MFS605/EE605
Systems for Factory Information and Control

Lecture 3
Fall 2005

Larry Holloway
Department of Electrical and Computer Engineering,
UK Center for Manufacturing

- **References:**
  - **Modeling and Analysis of Mfg. Systems, Askin and Standridge (John Wiley, 1993)**
  - **Factory Physics, by Hopp and Spearman (McGraw Hill 1996)**
  - **Simulation Modeling and Analysis, 2nd ed. Law and Kelton, 1991 (McGraw Hill)**
  - **Probability and Random Processes for Electrical Engineering, 2nd edition, A. Leon-Garcia, 1994 (Addison-Wesley)**
  - **Nelson, "Stochastic Modeling, Analysis, and Simulation", 1995**
  - **Hillier & Lieberman, "Intro to Oper. Research, 6th edition", 1995**

Previous Class:

Previous Class:

- **Considered serial assembly systems**
  - **Paced vs. unpaced**

- **Line Balancing Problem**
  - ***Given task times and precedence relations, how do we allocate tasks to least number of workstations?***
    - Search methods
    - RPW
    - COMSOAL
    - Etc…

But, each of these assumed deterministic task times!

Randomness…

Sources of Variability:

- Natural variation
  - operator difference, tool wear, variation of material

- Breakdown/repair and other unexpected delays

- Setups and other irregular but expected delays

- Quality problems…

To study variability/randomness, we need basics of probability…

Basic Probability (see appendix A)

Probability

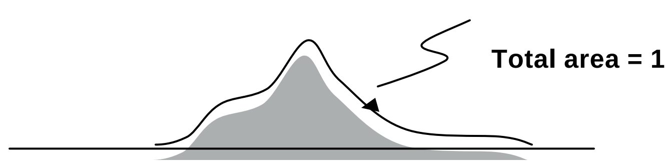
- We have an "experiment" and a set of possible "outcomes".
- Probability represents notion of "relative frequency"
  - how frequently does a given outcome occur given repeated experiments?

- Axioms of Probability: Given some subset A of outcomes:
  - (a.) $0 \le \text{Prob}[A] \le 1$ for any subset A of outcomes
  - (b.) Prob[set of all possible outcomes] = 1
  - (c.) If A and B are mutually exclusive sets of outcomes, then Prob[A U B] = Prob[A] + Prob[B]

Probability Distributions

- Random Variable: A map from set of outcomes to the real line
  - (assumes some fine restrictions are covered)
  - If experiment outcomes are real numbers anyway, then map is simple.
  - Example: processing times, numbers in queue, …

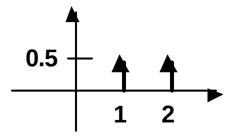
- Probability Distribution:
  - Intuitive notion: integrating between two points gives the probability of the random variable having value in that range.



Probability Distributions

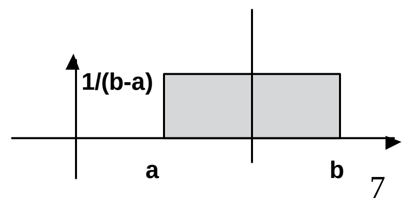
- **Probability Distributions are either:**
  - **Discrete: random variable only takes on countable number of possible values (nonzero probability at only specific points)**

    examples:
    - # of parts in buffer
    - # of machines needing repair
    - "heads" vs. "tails"



  - **Continuous: random variable takes on continuous range of possible values**

    examples
    - temperature, voltage, pressure
    - time to process
    - time between customers

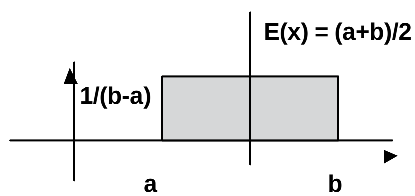


Expected Value

- **E(X) is *expected value* of random variable x**

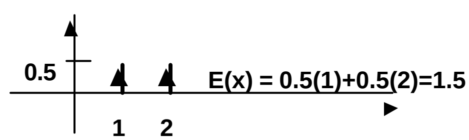
Continuous:

