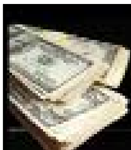
EMC 445 EMS SYSTEMS MANAGEMENT

EMS System Staffing I:

Queuing Theory



Unit Objectives

- Upon completion of this unit, you should be able to:
  - Define queuing theory.
  - Describe the components of queuing theory such as arrival rate, service time, availability reliability, queue discipline, and channels.
  - Calculate the probability of a patient encountering a queue delay.
  - Determine the number of ambulances needed for a stated availability goal.
  - Use queuing theory to develop staffing plans for a response district.

Introduction to Queuing Theory

- Response time is determined by:
  – Ambulance availability
  – Ambulance location vs. call location
- Historically, staffing levels determined by unsophisticated analysis of demand
- Queuing Theory is a method of analyzing availability

Review of Mathematical Concepts

- Factorial (!) – to calculate the factorial of any number, take the number and multiply it by one less that number, then take that quantity and multiply it by two less than the original number. Repeat this process until you are finally multiplying the quantity by 1.

  Examples:

  3! = 3x2x1 = 6
  4! = 4x3x2x1 = 24
  5! = 5x4x3x2x1 = 120

  Note: 0! = 1
  1! = 1

- Exponentiation (X^y) – multiply x by itself y times

  Examples:

  2^2 = 2x2 = 4
  2^3 = 2x2x2 = 8
  2^4 = 2x2x2x2 = 16
  Note: $X^0 = 1$ any number raised to the power of zero equals 1 ($3^0 = 1$, $10^0 = 1$)
  $X^1 = 1$ any number raised to the power of 1 equals itself ($3^1 = 3$, $10^1 = 10$)

Review of Mathematical Concepts continued

$$\sum_{n=0}^{n=k-1} 5 x n$$

- Summation (∑) – asks you to repeat a mathematic operation, replacing certain values each time, then adding together (or summing) the results of each repetition. The summation sign (∑) may also have an index. The index tells you which variable will be substituted in the mathematical operation, and what values to use during each substitution.

- For example:

$$\sum_{n=0}^{n=k-1} 5n$$

- Suppose we know that the value of k is 4. The notation under the summation sign indicates that we repeat the function (5*n), beginning with n = 0. Then we increase the value of n by 1 for each repetition until we reach the value above the summation sign. In this case, n = (4-1) = 3. Once we have calculated the value of the operation for each iteration, we SUM all of the values across all iterations.

- Example:

For n = 0	5 x 0 = 0
For n = 1	5 x 1 = 5
For n = 2	5 x 2 = 10
For n = 3	5 x 3 = 15

- Then we add together the results (0 + 5 + 10 + 15) = 30.

Components of Queuing Theory

- **Arrival rate** - the number of customers (ambulance calls) received during a given period of time (usually calculated on an hourly basis). For most problems, we assume that arrivals follow a particular probability distribution called a **Poisson pattern** or **Poisson probability distribution**. The Poisson probability distribution is described by the following equation:

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

where,

x = number of arrivals in a specified period of time
λ = average or expected number of arrivals for the specific period of time
e = 2.71828

Components of Queuing Theory continued

- Example: During a given hour, past data reveal that the average number of ambulance calls received is 3. With this single parameter, we can calculate the probability of receiving any specified number of calls during this given hour.
- The probabilities of receiving 0 thru 4 calls during this hour are as follows:

$$P(x=4) = \frac{3^4 e^{-3}}{4!} = \frac{81 x 2.71828^{-3}}{4 x 3 x 2 x 1} = 0.1680$$

$$P(x=3) = \frac{3^3 e^{-3}}{3!} = \frac{27 x 2.71828^{-3}}{3 x 2 x 1} = 0.2241$$

$$P(x=2) = \frac{3^2 e^{-3}}{2!} = \frac{9 x 2.71828^{-3}}{2 x 1} = 0.2241$$

$$P(x=1) = \frac{3^1 e^{-3}}{1!} = \frac{3 x 2.71828^{-3}}{1 x 1} = 0.1494$$

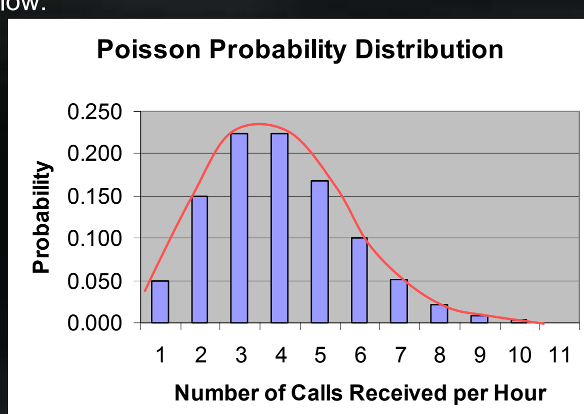
$$P(x=0) = \frac{3^0 e^{-3}}{0!} = \frac{1 x 2.71828^{-3}}{1} = 0.0498$$

