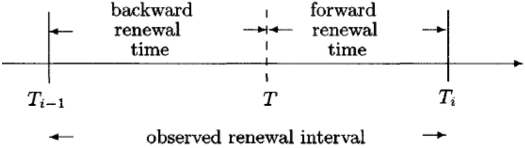
$M/G/1$ queue

Tejas Bodas

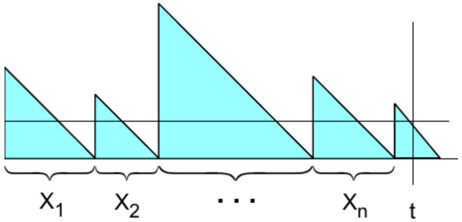
Assistant Professor, IIIT Hyderabad

Age and Residual life of a Renewal process



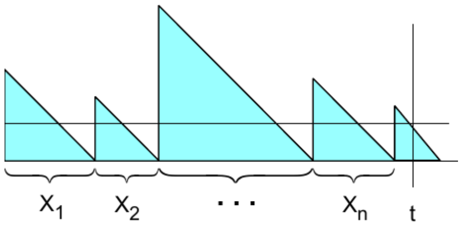
- Let $A(t)$ and $R(t)$ denote the age and the residual life of the renewal process at time t.
- Assume you arrive at a Metro station at time t.
- $A(t)$ is the time since the last metro departed.
- $R(t)$ is the time till the next Metro arrives.
- Assume that you arrive uniformly at random to the Metro.
- What is your average waiting time $\bar{R}$? $\bar{R} = E[X]/2$?

Hitchhiker's Paradox!



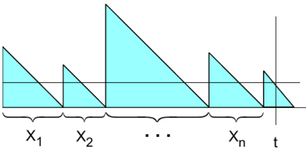
- Consider $\bar{R} = \lim_{t\to\infty} \frac{Y(t)}{t}$ where $Y(t) = \int_0^t R(t)$.
- Using Renewal reward theorem, $\bar{R} = \frac{E[Y]}{E(X)} = \frac{E[X^2]}{2E[X]} \neq E[X]/2$.
- $\frac{E[X^2]}{2E[X]} = E[X]/2$ only when interarrival times are deterministic.
- Consider $\bar{A} = \lim_{t\to\infty} \frac{Y(t)}{t}$ where $Y(t) = \int_0^t A(t)$. $\bar{A}$?.
- What is $\bar{R}$ or $\bar{A}$ when $X_i \sim exp(\lambda)$?

PASTA



- The key assumption to Hitchhikers paradox was that you arrive uniformly at random at the busy/metro stop.
- Now suppose there is a signboard at the metro that tells you the residual time till the next metro.
- Suppose that you note the residual time after every 5 min interval and compute an empirical average of the residual times.
- Will this be equal to $\bar{R}$? No!
- You do not sample $(0, t)$ uniformly.

PASTA

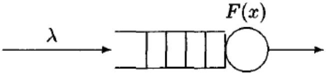


- What is you make the residual time readings after a random time which is $exp(\lambda)$ distributed.
- Your observation process is a Poisson(λ) process.
- In this case, the empirical average will equal $\bar{R}$.

For a Poisson process, given $N(t) = n$, the arrival times $S_1, \ldots S_n$ have the same distribution as the order statistics of n i.i.d uniform points over $(0, t)$. (Thm 2.3.2 Sheldon ross)

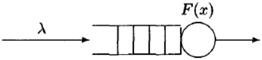
Poisson arrivals see time average! (PASTA)

$M/G/1$ queue



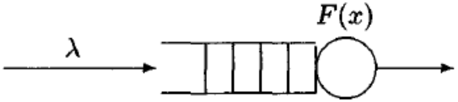
- Arrival process – Poisson (λ)
- Single server with unit service rate
- Arriving jobs require a random service time with arbitrary distribution $F(\cdot)$ with mean b.
- For an $M/M/1$ queue, $F(\cdot) \sim exp(\mu)$ and $b = \frac{1}{\mu}$
- For an $M/G/1$ queue, is $N(t)$ (number of jobs in the system) a Markov chain ? no!