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$



Discrete:

$$E(x) = \sum_k k p(k)$$

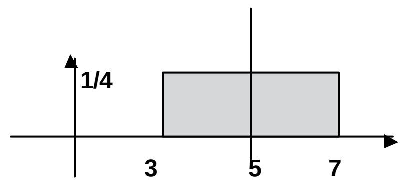


- **expected value of an Rand.Var. is also known as its *mean value,*** m
  - (*typically* this is the same as the "average value")

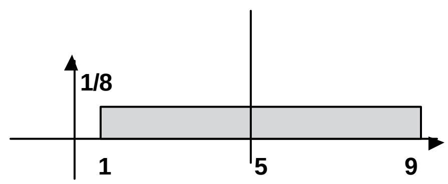
Variance

- **Variance of X is: Var(X) = s^2 = E((x – m) ^{2})**
  - **Standard Deviation of X is square root of variance: = s**
- **Variance indicates spread of the distribution.**

- **Example:**



E(x) = 5
Var(x) =1.33
StdDev = 1.15



E(x) = 5
Var(x) =5.33
StdDev = 2.3

Poisson Distribution

- **Poisson Distribution:**

  P{X=k} = (l ^{k} e^{-l}) / k!
  - **discrete distribution**
  - **parameter l is expected value (mean) of Poisson random variable X**
    - "avg rate of occurrences"
    - $\lambda = \mu$, *and* $\lambda = \sigma^2$
  - **used when events occur "completely at random" in time or space**
  - **Examples:**
    - # of customers walking into a bank / unit time
    - # of customers requesting starting a telephone call over period
    - # of failures occurring (or repair requests, etc.)
    - # of products with given option requested during period
    - # of parts in a batch of random size
    - # of parts demanded from inventory

Poisson Distribution

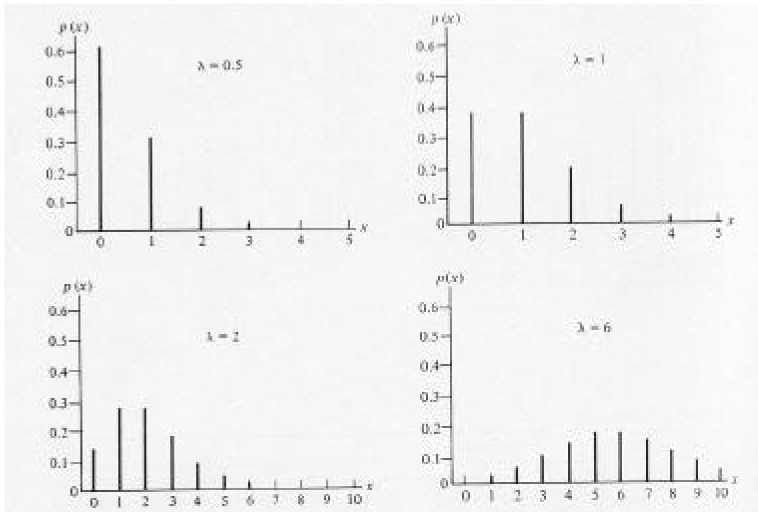


FIGURE 6.16
Poisson(λ) mass functions.

- From Law and Kelton, 1991

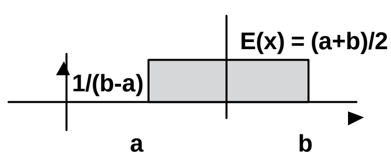
Continuous Random Variables

Continuous random variables

- **--> no probability associated to specific values, instead associated with intervals of values**
  - **probability is area under curve of distribution function**

<u>Uniform Distribution:</u> **$f_X(x) =$** $\begin{cases} 1/(b-a) & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$

E(X) = m = (a+b)/2
Var(X) = s^2 = (b-a)^2 / 12

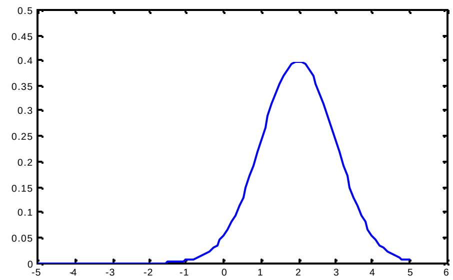


Used as a "first" model for a quantity that is felt to be randomly varying between a and b, but about which little is known