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

This means that we would expect to receive no calls during this hour 4.98% of the time, one call 14.94% of the time, and 2 calls 22.41% of the time, and so on

Components of Queuing Theory continued

- If we continued on and calculated the probabilities for receiving 5, 6, 7, etc., calls, we would get a probability distribution such as the one below:



Components of Queuing Theory continued

- **Service Time Distribution** - Similar to describing the distribution of call arrival rates, we need a mechanism to describe how long it takes to service the call. Because of the variation in response times, the time to treat patients at the scene, and transport times, the total time required to complete or "service" a call varies from call to call. In general, we can describe the variation in service time using the **exponential probability distribution**. The exponential probability distribution is defined as follows:

$$f(x) = \mu^{e-\mu x}$$

where,

x = service time
μ = the number of calls that can be handled during a specified period of time, calculated as 60 divided by the average time required to service a call.

Example: An EMS system takes an average of 45 minutes to complete a call. μ = 60 minutes/45 minutes = 1.33. Therefore, the service capacity of a single ambulance is 1.33 calls per hour.

With this equation, we can compute the probability of any one call taking *x* minutes to complete, although we really don't need to for our purposes.

Components of Queuing Theory continued

- **Queue Discipline**

When the call arrival rate exceeds the service capacity of an EMS system, calls will "back up", "stack", or more accurately, queue. How these calls are handled is referred to as queue discipline. Examples of queue disciplines include:

FIFO
- "first in, first out". This is what we call "first come, first served" or "serviced in the order in which they are received".

LIFO
- "last in first out". Similar to the queue discipline of an elevator.

Truncated
- we assume that calls are not allowed to queue, i.e., they are serviced by someone else.

Infinite
- calls are allowed to pile up until all are serviced.

Prioritized
- service certain calls before others, such as emergency calls before non-emergency

- For our purposes, the equations we use will assume FIFO with infinite queuing. Although other methods exist for dealing with the other queue disciplines, they are mathematically complex and some queue disciplines cannot be solved mathematically; they must be modeled using computer simulation.

Components of Queuing Theory continued

- **Channels**

Queuing systems are either single or multi-channeled. Single channel systems means that there is a single queue and a single service facility (i.e., ambulance). A multiple channel queuing system means that there is more than one ambulance. Although the equations for analyzing a single channel queuing system are simpler than those for the multiple channel system, they have limited applications in EMS (maybe for calculating service times for a single 911 dispatcher). So, we will focus only on a multiple channel queuing system with Poisson arrivals and exponential service times, where:

k = number of channels (ambulances)

λ = mean arrival rate for the system **for a specific hour**

μ = mean service **rate** for each channel (can assume to be identical across all ambulances within a given system or district, but will probably vary across districts due to variation in response times and transport times)

Queuing Calculations

- We can learn much about how an EMS system operates using queuing theory and the equations given below. We can use queuing theory to examine issues related to staffing, waiting times, and queue lengths, all of great importance to an EMS administrator.

k = number of channels (ambulances)

λ = mean arrival rate (calls per hour)

μ = mean service time per call

1. The probability that all ***k*** service channels (ambulances) are idle (i.e., the probability of zero calls in the system): **for kμ > λ**

$$P_0 = \frac{1}{\left[\sum_{n=0}^{n=k-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \frac{k\mu}{k\mu-\lambda}}$$

2. The probability that the next call has to wait for service:

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

Queuing Calculations continued

- **Sample Problem**

  An EMS district has three ambulances staffed 24 hours per day. This particular district has suffered from poor response times and the crews are complaining about being overworked. You are considering increasing staffing of this district from three to four ambulances. You collect data for the busiest hour during the day (i.e., where performance is likely to be the poorest), and determine that the average call arrival rate is 3 calls per hour and that the average time required to service a single call is 35 minutes. Analyze the **current** performance of the system.

