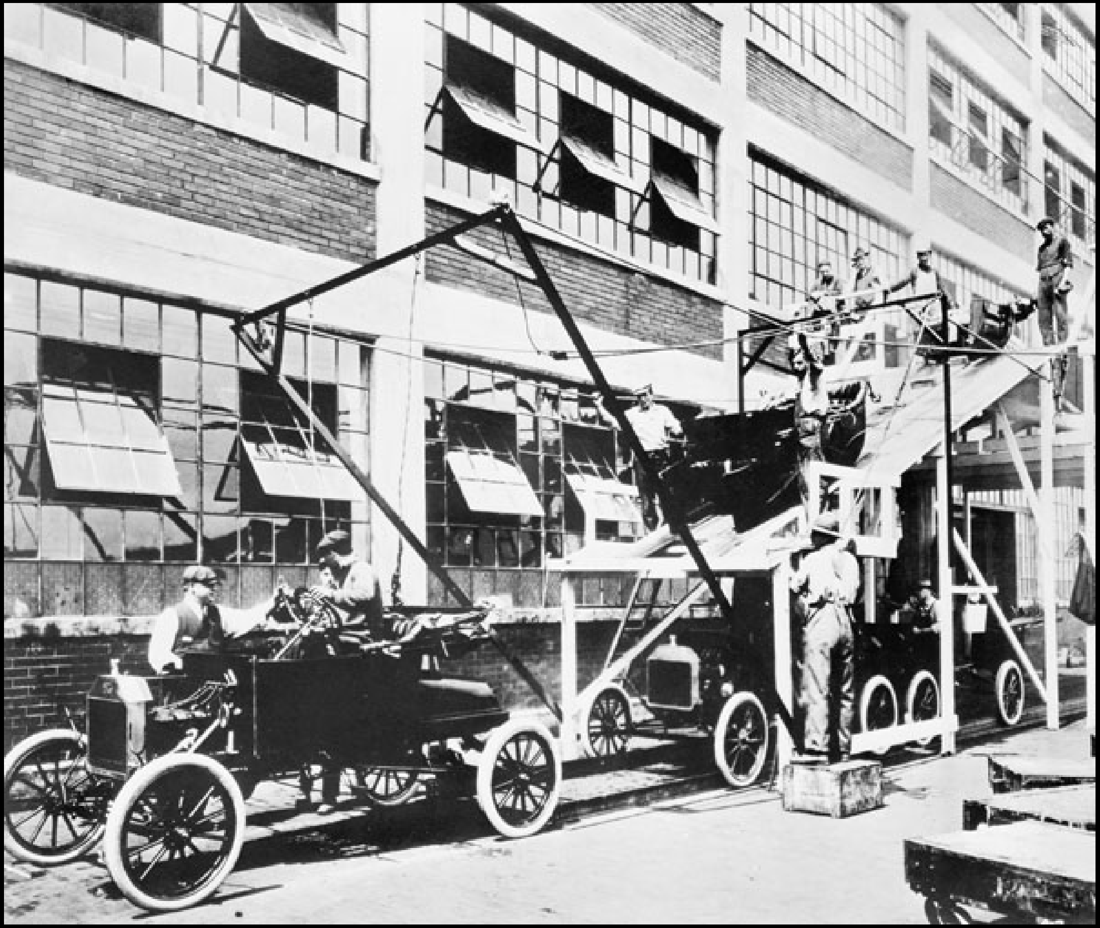
ESD.86. Markov Processes and their Application to Queueing II



Richard C. Larson

March 7, 2007

Photo: US National Archives

Outline

- Little's Law, one more time
- PASTA treat
- Markov Birth and Death Queueing Systems