Gaussian (Normal) Distribution:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2}$$

- μ **is mean = E(x)**
- σ^2 **is variance = E((x-m)²)**



Example applications:

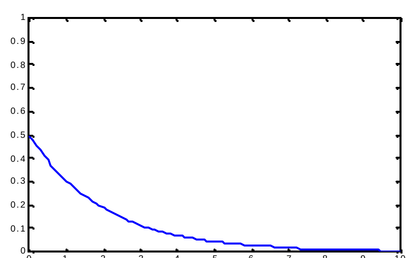
Errors of various types

Quantities that are sums of other quantities

"Central Limit Theorem" says that the <u>sums</u> of independent R.V.s will approach a normal distribution (approaches as we increase number of terms in the sum)

Exponential distribution: $\quad f_X(x) = \begin{cases} 0 & \text{if } x<0 \\ \lambda e^{-\lambda x} & \text{for } x \geq 0 \end{cases}$

- **mean = 1/** λ
- **Var = (1/** λ **)²**



- **Typically models time between occurrences of events that occur completely at random but at with constant avg. rate (** λ **)**
- **Examples:**
  - **time between initiation and completion of "service"**
  - **time between customer arrivals**
  - **time between failures**

Poisson vs. Exponential

A process with *Poisson arrival distribution* has *exponentially distributed time* between events

- arrival mean of λ parts/hour has
  interarrival time mean = $1/\lambda$ (hours/ part)
- *Key point:* *Memoryless property*
  - probability of having to wait h more time units before next arrival is same *regardless of when previous arrival occurred.*
    P[X > $t+h$ given X>t] = P [X > h]

Other Distributions:

Gamma and Wiebull distributions

- **Application: time to complete a task, such as customer service or machine repair (typically done in reasonable time, but sometimes takes awhile)**
- **Very versatile by changing parameters**
- **No negative tail**

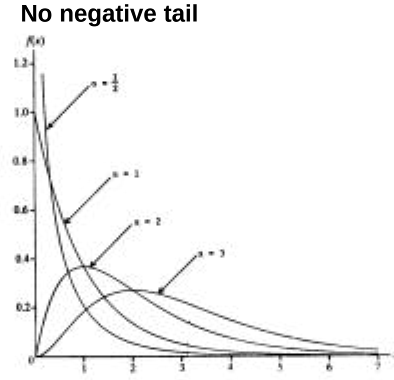


FIGURE 6.3
gamma(α,1) density functions

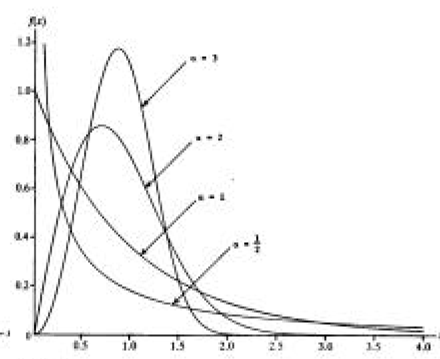


FIGURE 6.4
Weibull(α,1) density functions

- From Law and Kelton, 1991

Randomness…(again)

Sources of Variability:

- Natural variation
  - operator difference, tool wear, variation of material
- Breakdown/repair and other unexpected delays
- Setups and other irregular but expected delays
- Quality problems…

Coefficient of Variation

Characterizing variation:

- **Coefficient of variation:**

CV = (standard deviation) / (mean) = σ / μ .

= sqrt(Variance) / mean

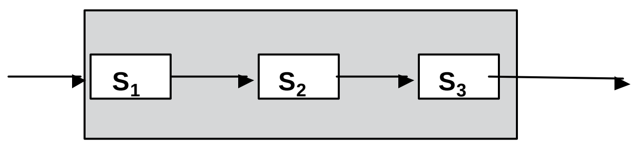
Examples:

Handling time variation - synchronous lines

- **Last class: Line balancing of synchronous lines**
  - **Workstation tasks must be done in allotted time *C.***
  - **Effectively no buffers**