$M/G/1$ queue



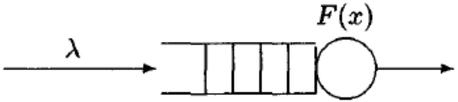
- General service times do not have the memoryless property.
- Let $h(x)$ denote the probability that the job will finish service now given that it has received x units of service already. (a.k.a hazard rate)
- $h(x) = \frac{f(x)}{\bar{F}(x)}$. $h(x) = \mu$ for exponential distribution.
- Therefore the instantaneous rate of going from $N(t) = n$ to $N(t^+) = n - 1$ depends on the age of the job which is in service at time t.
- $N(t)$ is therefore not sufficient to describe the evolution of the state.
- $(N(t), age(t))$ is a valid descriptor for a Markov chain.We do not study this in the course!

$M/G/1$ queue: Notations



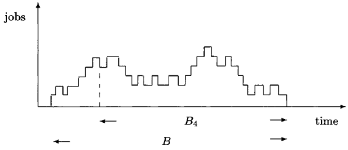
- L denotes the mean number of jobs in the system.
- let l denote the mean number of waiting jobs in the queue.
- $r = L - l$ denotes the mean number of jobs receiving service. This is same as the probability that the server is busy.
- w denotes the mean time spent by any job waiting for service while W denotes the mean sojourn time.
- $W = w + b$.

$M/G/1$ queue: Little's law



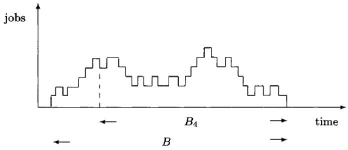
- $L = l + r$ & $W = w + b$.
- Using Little's law we have, $L = \lambda W$.
- Similarly, $l = \lambda w$.
- This gives us $r = \lambda b$.
- The probability that the server is busy is λb.
- Recall that for an $M/M/1$ queue, $1 - \pi(0) = \frac{\lambda}{\mu} = \lambda b$.

Busy period analysis for $M/G/1$



- What is the mean length of busy period, i.e., $E[B]$?
- What is the probability that the server is busy? $(1 - \pi_0 = \frac{\lambda}{\mu})$
- The time average that the server is busy is $\lim_{t\to\infty} \frac{1}{t} \int_0^t 1_{\{N(t)>0\}} dt$.
- Let $Y(t) = \int_0^t 1_{\{N(t)>0\}} dt$. This denotes the time for which the server is busy till time t.
- Using RR theorem, $Y(t)/t$ approaches $\frac{E[B]}{E[B]+\frac{1}{\lambda}}$.

Busy period analysis for $M/G/1$



- Using RR theorem, $Y(t)/t$ approaches $\frac{E[B]}{E[B]+\frac{1}{\lambda}}$
- Equating the two averages give us $E[B] = \frac{b}{1-\rho}$ where $\rho = \lambda b$.
- Mean number of jobs served in a busy period $n_B = E[B]/b = \frac{1}{1-\rho}$.

Mean value formulas for $M/G/1$

- $W = w + b$. $L = l + \rho$.
- Poisson arrival see time average. So consider a tagged arrival.
- For this job $w = w_0 + A$ where w_0 is the residual service time of job in server and A is the delay due to waiting jobs.
- $A = lb = \lambda wb$.
- What is w_0? Use Hitchhikers paradox!
- $w_0 = \rho \frac{M_2}{2b}$ where M_2 is the second moment of distribution F.
- $w = \frac{\lambda M_2}{2} + \lambda wb$.
- $w = \frac{\lambda M_2}{2(1-\rho)}$.
- These formulas are called Pollaczek-Khinchin formulas.
- We used Little's law to obtain mean performance metrics!!

For those interested in Honors/DD

1. Stochastic Optimization
   - Bayesian Optimization (Gaussain processes for ML)
   - Reinforcement learning (Markov Decision Process under unertainty)
   - Multi-arm bandit optmization (UCB, Thompson, Gittins index)
   - Probabilistic Machine learning
2. Operations Research
   - Performance modeling (this course)
   - Pricing (Data driven approaches, estimating distributions)
   - Choice modeling, Assortment Optimization, Recommender Systems
3. Financial Engineering
   - Porfolio Optimization
   - Optimal stopping for Option pricing
   - Brownian motion, Black Sholes formula, Stochastic Differential Equations