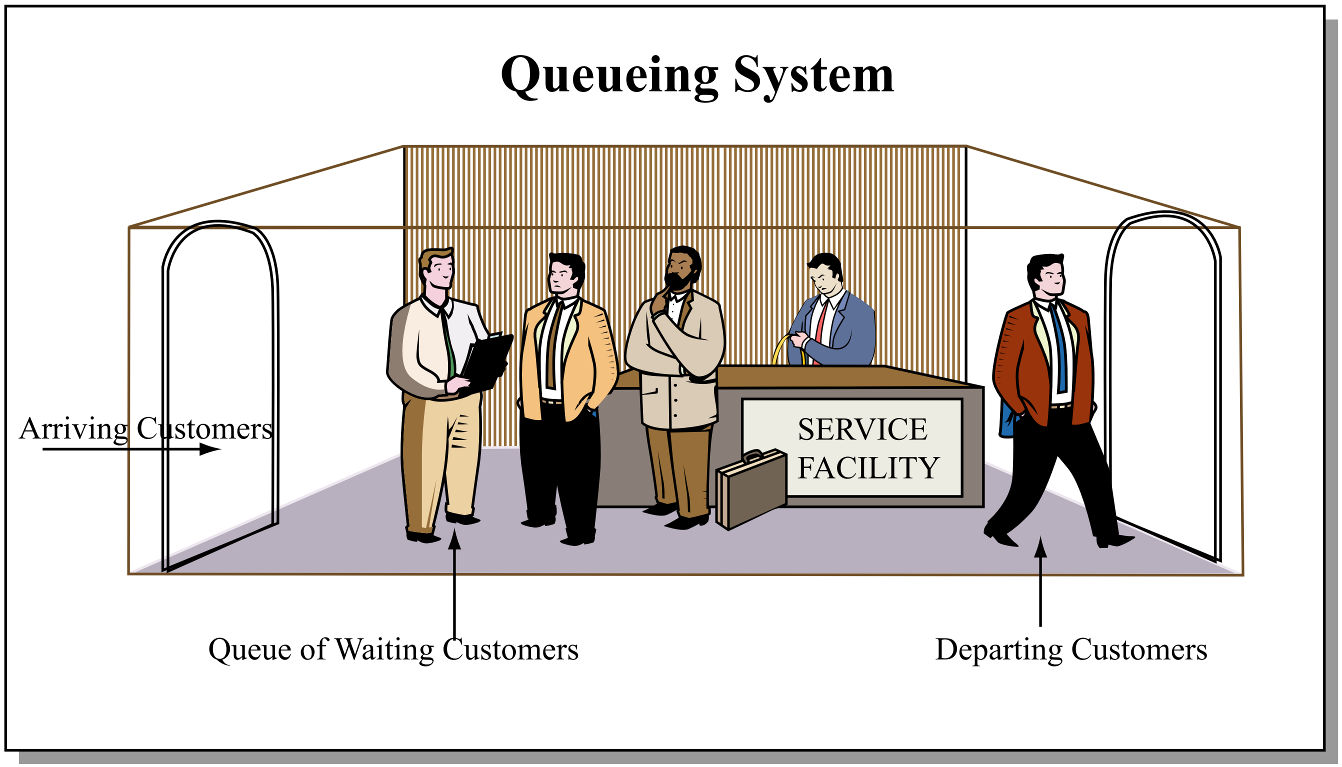
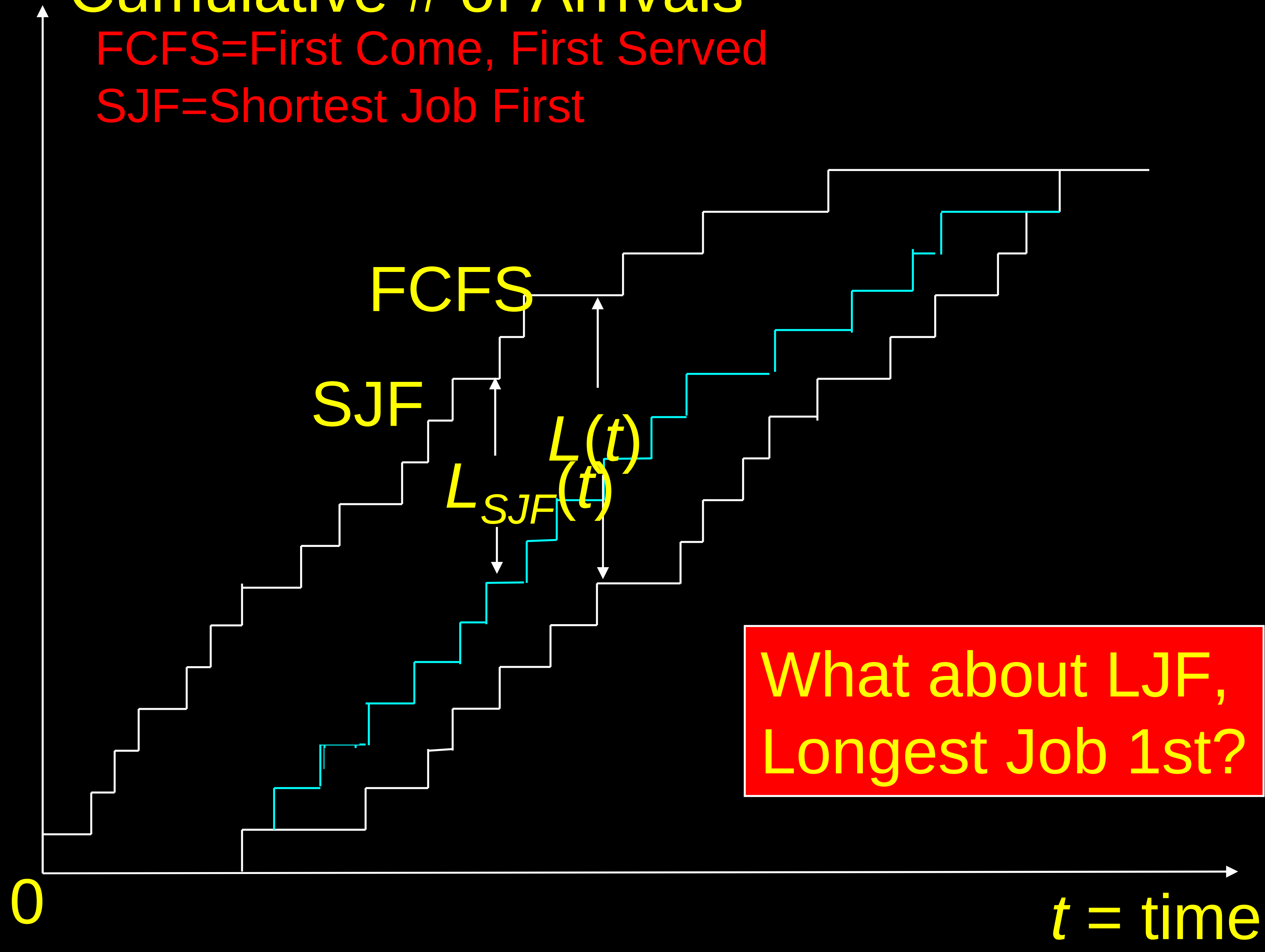