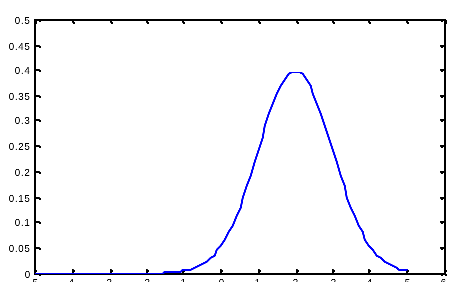
- **What if we have variation in these task times?**

- **Possible solutions:**
  1. **(Best solution) Try to reduce/eliminate this variation**
  2. **Accept that sometimes tasks won't get done:**
     - Set up emergency response system to help (impractical for fast lines – so may lead to some starvation or blockages)
     - Rework stations
     - Line stoppages ("stop the line")
  3. **Consider variation slack-time in our balancing**



Slack Time in Synchronous Lines

- If we knew the probability density function (pdf) for the times, then we could determine a "max" time for our tasks
- "max" is not true max, but is a cutoff covering sufficient probability.



Example: Suppose normal distribution on time.

- **From standard normal table we have:**
  - **For 95% below: less than <u>1.65</u> std. deviations above mean**
  - **For 99% below: less than <u>2.33</u> std. deviations above mean**
  - **For 99.9% below: less than <u>3.1</u> std. deviations above mean**

- **Thus, for probability of at least 95% of being able to complete tasks within cycle, the time to use is:**

  E(X) + 1.65*Std.Dev(X)

- **Thus, for probability of at least 99% of being able to complete tasks within cycle, the time to use is:**

  E(X) + 2.33*Std.Dev(X)

- **(Note this can reduce utilization of stations, so it is of course better to try to remove the variation as much as possible)**

- **Example:** tasks assigned to station have average time of 60 seconds, with variance of 4. To ensure 99% of time the tasks are done, we must have cycle time greater than

  60 + 2.33 *(2) = 64.66

  If we have a cycle time *C* of 65 seconds, how much idletime is there on average?

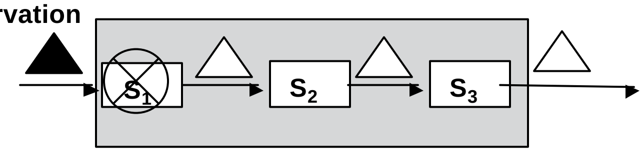
- How reasonable was it to use the normal distribution here?
  - **Reasonable if the individual task times are normal**

  or

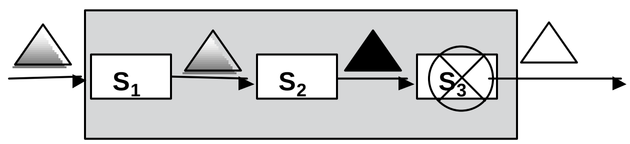
  - **Reasonable if the time is sum of other (independent) times**
  - **Why?**
- If we don't know distribution, we can still consider other ways of including this slack time to cover variability

The need for buffers

- **WIP buffers reduce the impact of stations on each other.**
  - **Starvation**



  - **Blocking**



Where are buffers most needed?

Random Processing Times without buffers

- Effect of Random Processing Times in balanced, unbuffered lines with no breakdowns (Askin/Standridge)
- CV is Coefficient of Variation = (std.dev.) / mean

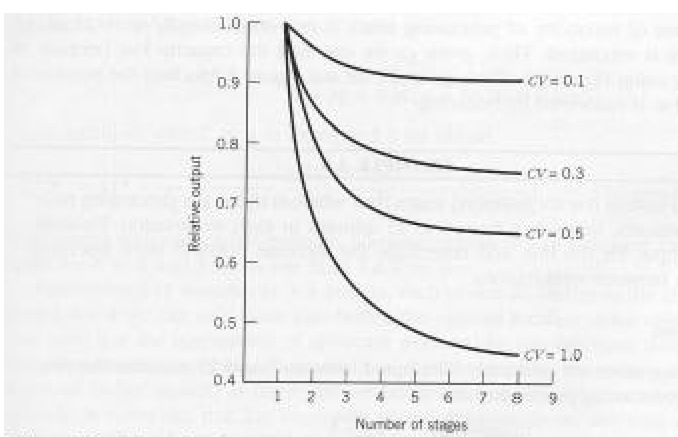


Figure 3.6 Effect of random processing time in balanced, unbuffered lines.

