Review of Queuing Models

Recitation, Apr. 1st
Guillaume Roels
15.763J Manufacturing System and Supply Chain Design

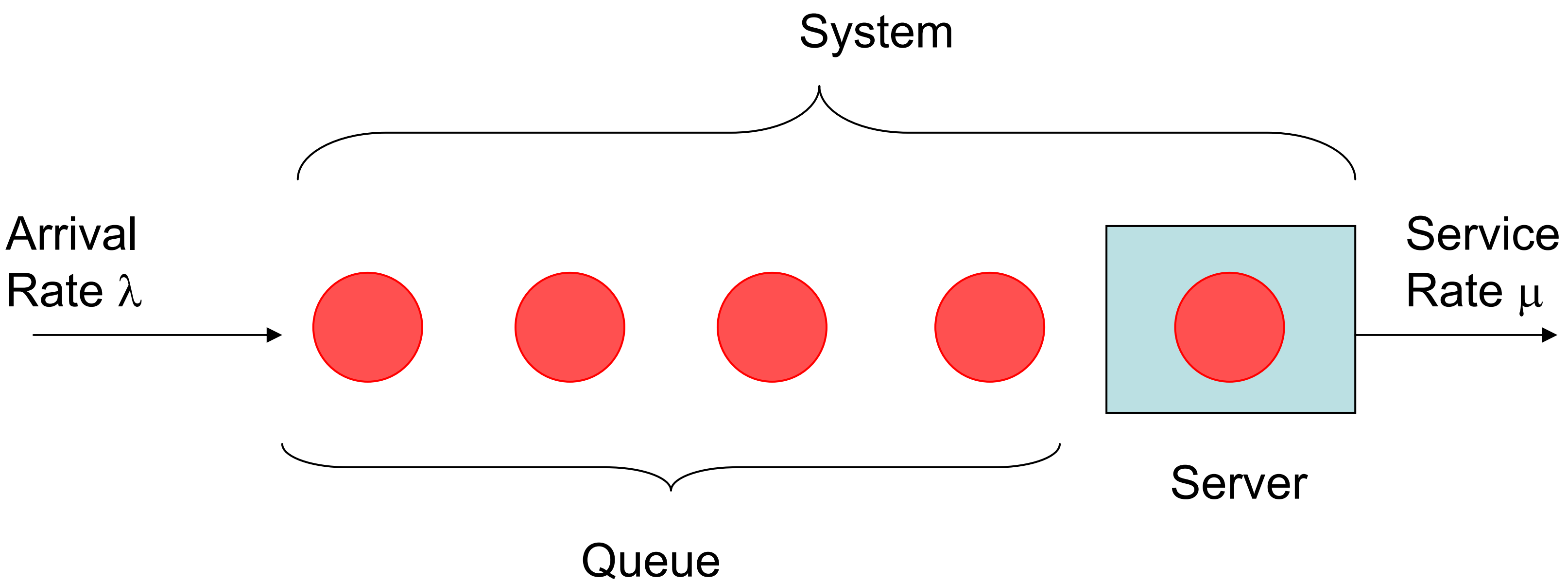
Outline

- Overview, Notations, Little's Law
- Counting Process vs. Interarrival Times
  - Memoryless Process
- Markovian Queues
  - M/M/1
  - M/M/k
- General Queues
  - M/G/1
  - M/G/k
  - M/G/∞
  - M/G/k/k
  - GI/G/k

Queuing Applications

Situation	Customers	Server
Bank	Customers	Tellers
Airport	Airplanes	Runaway
Telephone	Calls	Switches, routers

