops)

– we want to get 3 replications to be confident of our model

what is the expected time this will take?

43 (a) $3 \times \frac{1}{10}$

29 (b) $\frac{1}{10} + \frac{1}{9} + \frac{1}{8}$ ✓

27 (c) $\frac{1}{10^3}$

(d) $\frac{1}{30}$

Let $T_i \sim Exp(1) \equiv$ time for machine i

τ_i = Time for i^{th} machine to finish

$\Rightarrow \tau_1 = \min\{T_i, i \in \{1,2,\ldots,10\}\}$

$\Rightarrow \mathbb{E}[\tau_1] = 1/10$

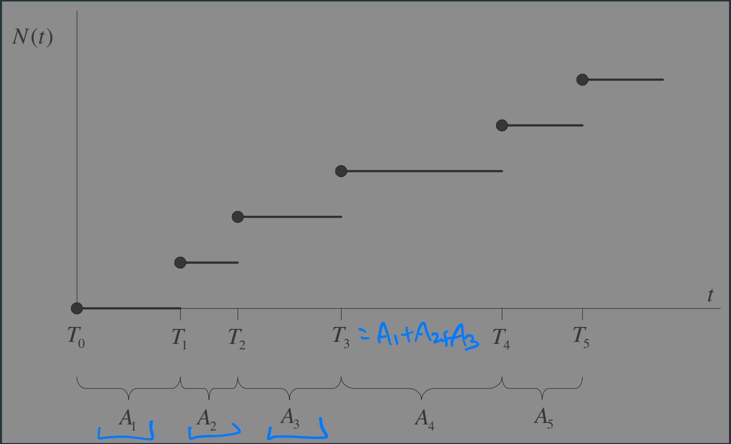
Also $\tau_2 - \tau_1 \equiv$ min of 9 exponentials

$\Rightarrow \mathbb{E}[\tau_2] = \mathbb{E}[\tau_1] + \mathbb{E}[\tau_2 - \tau_1]$

$= 1/10 + 1/9$

Poisson process

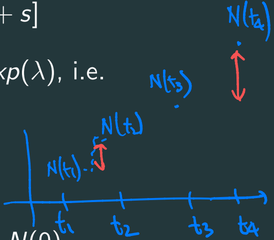
$A_1, A_2, \ldots$ i.i.d. $Exp(\lambda)$ $\implies$ Poisson process of rate λ
– denoted $PP(\lambda)$



iid $Exp(1)$ $Exp(1)$ $Exp(1)$

Poisson process: properties

note: $N(t+s) - N(t) = \#$ arrivals in time interval $(t, t+s]$



- Exponential interrarival times: $A_n = T_n - T_{n-1} \sim Exp(\lambda)$, i.e.

$$\mathbb{P}[A_n \leq t] = 1 - e^{-\lambda t}$$

- independent increments: $N(t+s) - N(t) \perp\!\!\!\perp N(t) - N(0)$

  more generally, for $t_1 \leq t_2 \leq t_3 \leq t_4$, $N(t_4) - N(t_3) \perp\!\!\!\perp N(t_2) - N(t_1)$

mean = interval size × rate

- Poisson arrivals: $N(t+s) - N(t) \sim Poisson(\lambda s)$, i.e.,

$$\mathbb{P}\left[N(t+s) - N(t) = k\right] = \frac{e^{-\lambda s}\,(\lambda s)^k}{k!}$$

moreover, $\mathbb{E}$[Arrivals in any interval of length s] $= \lambda s$.

$Poi(\lambda) \equiv$

$p(k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$

$\forall\, k \in \{0, 1, 2, \ldots\}$

pmf Poisson

Poisson process computations

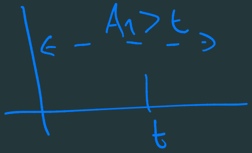
$N(t) \sim PP(\lambda)$

these properties of PP are useful for computations; Eg. $\mathbb{P}$[no arrivals in $[0, t]$]?