Figure by MIT OCW.

Cumulative # of Arrivals

FCFS=First Come, First Served

SJF=Shortest Job First

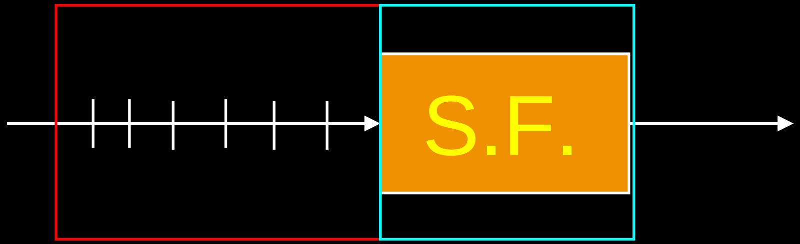


What about LJF, Longest Job 1st?

"System" is General

$$L = \lambda W$$

- *Our results apply to entire queue system, queue plus service facility*
- *But they could apply to queue only!*



$$L_q = \lambda W_q$$

- *Or to service facility only!*

$$L_{SF} = \lambda W_{SF} = \lambda / \mu$$

$$1/\mu = \text{mean service time}$$

All of this means,

"You buy one, you get the other 3 for free!"

$$W = \frac{1}{\mu} + W_q$$

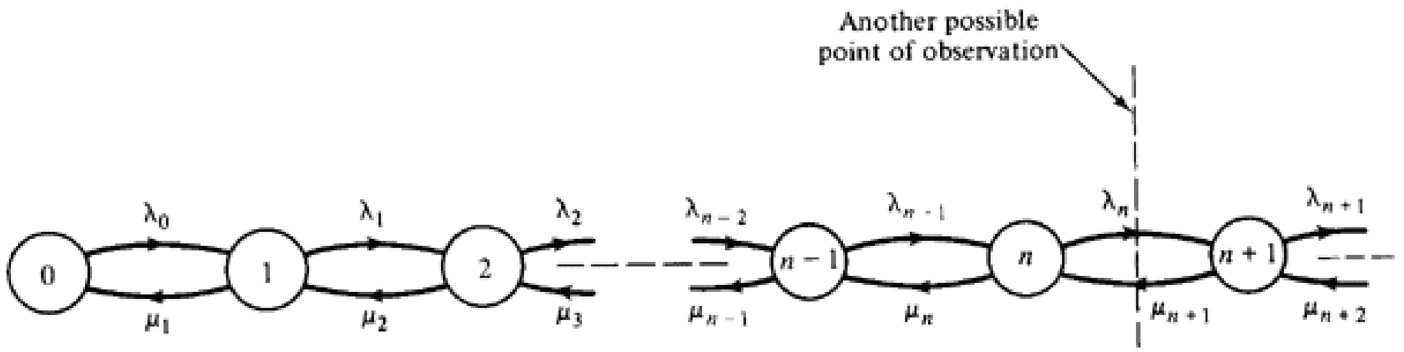
$$L = L_q + L_{SF} = L_q + \frac{\lambda}{\mu}$$