Example:

- five stages, each with 10 minutes avg. time and CV=1 (std.dev = 10). Then rate is only 50%.

- If CV reduced to 0.1 (std.dev. = 1 min.), then rate is at 90%.

- Observations:
  – Throughput decreases as # of stations increase, but then levels off.
  – Reducing CV will reduce our losses.
  – We sacrifice output by not having buffers

Recovery of lost output from buffering

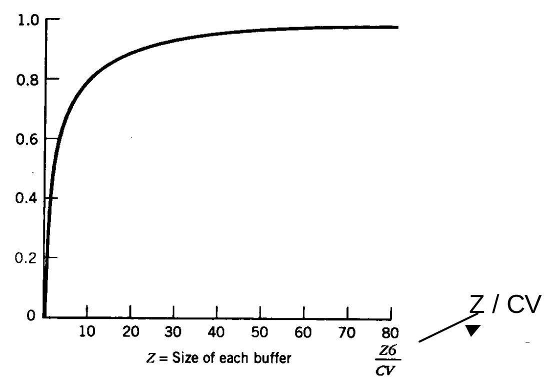


Figure 3.7 Proportion of lost output recovered by buffering in balanced lines.

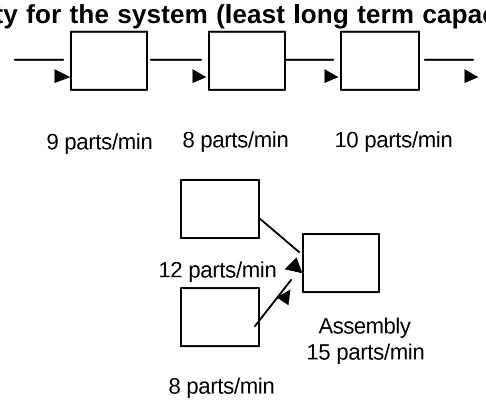
From Askin and Standridge

Definitions (review from class 1)

- **Flow Time is time for an order to move through the system**
- ***In this class, we use the term <u>Flow Time</u>, <u>Throughput Time</u>, and <u>Manufacturing Lead Time</u> (Askin&Standridge) to mean the same thing.***
- **Problem: Time spent waiting for machine setup, transportation, machine availability, etc.**
  - **---> includes *Non-value added activities***
    - *Waiting in queue*
    - *Waiting for transportation*
    - *Waiting for setup*
  - **Flow time directly increases with respect to batch size, and indirectly due to longer waits for other batches**

Bottlenecks

- ***<u>Bottleneck</u>*: The portion of the system that constrains capacity for the system (least long term capacity).**



- What is the system capacity?
- (Additional Terms: *Blocking* and *Starvation*

Critical WIP

- *Bottleneck*: constrains capacity for the system (least long term capacity). Let its rate be r_b
- *Raw Process Time:* T_0 : Sum of average process times of workstations along the line. This is theoretical minimum Mfg. Lead Time.
- *Critical WIP*: W_0 : WIP s. t. when *no variability*, we have maximum throughput rate (r_b) with minimum time (T_0)

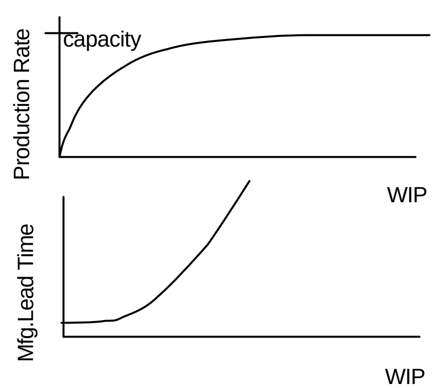
$$W_0 = r_b T_0$$

(This is from Little's Law of WIP = production rate x throughput time where we look at max. throughput rate and minimum time

Note: Critical does not necessarily imply "optimal"

Implication of Little's Law

WIP = Production Rate x Throughput Time

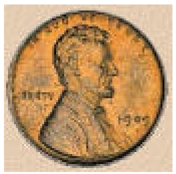


Implications:

- **If not near capacity, then increasing WIP increases rate without time increase. (Everything keeps busy).**
- **If near capacity, then rate cannot increase more – so increasing WIP increases throughput time!**

Example: Penny Fab 1

(from Hopp and Spearman)



Produce giant novelty pennies.