We know that:

- μ = 60/35 = 1.7 calls per hour (service rate)
- λ = 3 calls per hour
- k = 3 ambulances

Queuing Calculations continued

$$P_0 = \frac{1}{\left[\sum_{n=0}^{n=k-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right] + \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \frac{k\mu}{k\mu-\lambda}}$$

μ = 60/35 = 1.7 calls per hour
λ = 3 calls per hour
k = 3 ambulances

$$\left[\sum_{n=0}^{n=k-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n\right]$$

$$n=0: \quad \frac{1}{0!}\left(\frac{3}{1.7}\right)^0 = 1$$

$$n=1: \quad \frac{1}{1!}\left(\frac{3}{1.7}\right)^1 = 1.76$$

$$n=2: \quad \frac{1}{2!}\left(\frac{3}{1.7}\right)^2 = (0.5)(3.114) = 1.557$$

$$1 + 1.76 + 1.557 = 4.317$$

$$P_0 = \frac{1}{4.317 + \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \frac{k\mu}{k\mu-\lambda}}$$

$$P_0 = \frac{1}{4.317 + \frac{1}{3!}\left(\frac{3}{1.7}\right)^3 \frac{(3)(1.7)}{(3)(1.7)-3}}$$

$$P_0 = \frac{1}{4.317 + (0.167)(5.495)\left(\frac{5.1}{2.1}\right)}$$

$$P_0 = \frac{1}{4.317 + 2.228} = \frac{1}{6.545} = 0.152$$

Queuing Calculations continued

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

$$P_w = \frac{1}{3!}\left(\frac{3}{1.7}\right)^3 \frac{(3)(1.7)}{(3)(1.7)-3} (.152)$$

$$P_w = \frac{1}{6}(1.764)^3\left(\frac{5.1}{2.1}\right)(.152)$$

$$P_w = (.1667)(5.489)(2.43)(.152)$$

$$P_w = 0.337$$

Use of Queuing Theory in Developing Staffing Plans

- Previous calculations must be repeated across each hour of the day, each day of the week, and for each district.
- This will yield a table similar to the one below.
- You may also want to vary the unit availability goal to determine the impact on staffing levels.
- Unit availability is not a linear function!

EMS SYSTEM STAFFING PATTERN ANALYSIS
M/M/N QUEUEING SYSTEM MIN. UNIT AVAILABILITY = .90

DISTRICT 1

DAY/HR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	2	0	1	1	0	0	1	1	1	2	2	2	1	1	0	1	1	0	0	1	1	0	1
2	1	0	0	0	0	1	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
3	0	0	1	0	0	0	0	1	2	2	0	1	0	0	2	0	1	0	1	0	1	0	0	1
4	0	1	0	1	1	1	0	1	0	2	1	2	1	1	2	0	1	0	1	1	0	0	0	1
5	0	0	0	0	0	0	0	1	0	1	2	0	1	1	0	1	0	0	0	1	2	0	0	1
6	1	1	1	1	1	1	0	2	0	1	1	0	0	1	1	0	1	0	0	2	0	1	0	1
7	1	0	0	0	0	0	0	1	1	0	2	0	0	0	0	0	1	1	1	0	0	0	0	0

Use of Queuing Theory in Developing Staffing Plans

EMS SYSTEM STAFFING PATTERN ANALYSIS
M/M/N QUEUEING SYSTEM   MIN. UNIT AVAILABILITY = .90

DISTRICT 1

DAY/HR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	2	0	1	1	0	0	1	1	1	2	2	2	1	1	0	1	1	0	0	1	1	0	1
2	1	0	0	0	0	1	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
3	0	0	1	0	0	0	0	1	2	2	0	1	0	0	2	0	1	0	1	0	1	0	0	1
4	0	1	0	1	1	1	0	1	0	2	1	2	1	1	2	0	1	0	1	1	0	0	0	1
5	0	0	0	0	0	0	0	1	0	1	2	0	1	1	0	1	0	0	0	1	2	0	0	1
6	1	1	1	1	1	1	0	2	0	1	1	0	0	1	1	0	1	0	0	2	0	1	0	1
7	1	0	0	0	0	0	0	1	1	0	2	0	0	0	0	0	1	1	1	0	0	0	0	0

REGIONAL EMS SYSTEM STAFFING PATTERN ANALYSIS
M/M/N QUEUEING SYSTEM   MIN. UNIT AVAILABILITY = .90

DISTRICT 2

DAY/HR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	3	2	3	3	3	2	3	3	2	3	4	3	3	4	4	3	2	3	3	3	3	3	3	2
2	2	3	2	2	2	2	3	3									3	4	3	3	3	2	3	2
3	3	2	3	2	2	2	2	2									3	3	4	3	2	2	2	2
4	2	2	3	2	3	2	2	3									4	3	3	3	3	3	3	2
5	2	2	2	2	2	0	2	2									4	3	2	2	4	3	3	2
6	2	3	3	3	3	2	3	2									4	3	3	4	3	3	4	2
7	2	3	2	2	2	0	2	2	2	3	2	4	3	2	5	4	2	3	3	3	3	3	3	3