Queue Representation



L: expected number of people in the system

W: expected time spent in the system

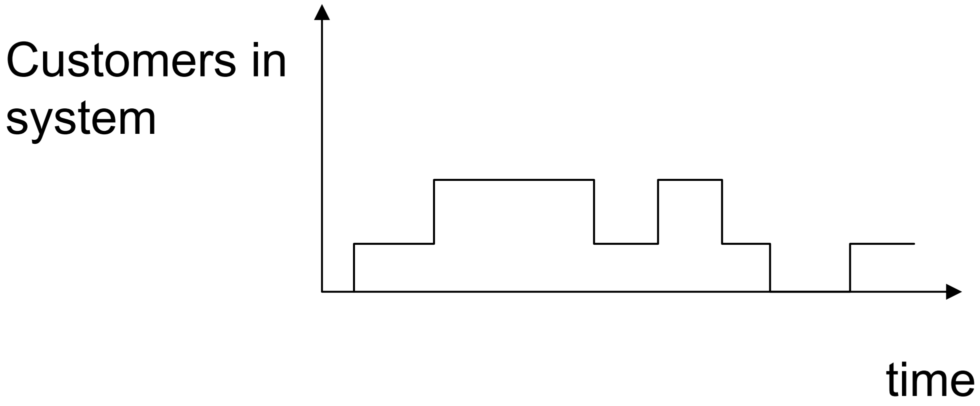
Q: expected number of people in queue

D: expected time spent in queue

Service time/rate

- Service rate: μ (Customers/minute)
- Average service time: $1/\mu$ (Minutes/cust.)

- Service Process is equivalent to Departure Process only if the queue is always nonempty.



Little's Law

- $L=\lambda W$ (system view)

or

- $Q=\lambda D$ (waiting line view)

Also, $W=D+1/\mu$

$1/\mu$ → Service Time

W → Time spent in the system

D → Time spent in the queue

Therefore, compute one quantity (say, L), and get the three others (W, D, Q) for free!

Notations: A/B/m

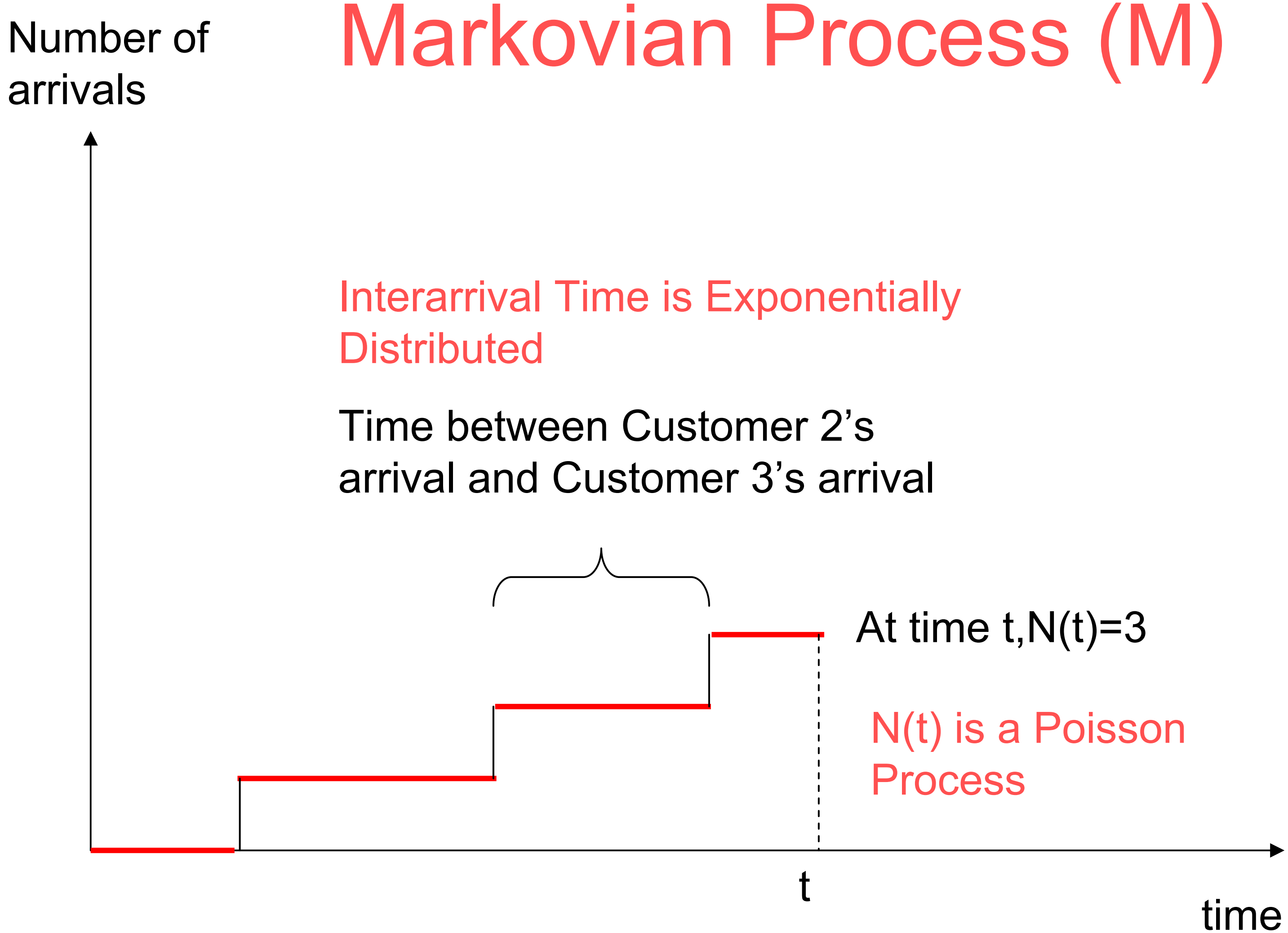
- A: Arrival Process
  - M: Memoryless (or Markovian or Poisson)
  - G: General
- B: Service Process
  - M: Memoryless
  - G: General
- m: Number of servers
- Also: A/B/m/k if system has capacity k