$$L = \lambda W$$

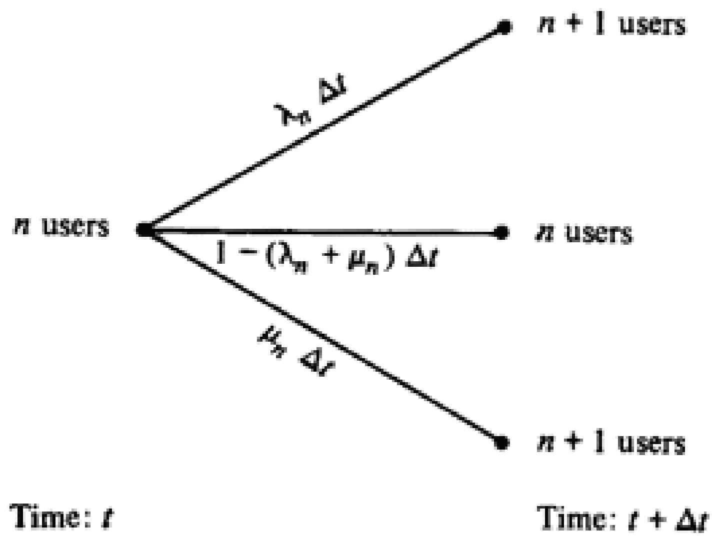
Markov Queues

Markov here means, “No Memory”

Source: Larson and Odoni, *Urban Operations Research*



State-transition diagram for the fundamental birth-and-death model.



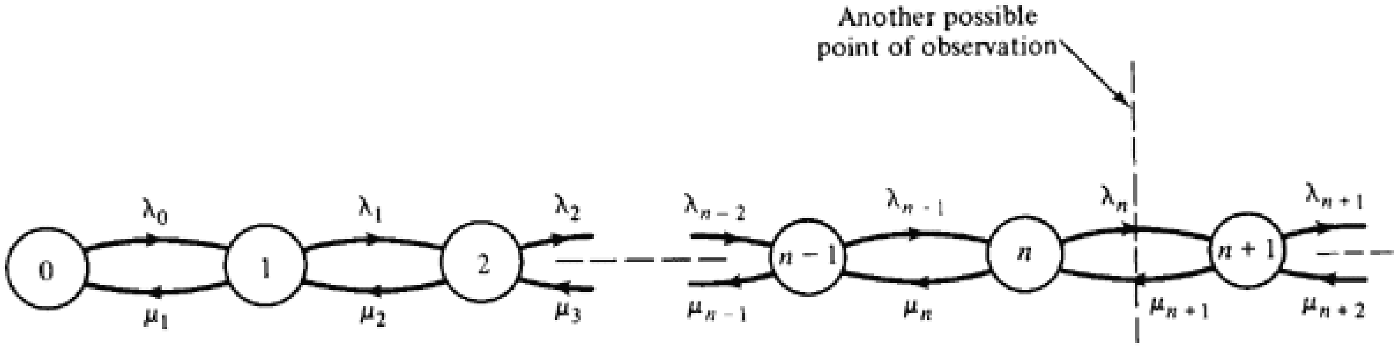
Probabilities of transitions for birth-and-death model in time Δt.

Balance of Flow Equations

$$\lambda_0 P_0 = \mu_1 P_1$$

$$(\lambda_n + \mu_n)P_n = \lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} \text{ for } n = 1,2,3,\ldots$$

Another way to balance the flow:



State-transition diagram for the fundamental birth-and-death model.

Source: Larson and Odoni, *Urban Operations Research*

$$\lambda_n P_n = \mu_{n+1}P_{n+1} \quad n = 0,1,2,\ldots$$

$$\lambda_0 P_0 = \mu_1 P_1$$

$$\lambda_1 P_1 = \mu_2 P_2 \ \ldots$$

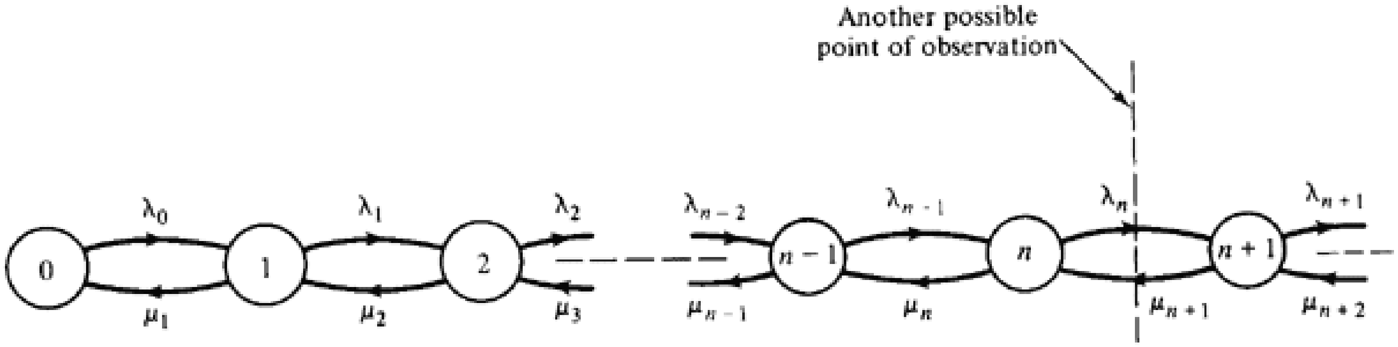
$$\lambda_n P_n = \mu_{n+1} P_{n+1}$$

$$P_1 = (\lambda_0 / \mu_1) P_0$$

$$P_2 = (\lambda_1 / \mu_2) P_1 = (\lambda_0 / \mu_1)(\lambda_1 / \mu_2) P_0 = (\lambda_0 \lambda_1 / [\mu_1 \mu_2]) P_0$$

$$P_{n+1} = (\lambda_n / \mu_{n+1}) P_n = (\lambda_0 \lambda_1 \ldots \lambda_n / [\mu_1 \mu_2 \ldots \mu_{n+1}]) P_0$$