- **Process 1: Punch press cuts penny blanks**
- **Process 2: Stamping of images**
- **Process 3: Rim put on penny**
- **Process 4: Cleaning and deburring**

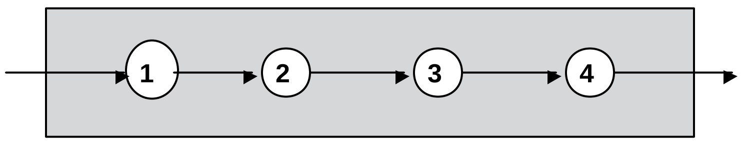
Suppose two hours per machine *("perfectly balanced")*

24 hours /day

r_b = ?

Raw Process Time, T_0 = ?

Critical WIP = ?



Example: Penny Fab 2

(from Hopp and Spearman)

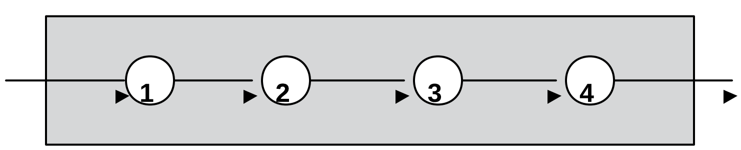
Station no.	# of Machines	Process Time	Station Rate (calc)
1	1	2 hr	0.50 jobs/hr
2	2	5 hr	0.40 jobs/hr
3	6	10 hr	0.60 jobs/hr
4	2	3 hr	0.67 jobs/hr

Station Rate: (#machines)/(time per part per machine)

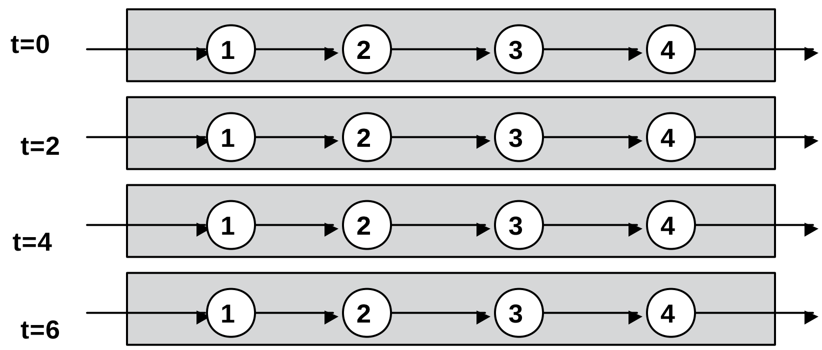
Bottleneck rate = r_b = ?

Total Process Time (excludes waits), T_0 = ?

"Critical WIP" = $r_b T_0$ =?



Best Case Performance: Penny fab 1



WIP	Flow Time	$\%T_0$	TH rate	$\%r_b$
1	8	100	0.125	25
2	8	100	0.250	50
3	8	100	0.375	75
4	8	100	0.500	100
5	10	125	0.500	100

Little’s Law

Key Points about Penny Fab 1:

- **Once we reach enough WIP for max throughput, then added WIP just adds to flow time. No added benefit of additional WIP (in deterministic case)**
- **WIP below critical level: we lose throughput.**

- **This is illustration of *Little’s Law:***

Throughput rate = WIP / Flow time

→ **As WIP grows, either throughput rate or flow time grows**
→ **Reducing Flow time means reducing WIP, as long as TH remains constant**

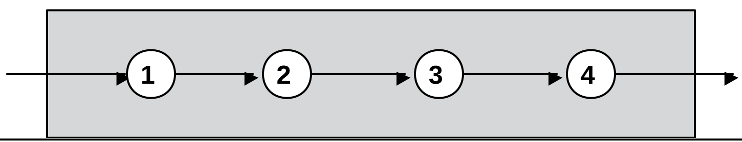
Little's Law applied to Penny Fab 2

- Little's law applies at stations or entire line.