Outline

- Overview, Notations, Little's Law
- Counting Process vs. Interarrival Times
  - Memoryless Process
- Markovian Queues
  - M/M/1
  - M/M/k
- General Queues
  - M/G/1
  - M/G/k
  - M/G/∞
  - M/G/k/k
  - GI/G/k

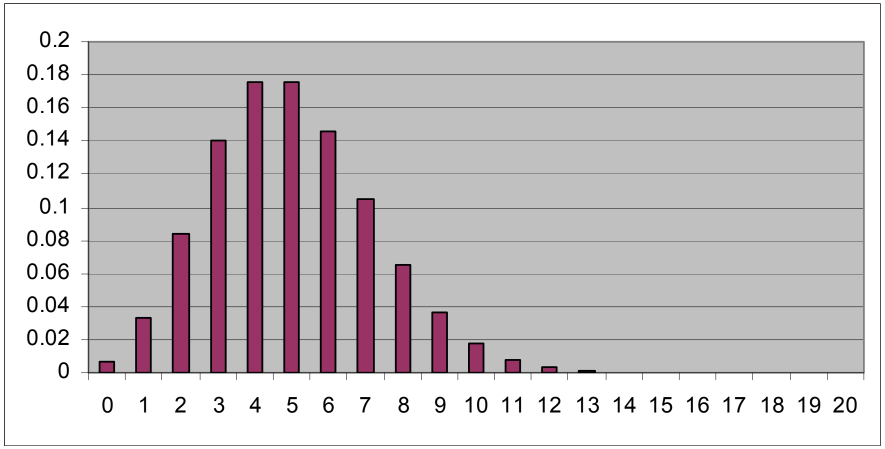
Counting Process vs. Interarrival Times

Markovian Process (M)



Markovian Arrival Process

- Poisson Counting Process (λ=5)



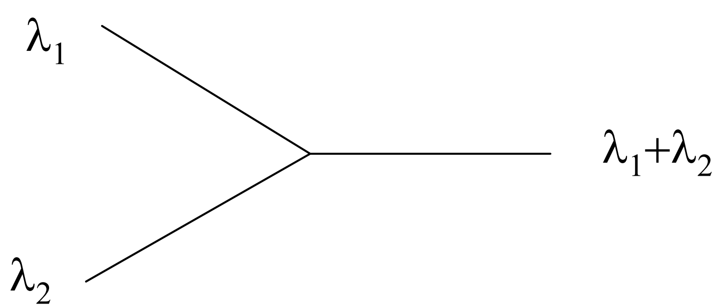
- Counts the number of people that have arrived in a time interval t (Discrete Distribution)
- Memoryless: the number of people who arrive in [t, t+s] is independent of the number of people who have arrived in [0,t]

Markovian Arrival Process

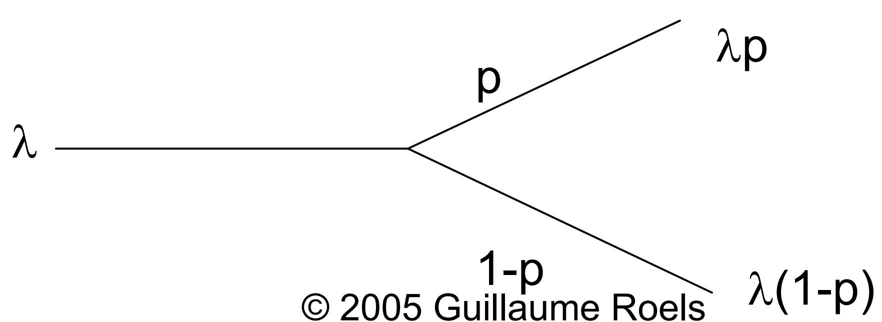
- $P(N(t)=n)=\exp(-\lambda t)(\lambda t)^n/n!$
- $E[N(t)]=\lambda t$; $Var[N(t)]= \lambda t$
- Excel: =POISSON(n,λ,0)
- When $\lambda t>20$, very close to a Normal distribution

Properties of a Poisson Process

- Merging two Poisson processes, with rates λ_1 and λ_2 gives rise to a Poisson process with rare $\lambda_1+\lambda_2$.

