The Theory of Queues

Models of

Waiting in line

Queuing Theory

- Basic model:
- Customers ➔ Queue ➔ Served ➔ Done

- How long will it take to get done?
- How many customers are waiting?

Customers ➔ Queue ➔ Served ➔ Done

- How long will it take to get done?
  - Time allocation
    - For customers
    - For service
- How many customers are waiting?
  - Space allocation
  - Social dynamic

Customers ➔ Queue ➔ Served ➔ Done

- λ Lambda – Average arrival rate
- μ Mu – Average service rate

- $1/\lambda$ = Average time between arrivals
- $1/\mu$ = Average time for service

Customers ➔ Queue ➔ Served ➔ Done

- λ Lambda – Average arrival rate
- μ Mu – Average service rate

- μ > λ – For all models, the service rate is faster than the arrival rate, on average.
  - If λ > μ, the queue will grow and grow.

Y a Q if $\lambda < \mu$?

- If customers are served faster than the time between arrivals, why is there a queue?

Y a Q if $\lambda < \mu$?

- If customers are served faster than the time between arrivals, why is there a queue?

- Random variation in arrival and service times.

Randomness Modeled

- Poisson distribution
- - random variation with no limit
  - extremes are unlikely but possible
- Event can happen at any time
- Events are independent
- We know the average rate of occurance

Poisson distribution

- Limit of the binomial
- Cut t amount of time into smaller and smaller pieces
- Only one arrival allowed during each piece.
- The probability of an arrival during the piece is λ/N, where N is the number of pieces.
- As $N \rightarrow \infty$, the probability of x arrivals during t amount of time is …

Poisson distribution

- Prob (x arrivals during time t) =

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

Poisson distribution

- Application: Running total of Poisson distribution can show how much inventory is needed.

Exponential Distribution

- Probability (next ARRIVAL during the next t amount of time)=

$$1 - e^{-\lambda t}$$

Exponential Distribution

- Probability (SERVER will finish current customer during the next t amount of time)=

$$1-e^{-\mu t}$$

Steady State

- When the distribution of probabilities of various numbers in the system stabilizes
- When startup conditions are far behind
- $\lambda<\mu$ required for steady state to exist
- $\rho=\lambda/\mu$ is how busy the server is
- ρ is server busy time ÷ all time
- $\rho < 1$ for steady state to be possible

Steady State

- For single-server single-stage queue with Poisson arrivals and service:
- Probability of n in the system (waiting or being served)= $(1-\rho)\rho^n$

Customers in System

- L – customers in system = $\rho/(1-\rho)$
- L_q – customers in queue = ρL
- These are averages, once the steady state is reached.
- There are usually fewer in the system than L because the prob of n in system is skewed.

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

- When the arrival rate nearly equals the service rate, the line can get very long.
- If the arrival rate nearly equals the service rate, small changes in either rate can make a big difference.

Time in System

- W – time in system = L/λ

- W_q – time waiting to be served = ρW

- These are averages, once the steady state is reached.
- Most customers get through in less time.

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

- When the arrival rate and the service rate are nearly equal, the wait can be very long.
- When the arrival rate and the service rate are nearly equal, small changes in either rate can make a big difference.

Economic Analysis

- Weigh
- The cost of waiting
  - Against
- The cost of speeding up service
  - or reducing arrivals.