Telescoping!



State-transition diagram for the fundamental birth-and-death model.

Source: Larson and Odoni, *Urban Operations Research*

$$\lambda_n P_n = \mu_{n+1} P_{n+1} \quad n = 0, 1, 2, \ldots$$

$$\lambda_0 P_0 = \mu_1 P_1$$

$$\lambda_1 P_1 = \mu_2 P_2 \ \ldots$$

$$\lambda_n P_n = \mu_{n+1} P_{n+1}$$

$$P_1 = (\lambda_0 / \mu_1) P_0$$

$$P_2 = (\lambda_1 / \mu_2) P_1 = (\lambda_0 / \mu_1)(\lambda_1 / \mu_2) P_0 = (\lambda_0 \lambda_1 / [\mu_1 \mu_2]) P_0$$

$$P_{n+1} = (\lambda_n / \mu_{n+1}) P_n = (\lambda_0 \lambda_1 \ldots \lambda_n / [\mu_1 \mu_2 \ldots \mu_{n+1}]) P_0$$

Telescoping!

$$P_0 + P_1 + P_2 + \ldots = \sum_{n=0}^{\infty} P_n = 1$$

$$P_0 + (\lambda_0 / \mu_1) P_0 + (\lambda_0 \lambda_1 / [\mu_1 \mu_2]) P_0 + \ldots + (\lambda_0 \lambda_1 \ldots \lambda_n / [\mu_1 \mu_2 \ldots \mu_{n+1}]) P_0 + \ldots = 1$$

$$P_0 \{1 + (\lambda_0 / \mu_1) + (\lambda_0 \lambda_1 / [\mu_1 \mu_2]) + \ldots + (\lambda_0 \lambda_1 \ldots \lambda_n / [\mu_1 \mu_2 \ldots \mu_{n+1}]) + \ldots\} = 1$$

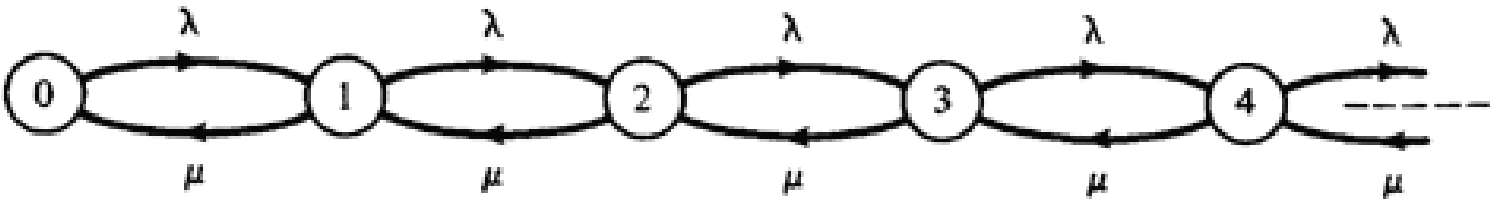
Now, you easily solve for P_0 and then for
All other P_n's.

PASTA: Poisson Arrivals See Time Averages

Time to Buckle your Seatbelts!

http://www.census.gov/pubinfo/www/multimedia/img/seatbelt-lo.jpg

The M/M/1 Queue



State-transition diagram for a *M/M/1* queueing system with infinite system capacity.

Source: Larson and Odoni, *Urban Operations Research*

$$P_0 + (\lambda_0/\mu_1)P_0 + (\lambda_0\lambda_1/[\mu_1\mu_2])P_0 + \ldots + (\lambda_0\lambda_1\ldots\lambda_n/[\mu_1\mu_2\ldots\mu_{n+1}])P_0 + \ldots = 1$$