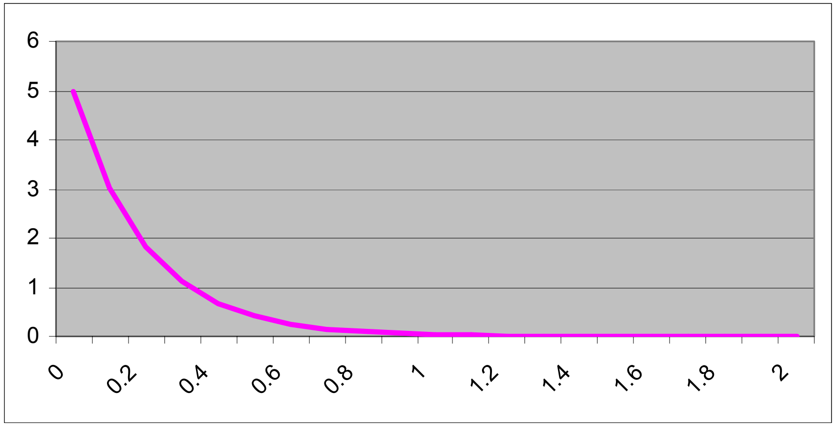


- Randomly splitting a Poisson process with rate λ, according to probabilities p and (1-p), gives rise to two Poisson processes with rates λp and $\lambda(1-p)$.



Markovian Arrival Process

- Exponential Interarrival times (λ=5)



- Time between two arrivals; time between now and the next arrival (Continuous Distribution)
- Memoryless: the time between now and the next arrival is independent of when was the last arrival!

Markovian Arrival Process

- $P(T \leq t) = 1 - \exp(-\lambda t)$; $t > 0$
- $E[T] = 1/\lambda$; $Var[T] = (1/\lambda)^2$
  Coeff. Of Var=1 (highly random)
- Excel: =EXPONDIST(t,λ,1)

Example

The number of glasses of beer ordered per hour at Dick's Pub follows a Poisson distribution, with an average of 30 beers per hour being ordered.

1. Find the probability that exactly 10 beers are ordered between 10 PM and 10:30PM.

   Poisson with parameter (1/2)(30)=15.

   Probability that 10 beers are ordered in 1/2 hour is

$$\frac{e^{-15}15^{10}}{10!} = .048$$

Example from Winston, *Operations Research, Applications and Algorithms* (1993)

Example cont'd

2. Find the mean and standard deviation of the number of beers ordered between 9 PM and 1 AM.

   $\lambda=30$ beers per hour; t=4 hours.