Station no.	# of Machines	Process Time	Station Rate
1	1	2 hr	0.50 jobs/hr
2	2	5 hr	0.40 jobs/hr
3	6	10 hr	0.60 jobs/hr
4	2	3 hr	0.67 jobs/hr

Suppose system running at bottleneck rate of .4/hr

- **Then station 1 has WIP = rate*FT = .4 j/hr*2hrs = .8 jobs**
- **Then station 2 has WIP = rate*FT = .4 j/hr*5hrs = 2 jobs**
- **Then station 3 has WIP = rate*FT = .4 j/hr*10hrs = 4 jobs**
- **Then station 4 has WIP = rate*FT = .4 j/hr*3hrs = 1.2 jobs**



Best Case

Best case (zero variability) performance:

- **Minimum flow time for a given WIP given by:**

$$FT_{best} \begin{cases} = T_0 & \text{if } w < \text{critical} \quad \text{(no waiting)} \\ = w / r_b & \text{else} \end{cases}$$

Maximum Throughput rate for a given WIP given by:

$$rate_{best} \begin{cases} = w / T_0 & \text{if } w < \text{critical} \\ = r_b & \text{else} \end{cases}$$

This is the case for PennyFab 1
(perfectly balanced, no variability)

Worst Case (impractical)

Worst Case (impractical): Suppose average processing time per station was same, but actually all the time was spent with the first job.

Note: this is still not random, but we have picked the worst case possible

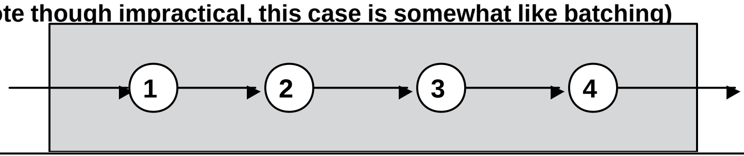
Example: 1st job = 8hrs; 2nd, 3rd, 4th take 0 hours.
Then: avg.=2 still, but 4th job must wait 8 hours at each station!

For given WIP w, define:

$FT_{worst} = w\,T_0$ (waiting on all other WIP)
$rate_{worst} = 1/T_0$

Example: For Penny Fab 1, we have to wait 8 hours each station
For w = 4, Flow time = 8 +8+8+8 = 32 hours = 4*8
Throughput rate = 4/32 = 1/8 jobs / hour

(Note though impractical, this case is somewhat like batching)



Practical Worst Case

- Real systems rarely operate like the best or worst case.
- To consider between the best and worst, we invoke probability.
- The derivation of this "practical worst case" will be done later. For now, we present the basic idea:
  - All stations perfectly balanced
  - All stations are single machines
  - Process times are random. In particular, assume that our system is "memoryless" – so the remaining time for a job at a station is independent of the time that the job has already been there.
- See graphs next page – derivations to be done with queueing material

WIP relationships

(Hopp and Spearman, 1996)

FIGURE 7.9
Cycle time versus WIP in Penny Fab Two

- Flow time vs. WIP

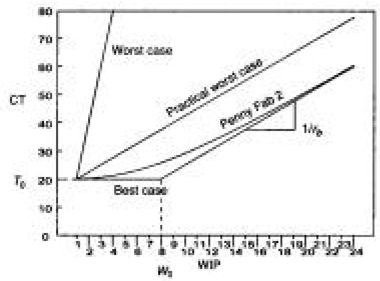
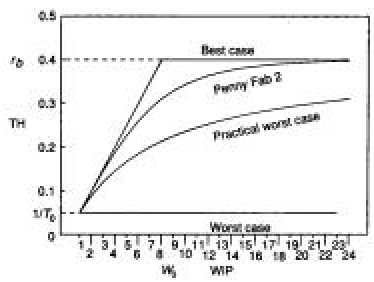


FIGURE 7.10
Throughput versus WIP in Penny Fab Two

- rate vs. WIP

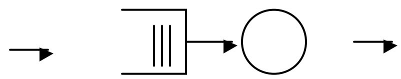


Queueing models

- **<u>Queueing Models</u>**
  - **Probabilistic -- primarily steady state averages**
  - **Advantages:**
    - Analytical solution to problems with randomness & uncertainty
    - --> fast solution (*if solution exists or is known!)*
  - **Disadvantages:**
    - Requires simplifying assumptions
    - Best for smaller models
    - solutions not always possible
    - most suited for steady state analysis