– using Exponential interarrival times:

$\mathbb{P}[\mathcal{N}(t) - \mathcal{N}(0) = 0] = \mathbb{P}[A_1 > t]$

$= 1 - F_{Exp(\lambda)}(t) = e^{-\lambda t}$



– using Poisson arrivals:

$\mathbb{P}[\mathcal{N}(t) - \mathcal{N}(0) = 0] = \mathbb{P}[Z = 0]$ where $Z \sim Poi(\lambda t)$

$$= \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

clicker question

cars are pass a speed camera on a highway according to a Poisson process
the chance that at least one car passes by in the next 4 minutes is 0.4
what is the chance that at least one car passes by in the next minute?

32 (a) $1 - 0.6^{\frac{1}{4}}$

45 (b) $1 - 0.6^{4}$

10 (c) 0.44

10 (d) 0.1

2 (e) cannot be determined from the given information

clicker question

cars are pass a speed camera on a highway according to a Poisson process
the chance that at least one car passes by in the next 4 minutes is 0.4
what is the chance that at least one car passes by in the next minute?

$$\mathbb{P}[N(4)=0] = 1 - \mathbb{P}[N(4)>0]$$
$$= 1-0.4 = 0.6$$
$$\Rightarrow e^{-\lambda 4} = 0.6$$

$$\mathbb{P}[N(1)>0] = 1-\mathbb{P}[N(1)=0]$$
$$= 1-e^{-\lambda}$$
$$= 1-(0.6)^{1/4}$$

Poisson process: formal definition

Poisson process $PP(\lambda)$

an arrival process $[N(t) : t \geq 0]$ is a Poisson process with rate λ if

1. $N(t+s) - N(t)$ independent of $N(t) - N(0)$.

2. $\mathbb{P}[N(t+s) - N(t) = 1] = \lambda s + o(s)$, where $o(s)$ denotes a function $g(\cdot)$ satisfying: $\lim_{s \to 0} \frac{g(s)}{s} = 0$.

3. $\mathbb{P}[N(t+s) - N(t) \geq 2] = o(s)$.

intuitively

$N(t+s) - N(t) \sim Ber(\lambda s)$ for s small

example: $P(N(t+s) - N(t) = 1) = \frac{e^{-\lambda s}(\lambda s)^1}{1!}$

$$= \left(1 - \lambda s + \frac{(\lambda s)^2}{2!} - \dots\right) \lambda s$$

$$= \lambda s + \underbrace{s^2(\dots)}$$

$g(s) =$ '$o(s)$', is $\lim_{s \to 0} \frac{g(s)}{s} = 0$

$$\mathbb{P}[N(t+s) - N(t) \geq 2] = 1 - e^{-\lambda s} - \lambda s e^{-\lambda s}$$

$$= \cancel{1} - (\cancel{1} - \cancel{\lambda s} + s^2(\,)) - \lambda \cancel{s}(1 - \lambda s + s^2)$$

$$= s^2(\dots) = o(s)$$

why Poisson process?

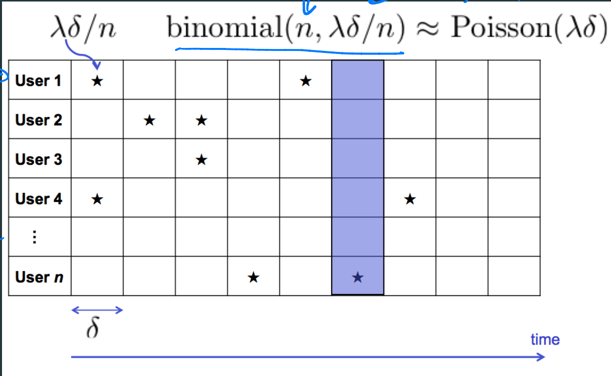
- easy to simulate!
- behavioral justifications: arrivals are modeled as PP because
  - memorylessness of interarrival times