   Mean=4(30)=120 beers

   Standard Deviation=$(120)^{1/2}$=10.95

3. Find the probability that the time between two consecutive orders is between 1 and 3 minutes.

   X=time between successive orders

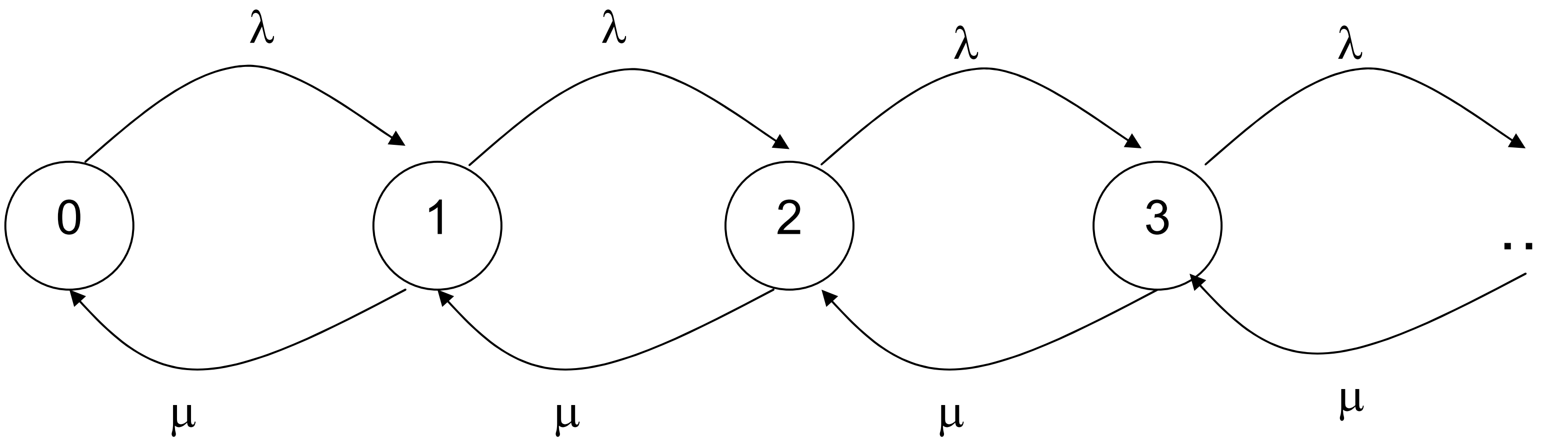
   X is exponential with rate 30/60=0.5 beers/min.

$$P(1 \leq X \leq 3) = \int_1^3 (0.5e^{-0.5t})dt = e^{-0.5} - e^{-1.5} = .38$$

Outline

- Overview, Notations, Little's Law
- Counting Process vs. Interarrival Times
  - Memoryless Process
- Markovian Queues
  - M/M/1
  - M/M/k
- General Queues
  - M/G/1
  - M/G/k
  - M/G/∞
  - M/G/k/k
  - GI/G/k

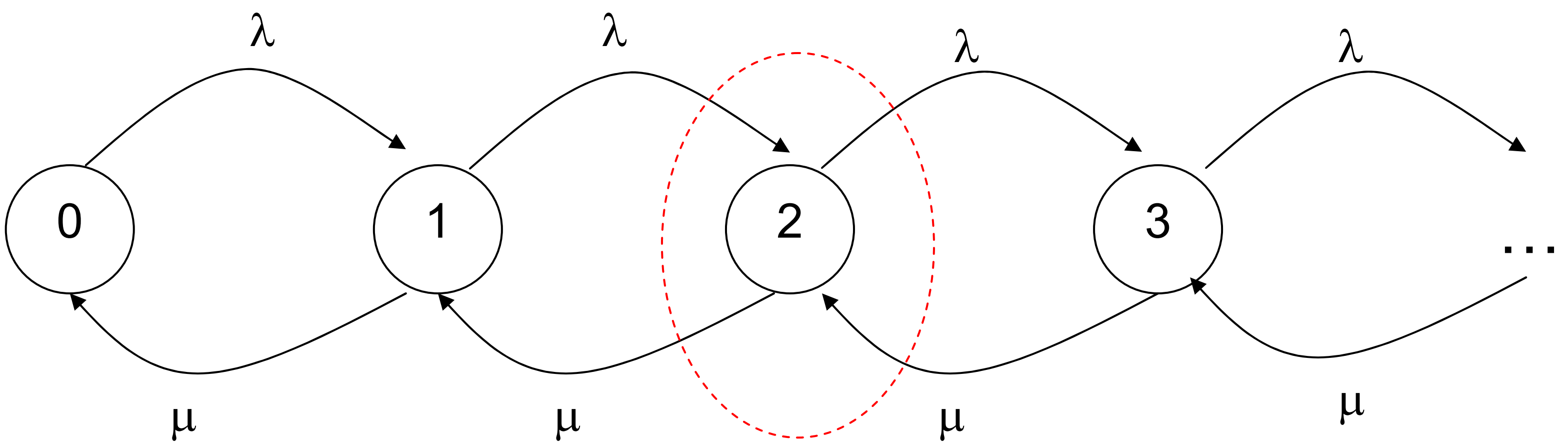
M/M/1 Queue



- Memoryless Queuing System:
- State of the system: number of people in the system
- Utilization Rate $\rho=\lambda/\mu$ (<1)

M/M/1 Balance Equations

- In steady state, the rate of entry into a state must equal the rate of entry out of a state, if $\rho<1$.