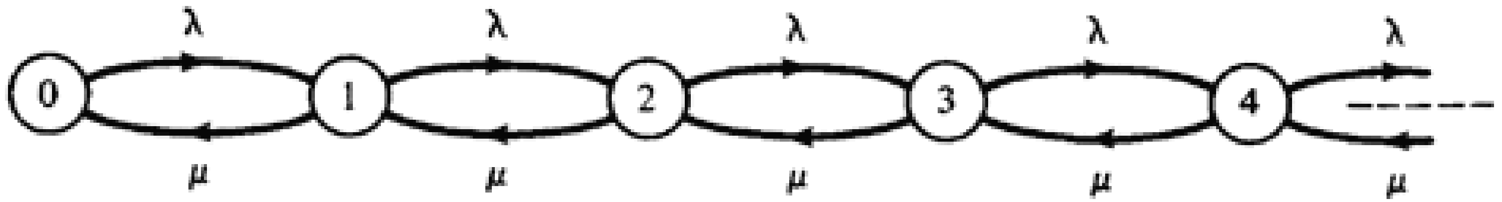
$$P_0\{1 + (\lambda_0/\mu_1) + (\lambda_0\lambda_1/[\mu_1\mu_2]) + \ldots + (\lambda_0\lambda_1\ldots\lambda_n/[\mu_1\mu_2\ldots\mu_{n+1}]) + \ldots\} = 1$$

$$P_0\{1 + (\lambda/\mu) + (\lambda^2/\mu^2) + \ldots + (\lambda^{n+1}/\mu^{n+1}) + \ldots\} = 1$$

$$\{1 + (\lambda/\mu) + (\lambda^2/\mu^2) + \ldots + (\lambda^{n+1}/\mu^{n+1}) + \ldots\} = 1/[1 - (\lambda/\mu)]$$

For $\lambda/\mu < 1$.

The M/M/1 Queue



State-transition diagram for a *M/M/1* queueing system with infinite system capacity.

Source: Larson and Odoni, *Urban Operations Research*

$$P_0 = 1 - \lambda/\mu \quad \text{for } \lambda/\mu < 1.$$

$$P_n = (\lambda/\mu)^n P_0 = (\lambda/\mu)^n (1 - \lambda/\mu) \text{ for } n = 1,2,3,\ldots$$

$$P_0\{1 + (\lambda/\mu) + (\lambda^2/\mu^2) + \ldots + (\lambda^{n+1}/\mu^{n+1}) + \ldots\} = 1$$

$$\{1 + (\lambda/\mu) + (\lambda^2/\mu^2) + \ldots + (\lambda^{n+1}/\mu^{n+1}) + \ldots\} = 1/[1 - (\lambda/\mu)]$$

For $\lambda/\mu < 1$.

The M/M/1 Queue

$$P^T(z) \equiv \sum_{n=0}^{\infty} P_n z^n = \sum_{n=0}^{\infty} (\lambda/\mu)^n (1-\lambda/\mu) z^n = \frac{1-\rho}{1-\rho z}$$

$$\frac{d}{dz} P^T(z) \Big]_{z=1} \equiv \sum_{n=0}^{\infty} n P_n = L = \frac{-(1-\rho)(-\rho)}{(1-\rho z)^2} \Bigg] = \frac{\rho}{1-\rho} \text{ for } \rho < 1$$

