ESD.86

Markov Processes and their Application to Queueing

Richard C. Larson

March 5, 2007

Photo courtesy of Johnathan Boeke: http://www.flickr.com/photos/boeke/134030512/
Outline

- Spatial Poisson Processes, one more time
- Introduction to Queueing Systems
- Little’s Law
- Markov Processes
Spatial Poisson Processes

Courtesy of Andy Long. Used with permission.
Spatial Poisson Processes

- Entities distributed in space (Examples?)
- Follow postulates of the (time) Poisson process
  - $\lambda dt = \text{Probability of a Poisson event in } dt$
  - History not relevant
  - What happens in disjoint time intervals is independent, one from the other
  - The probability of a two or more Possion events in $dt$ is second order in $dt$ and can be ignored

- Let’s fill in the spatial analogue…..
Set $S$ has area $A(S)$. Poisson intensity is $\gamma$ Poisson entities/(unit area). $X(S)$ is a random variable $X(S) = \text{number of Poisson entities in } S$

$$P\{X(S) = k\} = \frac{(\gamma A(S))^k}{k!} e^{-\gamma A(S)}, \; k = 0,1,2,...$$
Nearest Neighbors: Euclidean

Define $D_2$ = distance from a random point to nearest Poisson entity

Want to derive $f_{D_2}(r)$.

Happiness:

\[
F_{D_2}(r) = P\{D_2 \leq r\} = 1 - P\{D_2 > r\}
\]

\[
F_{D_2}(r) = 1 - \text{Prob}\{\text{no Poisson entities in circle of radius } r\}
\]

\[
F_{D_2}(r) = 1 - e^{-\gamma \pi r^2} \quad r \geq 0
\]

\[
f_{D_2}(r) = \frac{d}{dr} F_{D_2}(r) = 2r \gamma \pi e^{-\gamma \pi r^2} \quad r \geq 0
\]

Rayleigh pdf with parameter $\sqrt{2 \gamma \pi}$
Nearest Neighbors: Euclidean

Define $D_2 =$ distance from a random point to nearest Poisson entity

Want to derive $f_{D_2}(r)$.

\[
E[D_2] = (1/2) \sqrt{\frac{1}{\gamma}} \quad \text{"Square Root Law"}
\]

\[
\sigma_{D_2}^2 = (2 - \pi/2) \frac{1}{2 \pi \gamma}
\]

\[
f_{D_2}(r) = \frac{d}{dr} F_{D_2}(r) = 2r \gamma \pi e^{-\gamma \pi r^2} \quad r \geq 0
\]

Rayleigh pdf with parameter $\sqrt{2 \gamma \pi}$
Nearest Neighbor: Taxi Metric

\[
F_{D_1}(r) \equiv P\{D_1 \leq r\} \\
F_{D_1}(r) = 1 - \Pr\{\text{no Poisson entities in diamond}\}
\]
How Might you Derive the PDF for the $k^{th}$ Nearest Neighbor?

Blackboard exercise!
To Queue or Not to Queue,
That May be a Question!
Queueing System

Arriving Customers

Queue of Waiting Customers

Service Facility

Departing Customers

Figure by MIT OCW.
Finite or Infinite?

Queue Discipline: How queuers Are selected for service

Servers: Statistical Clones?

Source: Larson and Odoni, Urban Operations Research
What Kinds of Queues Occur in Systems of Interest to ESD?

ESD Queues?
Little’s Law for Queues

\[ a(t) = \text{cumulative \# arrivals to system in } (0, t] \]
\[ d(t) = \text{cumulative \# departures from system in } (0, t] \]
\[ L(t) = a(t) - d(t) \]
\[ L(t) = \text{number of customers in the system (in queue and in service) at time } t \]
Little’s Law for Queues

\[ \gamma(t) = \int_0^t [a(\tau) - d(\tau)] d\tau = \int_0^t L(\tau) d\tau \]

\( \gamma(t) = \) total number of customer minutes spent in the system

\( a(t) = \) cumulative # arrivals to system in \((0, t]\)

\( d(t) = \) cumulative # departures from system in \((0, t]\)

\( L(t) = a(t) - d(t) \)

\( L(t) = \) number of customers in the system (in queue and in service) at time \( t \)
Let’s Get an expression for Each of 3 Quantities

\[ \lambda_t \equiv \text{average customer arrival rate} = a(t)/t \]

\[ W_t \equiv \text{average time that an arrived customer has spent in the system} \]

\[ W_t = \frac{\gamma(t)}{a(t)} = \frac{\gamma(t) a(t)}{t} \]

\[ L_t = \frac{\gamma(t)}{t} = \frac{a(t) \gamma(t)}{t} = \lambda_t W_t \]

In the limit,

\[ L = \lambda W, \quad \text{Little's Law} \]
Key Issues

$L = \lambda W$

- $L$ in a time-average. Explain
- $\lambda$ is average of arrival rate of customers who actually enter the system
- $W$ is average time in system (in queue and in service) for actual customers who enter the system
More Issues

- Little’s Law is general. It does not depend on:
  - Arrival process
  - Service process
  - # servers
  - Queue discipline
  - Renewal assumptions, etc.

- It just requires that the 3 limits exist.

\[ L = \lambda W \]
Still More Issues

- What about balking? Reneging? Finite capacity?
- Do we need iid service times? Iid interarrival times?
- Do we need each busy period to behave statistically identically?
- Look at role of $\gamma(t)$. Can change queue statistics by changing queue discipline.

$L = \lambda W$
FCFS = First Come, First Served
SJF = Shortest Job First

What about LJF, Longest Job First?
“System” is General

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

\[
L = \lambda W
\]

\[
L_q = \lambda W_q
\]

\[
L_{SF} = \lambda W_{SF} = \lambda / \mu
\]

\[
1 / \mu = \text{mean service time}
\]
All of this means, “You buy one, you get the other 3 for free!”

\[
W = \frac{1}{\mu} + W_q
\]

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

\[
L = \lambda W
\]
Utilization Factor $\rho$

- Single Server. Set

$$Y = \begin{cases} 1 & \text{if server is busy} \\ 0 & \text{if server is idle} \end{cases}$$

$$E[Y] = 1 \times P\{\text{server is busy}\} + 0 \times P\{\text{server is idle}\}$$

$$E[Y] = 1 \times \rho + 0 = \rho = E[\# \text{ customers in SF}] = ?$$

- $E[Y]$ is time-average number of customers in the SF

- Buy Little’s Law,

$$\rho = \frac{\lambda}{\mu} < 1$$
Utilization Factor $\rho$

- Similar logic for $N$ identical parallel servers gives

\[
\rho = \left( \frac{\lambda}{N} \right) \frac{1}{\mu} = \frac{\lambda}{N\mu} < 1
\]

- Here, $\lambda/\mu$ corresponds to the time-average number of servers busy
Markov Queues

Markov here means, “No Memory”
State-transition diagram for the fundamental birth-and-death model.

Probabilities of transitions for birth-and-death model in time $\Delta t$.

Source: Larson and Odoni, Urban Operations Research
Balance of Flow Equations

\[ \lambda_0 P_0 = \mu_1 P_1 \]
\[ (\lambda_n + \mu_n)P_n = \lambda_{n-1}P_{n-1} + \mu_{n+1}P_{n+1} \text{ for } n = 1,2,3,... \]