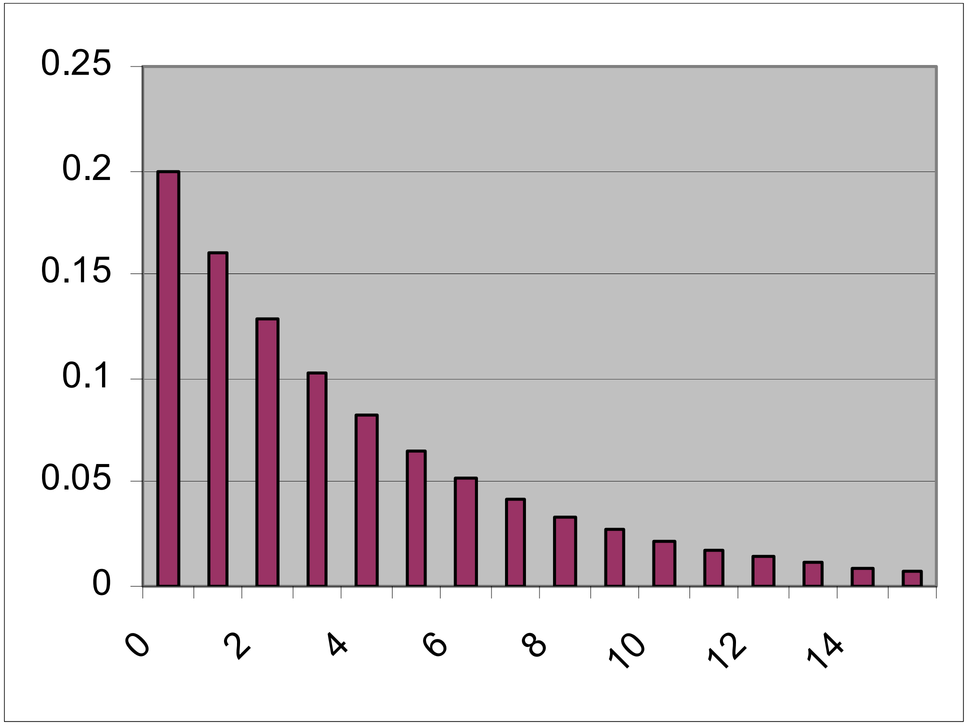


$$\lambda\Pi_1+\mu\Pi_3=(\lambda+\mu)\Pi_2$$

M/M/1: Solving the Balance Equations

$\Pi_i = (\lambda/\mu)^i \, \Pi_0 = \rho^i \, \Pi_0$

$$\sum_{i=0}^{\infty} \Pi_i = 1$$



- Solution

  $\Pi_0 = 1 - \rho$

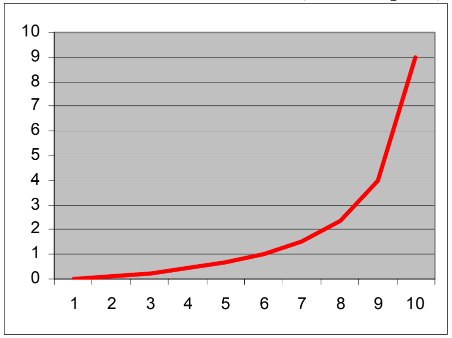
  $\Pi_i = (1-\rho)\rho^i$

  Geometric distribution

Performance Analysis

$$L = \sum_{i=0}^{\infty} i\Pi_i = \frac{\rho}{(1-\rho)}$$

$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)}$$



$$Q = \sum_{i=1}^{\infty} (i-1)\Pi_i = L - (1-P_0) = \frac{\rho^2}{(1-\rho)}$$

$$D = \frac{Q}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)}$$

Example

An average of 10 cars per hour arrive at a single-server drive-in teller. Assume the average service time for each customer is 4 minutes, and both interarrival times and service times are exponential.

M/M/1 with $\lambda=10$ cars/hour and μ=15 cars/hour.

Answer the following questions:

1. What is the probability that the teller is idle?

$\Pi_0=1-\rho=1-2/3=1/3$

Example from Winston, *Operations Research, Applications and Algorithms* (1993)

Example (cont'd)

2. What is the average number of cars waiting in line for the teller? (A car that is being served is not considered to be waiting in line)

$$Q = \frac{\rho^2}{(1-\rho)} = \frac{\left(\frac{2}{3}\right)^2}{1-\frac{2}{3}} = \frac{4}{3} customers$$

3. What is the average amount of time a drive-in customer spends in the bank parking lot (including service time)?

$$W = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{5} hour$$

Example (cont'd)