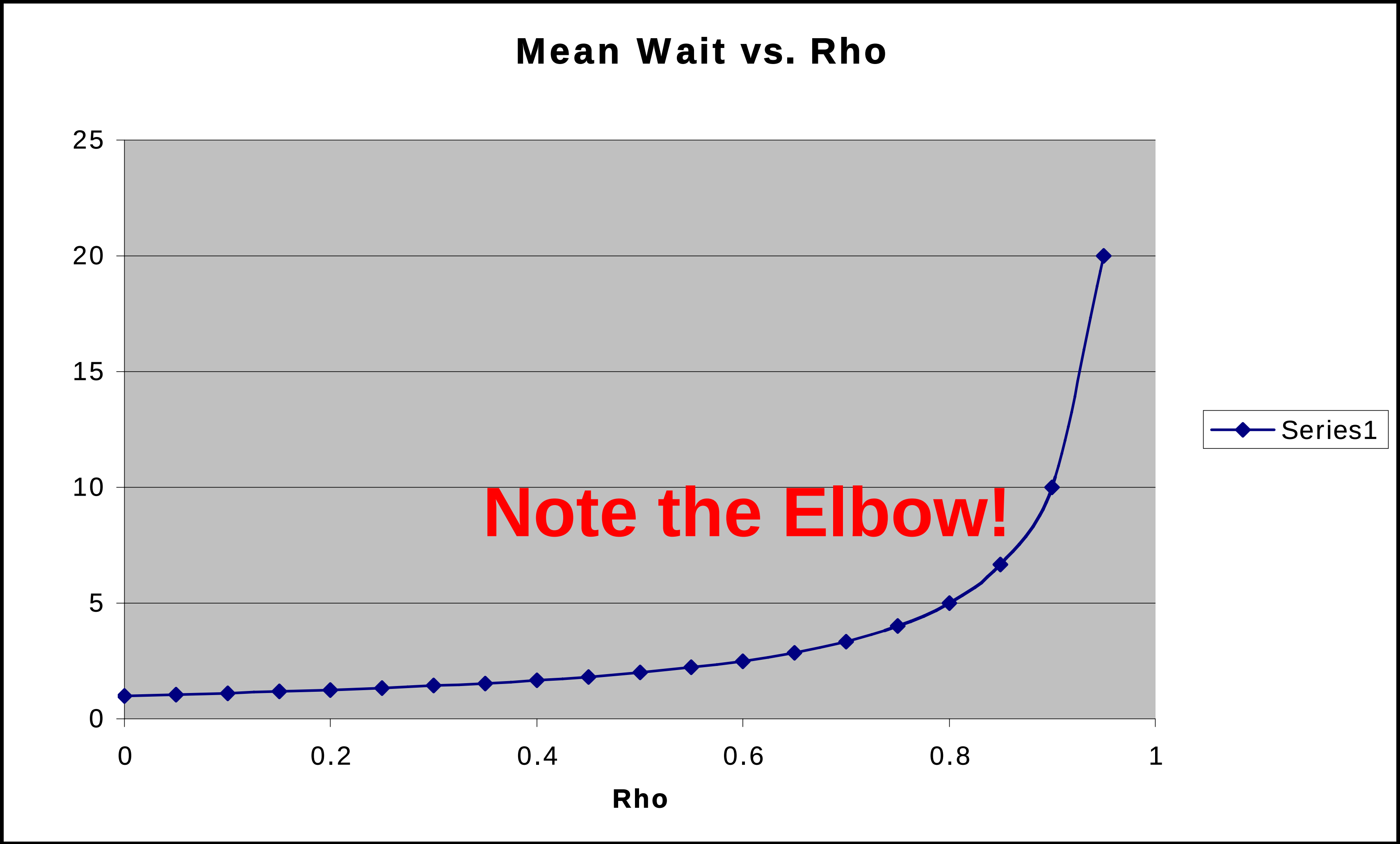
$$P_0 = 1 - \lambda/\mu \quad \text{for } \lambda/\mu < 1.$$

$$P_n = (\lambda/\mu)^n P_0 = (\lambda/\mu)^n (1-\lambda/\mu) \text{ for } n = 1, 2, 3, \ldots$$

$$L = \lambda W = \rho/(1-\rho)$$

implies $W = (1/\lambda)\rho/(1-\rho) = (1/\mu)/(1-\rho)$

$L_q = \lambda W_q$ etc.


Mean Wait vs. Rho
25
20
15
10
5
0
0
0.2
0.4
0.6
0.8
1
Rho
Series1
Note the Elbow!

More on M/M/1 Queue

Let $w(t)$ = pdf for time in the system (including queue and service)

Assume First-Come, First-Served (FCFS) Queue Discipline

$$w(t) = \sum_{k=0}^{\infty} w(t \mid k) P_k = \sum_{k=0}^{\infty} \frac{\mu^{k+1} t^k e^{-\mu t}}{k!} \rho^k (1-\rho)$$

Exercise: Do the same for Time in queue

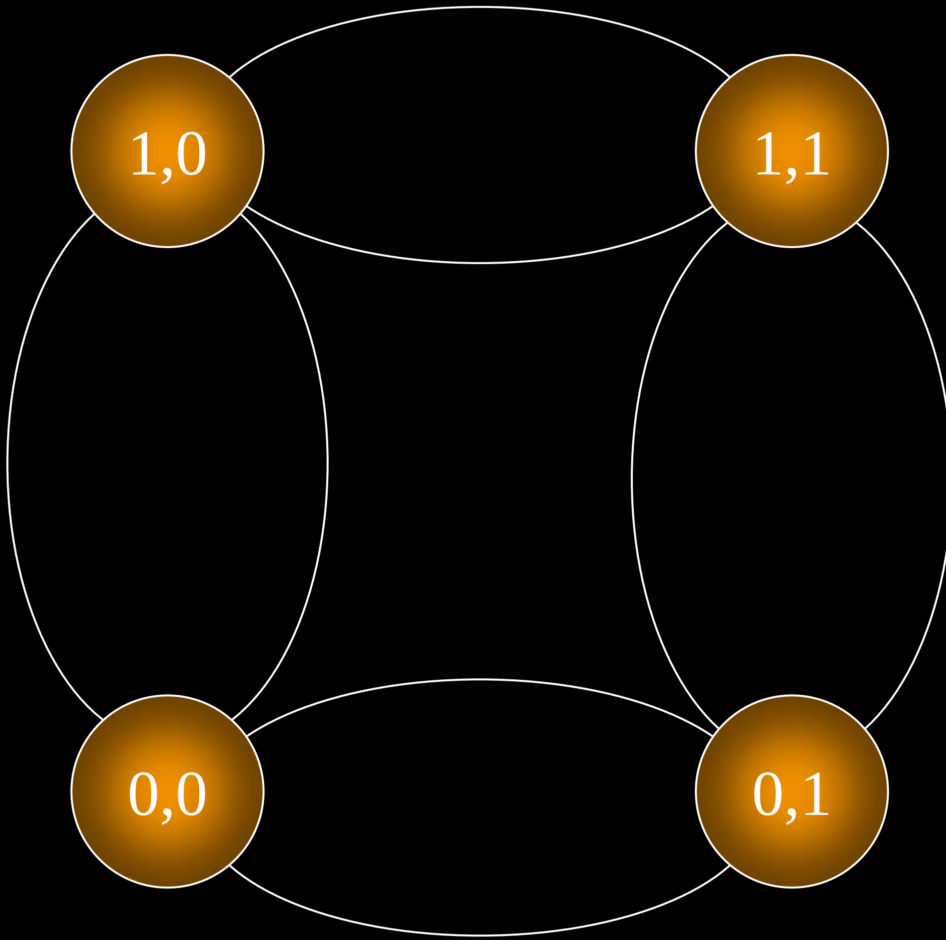
$$w(t) = \mu e^{-\mu t}(1-\rho) \sum_{k=0} \frac{(\mu t \rho)}{k!} = \mu(1-\rho) e^{-\mu t} e^{-\mu \rho t}$$