    · knowing T_k does not 'influence' A_{k+1}

  - the Palm-Khintchine theorem

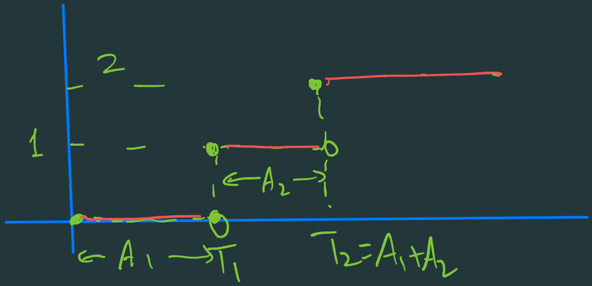
lots of events

small prob of each



generating samples of Poisson processes

1. set the arrival counter $n = 0$. Set $T_0 = 0$.
2. increment n by 1
let A_n be a sample from exponential distribution with parameter λ.
3. advance time
$$T_n = T_{n-1} + A_n.$$
4. return to Step 2.



thinning and superposition

two more important properties of Poisson processes

superposition

let $N_1(t) \sim PP(\lambda_1)$ and $N_2(t) \sim PP(\lambda_2)$ be two independent Poisson processes

then $N_1(t) + N_2(t)$ is a Poisson process with rate $\lambda_1 + \lambda_2$

thinning (random thinning)

given $N(t) \sim PP(\lambda)$, let $N_1(t)$ be the process generated by retaining each arrival of $N(t)$ independently with probability p, and let $N_2(t) = N(t) - N_1(t)$ be the rejected points

then $N_1(t) \sim PP(\lambda p)$ and $N_2(t) \sim PP(\lambda(1-p))$

moreover, $N_1(t) \perp\!\!\!\perp N_2(t)$!

these are very useful for discrete-event simulation!

nonstationary Poisson processes

- imagine that the arrival rate of the Poisson process is not constant, but changes with time.
- $\lambda(t)$ = arrival rate at time t.
- "time of the day" or "seasonality" effects.

Warning – inter arrival times no longer exponential

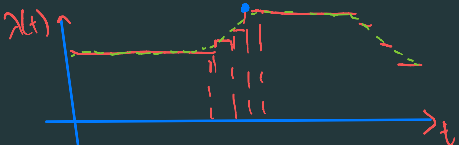
formal definition:

an arrival process $[\mathcal{N}(t) : t \geq 0]$ is called a nonstationary Poisson process with rate function $\lambda(\cdot)$ if

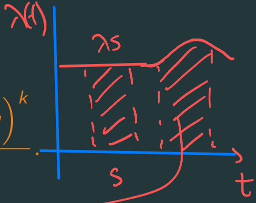
1. $\mathcal{N}(t+s) - \mathcal{N}(t)$ is independent of $\mathcal{N}(t) - \mathcal{N}(0)$.
2. $\mathbb{P}[\mathcal{N}(t+s) - \mathcal{N}(t) = 1] = \lambda(t)s + o(s)$.
3. $\mathbb{P}[\mathcal{N}(t+s) - \mathcal{N}(t) \geq 2] = o(s)$.

for small s

$N(t+s) - N(t) \sim Ber(\lambda(t).s) \sim Ber(\lambda(t+s).s)$



nonstationary Poisson process: properties



- $\mathcal{N}(t+s) - \mathcal{N}(t) \sim \mathit{Poisson}\left(\int_t^{t+s} \lambda(u)du\right)$, that is

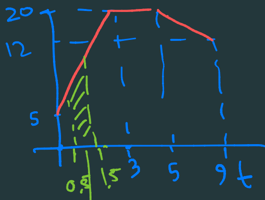
$$\mathbb{P}\left[\mathcal{N}(t+s) - \mathcal{N}(t) = k\right] = \frac{e^{-\int_t^{t+s}\lambda(u)du}\left(\int_t^{t+s}\lambda(u)\ du\right)^k}{k!}.$$

- $\mathbb{E}\Big[\#\text{ of arrivals in interval }(t, t+s]\Big] = \int_t^{t+s}\lambda(u)du.$
- distribution of number of arrivals in $(t, t+s]$ depends on t

example

let $[\mathcal{N}(t) : t \geq 0]$ have arrival rate function

$$\lambda(t) = \begin{cases} 5 + 5t & \text{if } 0 \leq t \leq 3 \\ 20 & \text{if } 3 \leq t \leq 5 \\ 20 - 2(t-5) & \text{if } 5 \leq t \leq 9. \end{cases}$$