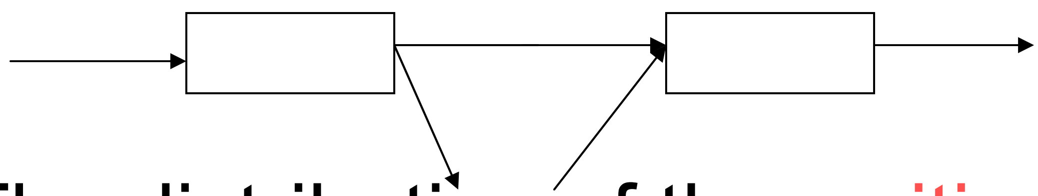
4. On the average, how many customers per hour will be served by the teller?

   If the teller were always busy, he would serve an average of μ=15 cust./hour.

   From (1), we know that he is only busy two-thirds of the time. Thus, during each hour, the teller will serve an average of (2/3) 15 = 10 customers.

M/M/1 Further Analysis

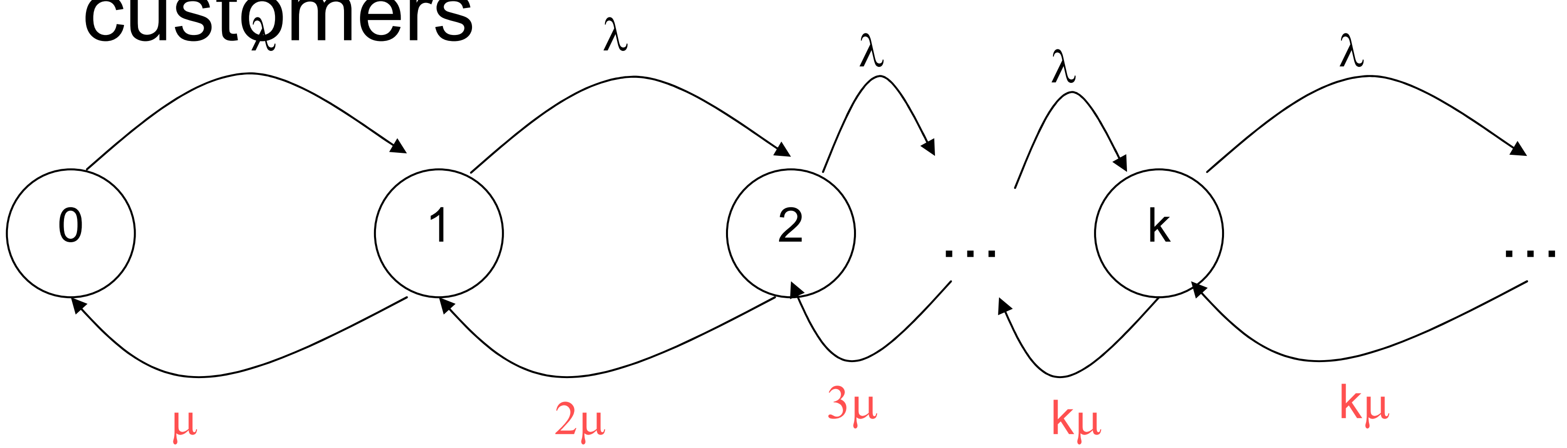
- If $\rho<1$, the departure process of an M/M/1 queuing system is Poisson with rate λ.
  - Same as arrival process!
  - Very useful for series of queues.



- The distribution of the waiting time in the system is exponential with rate $(\mu-\lambda)$.
  - Better measure of service.

M/M/k Queue

- k servers available
- Utilization rate $\rho=\lambda/(k\mu)$
- If less than k customers in the system, use as many servers as the number of customers



M/M/k Steady-State Probabilities

Same kind of derivation as for M/M/1

$$\Pi_0 = \left\{ \sum_{n=0}^{k-1} \frac{(k\rho)^n}{n!} + \frac{(k\rho)^k}{k!} \frac{1}{1-\rho} \right\}^{-1}$$

$$Q = \frac{k^k \rho^{k+1}}{k!(1-\rho)^2} \Pi_0$$

$$D = Q / \lambda \qquad W = D + 1/\mu$$

$$L = Q + k\rho$$

M/M/k Example

Consider a bank with two tellers. An average of 80 customers per hour arrive at the bank and wait in a single line for an idle teller. The average time it takes to serve a customer is 1.2 minutes. Assume that interarrival times and service times are exponential. Determine the expected number of customers present in the bank.