$$w(t) = \mu(1-\rho) e^{-\mu(1-\rho)t} \quad t \geq 0$$

Blackboard Modeling

- 3 server zero line capacity
- 3 server capacity for 4 in queue
- Same as above, but 50% of queuers balk due to having to wait in queue
- Single server who slows down to half service rate when nobody is in queue
- More?? ….

About the ‘cut’’ between states to write the balance of flow equations…



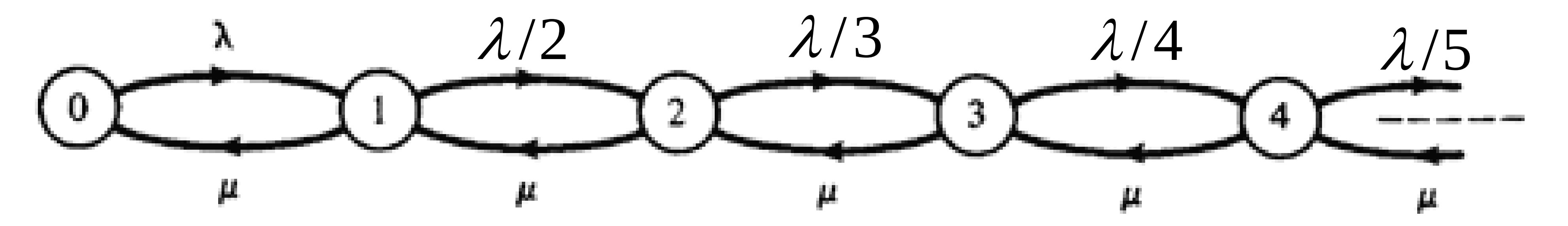
Optional Exercise:

Is it ‘’better’’ to enter a single server queue with service rate μ or a 2-server queue each with rate $\mu/2$?

Can someone draw one or both of the state-rate-transition diagrams?

Then what do you do?

Final Example:
Single Server, Discouraged Arrivals



State-Rate-Transition Diagram, Discouraged Arrivals

$$P_k = \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k P_0$$

$$P_0 = \left[1 + \left(\frac{\lambda}{\mu}\right) + \frac{1}{2!}\left(\frac{\lambda}{\mu}\right)^2 + \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3 + \ldots + \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k + \ldots\right]^{-1}$$

$$P_0 = \left(e^{\lambda/\mu}\right)^{-1} = e^{-\lambda/\mu}$$

$$P_0 = (e^{\lambda/\mu})^{-1} = e^{-\lambda/\mu} > 0$$

$$\rho = \text{utilization factor} = 1 - P_0 = 1 - e^{-\lambda/\mu} < 1.$$

$$P_k = \frac{(\lambda/\mu)^k}{k!} e^{-\lambda/\mu}, \quad k = 0,1,2,\ldots \text{ Poisson Distribution!}$$

$$L = \text{time-average number in system} = \lambda/\mu \quad \text{How?}$$

$$L = \lambda_A W \qquad \text{Little's Law, where}$$

$$\lambda_A \equiv \text{average rate of accepted arrivals into system}$$

Apply Little's Law to Service Facility

$$\rho = \lambda_A (\text{average service time})$$

$$\rho = \text{average number in service facility} = \lambda_A / \mu$$

$$\lambda_A = \mu\rho = \mu(1 - e^{-\lambda/\mu})$$

$$W = \frac{L}{\lambda_A} = \frac{\lambda/\mu}{\mu(1 - e^{-\lambda/\mu})} = \frac{\lambda}{\mu^2(1 - e^{-\lambda/\mu})}$$