- the number of arrivals between $t = 0.5$ and $t = 1.5$ has Poisson distribution with parameter

$$\int_{1/2}^{3/2} (5 + 5t)dt = 10.$$

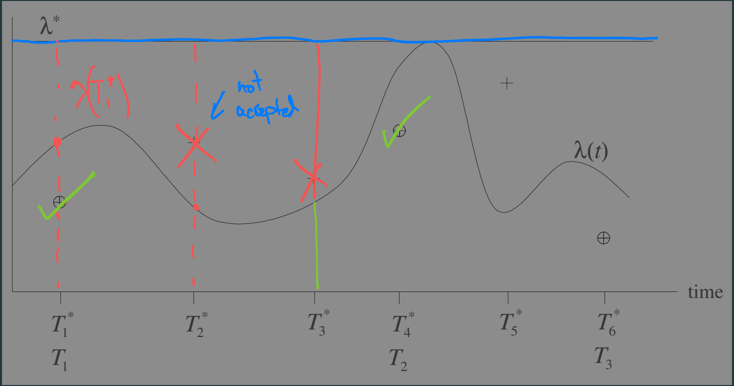
- the probability of having 7 customer arrivals between $t = 0.5$ and $t = 1.5$ is

$$\frac{e^{-10}\, 10^7}{7!}.$$

generating nonstationary Poisson processes via AR (ie thinning)

- $\lambda^* = \max\left[\lambda(t) : t \geq 0\right]$.
- generate a sample of a stationary $PP(\lambda^*)$
- suppose arrival times we obtain are $T_1^*, T_2^*, \ldots$
  accept each arrival time T_i^* with probability $\mathbb{P}[Accept] = \frac{\lambda(T_i^*)}{\lambda^*}$

generate $Z_i \sim U[0, \lambda^*]$,
accept if $Z_i \leq \lambda(T_i^*)$



generating nonstationary Poisson processes via AR

1. set $\lambda^* \geq \max[\lambda(t) : t \geq 0]$
2. set arrival counter $n = 0,\ T^* = 0,\ T_0 = 0$
3. generate $A \sim\ Exp(\lambda^*)$
4. update $\quad T^* = T^* + A.$
5. generate $U \sim U[0,1]$
6. If $U \leq \frac{\lambda(T^*)}{\lambda^*}$,, then increment n by 1 and let $T_n = T^*$
7. return to Step 2

Generate $PP(\lambda^*)$

AR

to show this works, need to verify the 3 properties:

1. $\mathcal{N}(t+s) - \mathcal{N}(t)$ is independent of $\mathcal{N}(t) - \mathcal{N}(0)$] This is true by construction
2. $\mathbb{P}[\mathcal{N}(t+s) - \mathcal{N}(t) = 1] = \lambda(t)s + o(s)$.] see next page
3. $\mathbb{P}[\mathcal{N}(t+s) - \mathcal{N}(t) \geq 2] = o(s)$] This is true because its true for the $PP(\lambda^*)$ process

generating nonstationary Poisson processes via AR

the main thing we need to check is property 2: Let $N^*(t) \sim PP(\lambda^*)$

$$\mathbb{P}[N(t+s)-N(t)=1] = \mathbb{P}[N^*(t+s)-N^*(t)=1, \text{ point accepted}]$$

$$+ \underbrace{\mathbb{P}[N^*(t+s)-N^*(t)>1, \ 1 \text{ point accepted}]}_{=o(s)}$$

$$= \mathbb{P}[N^*(t+s)-N^*(t)=1] \, \mathbb{P}[\text{point accepted} \mid N^*(t+s)-N^*(t)=1] + o(s)$$

$$= \underbrace{\left(\lambda s + o(s)\right)}_{\text{defn of } PP(\lambda^*)} \cdot \left(\frac{\lambda(t)}{\lambda^*}\right) + o(s)$$

combine

$$= \lambda(t) s + o(s)$$