Example from Winston, *Operations Research, Applications and Algorithms* (1993)

M/M/k Example

M/M/2 with λ=80 cust./hour and μ=50 cust./hour. Thus ρ=80/(2(50))=0.80<1.

$$\Pi_0 = \left\{ \sum_{n=0}^{k-1} \frac{(k\rho)^n}{n!} + \frac{(k\rho)^k}{k!} \frac{1}{1-\rho} \right\}^{-1} = \left\{ 1 + 2(0.8) + \frac{(2(0.8))^2}{2} \frac{1}{1-0.8} \right\}^{-1} = \frac{1}{9}$$

$$Q = \frac{k^k \rho^{k+1}}{k!(1-\rho)^2} \Pi_0 = \frac{2^2 (.8)^3}{2!(1-.8)^2} \frac{1}{9} = 2.84 customers$$

$$L = Q + 2(0.8) = 4.44 customers$$

M/M/k model is very useful for capacity planning (try different k's)

Outline

- Overview, Notations, Little's Law
- Counting Process vs. Interarrival Times
  - Memoryless Process
- Markovian Queues
  - M/M/1
  - M/M/k
- General Queues
  - M/G/1
  - M/G/k
  - M/G/∞
  - M/G/k/k
  - GI/G/k

M/G/1 Queue

- Exponential service time often unrealistic because of memoryless property.
- General Service Time T, with mean $E(T)=1/\mu$ and Variance $Var(T)=\sigma^2$ ($\rho=\lambda/\mu$).

$$Q = \frac{\lambda^2\{Var(T)+E(T)^2\}}{2(1-\lambda E(T))} = \frac{\lambda^2\sigma^2+\rho^2}{2(1-\rho)}$$

Not an approximation!

Observe that Q, L, D, and W increase with σ^2.

M/G/k Queue -- Approximation

SCV_s=squared coefficient of variation for service times

Q≈(Expected waiting time for M/M/k) [(1+SCV_s)/2]

M/G/∞ Queue

- Infinite number of servers; hence, D=Q=0.
- L has a steady-state Poisson distribution, with rate λ/μ.
- Applications of M/G/∞:
  – (S-1,S) inventory control; order one item as soon as you sell one.
  – Number of firms in an industry

M/G/k/k Queue

General service time with mean τ.

No waiting space. All potential customers that arrive when all k servers are busy depart the system. Blocked customers are cleared.

Steady-state distribution of customers in system:

$$P(L = n) = \frac{\frac{(\lambda\tau)^n}{n!}}{\sum_{i=0}^{k} \frac{(\lambda\tau)^i}{i!}}$$

M/G/k/k Queue: Loss Probability

- Loss Probability = P(L=k)

  Probability that all servers are busy. The rate of customers that observe this state of the system, λ P(L=k), will balk.

- If small loss probability, good approximation with $M/G/\infty$.

Example

An average of 20 ambulance calls per hour are received by Gotham City Hospital. An ambulance requires an average of 20 minutes to pick up a patient and take the patient to the hospital. The ambulance is then available to pick up another patient.

How many ambulances should the hospital have to ensure that there is at most 1% probability of not being able to respond immediately to an ambulance call? Assume that interarrival times are exponentially distributed.

Example cont'd

M/G/k/k model with λ=20 and τ=1/3.

Consider k=13.

$$P(L=13)=\frac{\frac{(20/3)^{13}}{13!}}{\sum_{i=0}^{13}\frac{(20/3)^{i}}{i!}}=0.01059$$

Consider k=14. P(L=14)=.005019

GI/G/k Queue -- Approximation

General (and independent) arrival process, general service time distribution. Assume $\rho<1$.

SCV_a=squared coefficient of variation for interarrival times

SCV_s=squared coefficient of variation for service times

Utilization Rate   Scale   Variability

$$D \approx \frac{\rho}{1-\rho}\frac{1}{\mu}\frac{SCV_a + SCV_s}{2}$$

Q=λ D, W=D+1/μ, L=λ W

GI/G/k Network of Queues

The departure process from one queue is the arrival process to the next one.

Approximate the SCV of the departure process as:

$SCV_d = (1-\rho^2)\ SCV_a + \rho^2\ SCV_s$

Conclusions

- Various models
  - Closed-form solutions only for simplistic models
  - If complex models, use
    - Approximations
    - Simulation
- Descriptive models
  - Building block for optimization models: size of waiting room, capacity, comparison of technology…