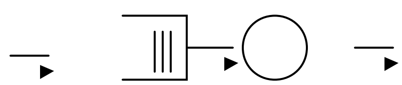
Queuing Theory: Basic Terms

- **Queuing Theory: Study of lines, waiting in lines**



- **Buffer or Queue**
- **Server -- service rate =** m
  - **machine**
  - **repair**
  - **(sometimes even parts)**

- **Customers --- arrival rate =** l
  - **parts waiting for machining**
  - **machines waiting for repair**
  - **...**

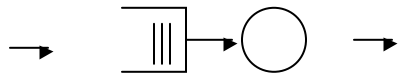


- **Issues: What is distribution of arrivals**
  - What is distribution of service time
  - What is policy for selecting next customer?

- **Queuing theory answers**
  - **expected number of customers in system**
  - **expected number of customers in queue**
  - **avg. customer waiting time in system**
  - **avg customer waiting time in queue**
  - **avg. utilization of server**

Queuing Theory: Basic Terms

- **Queuing Theory: Study of lines, waiting in lines**



- **Buffer or Queue**
- **Server -- service rate =** m
  - **machine**
  - **repair**
  - **(sometimes even parts)**

- **Customers -- arrival rate =** l
  - **parts waiting for machining**
  - **machines waiting for repair**
  - **...**

Queuing Theory

- **In this course, we focus *mostly* on simplest queue: M/M/1**
  - **M stands for *Markov* or *Memoryless***
  - **Notation: Arrival Process / Service Process / # of servers**

  - **M/M/1 -->**
    - Poisson arrivals: (exponential interarrival times)

      time distribution = $\lambda e^{-\lambda t}$ for t ³ 0

      λ = mean arrival rate (expected # arrivals / time)
    - Exponential service times

      time distribution = $\mu e^{-\mu t}$ for t ³ 0

      μ = mean service rate (expected # of customers completing service per unit time.

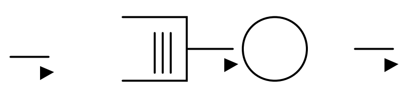
Exponential Assumption

- **Exponential distributions give simplest queuing analysis**
- **Is exponential distribution reasonable?**

  For service time distribution:
  - **inappropriate if service same over most parts (as if same parts)**
  - **appropriate if variety of customers each requiring different operations**

  For Interarrival time distributions:
  - **Appropriate if customers appear in very random manner**
  - **Inappropriate if customers typically appear in groups, or if customers may postpone arrival based on length of queue**

Utilization Assumption



- **We assume** l/m **< 1.**
  - **Why??**

$r = l/m$ **is called *utilization***

Birth-Death Model

- **Quantities of interest in queuing theory:**

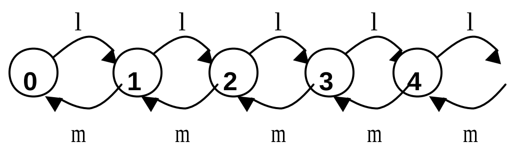
  P_n: probability of having *n* parts in system

  W: avg. wait in system

  L: avg. number of customers in system

  Consider steady-state averages

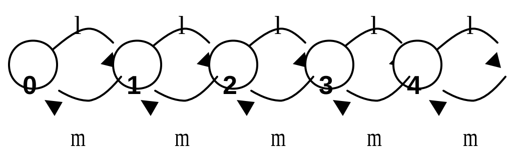
  Assume unbounded queue size



Consider steady-state averages: --> Freq. of entry to state same as frequency of departing from state.

Deriving P_n from birth-death model

- **Balance Equations: entry freq. = exit freq. for each state.**



- **Insert derivations**

Results

- $P_0 = 1 - l/m = 1 - r$
- $P_n = (l/m)^n (1 - l/m) = r^n (1 - r) = r^n P_0$ **(for n>0)**

- **Example 1:** l **= 5 parts/hr,** m **= 10 parts / hr.**

- **Example 2:** l **= 8 parts/hr,** m **= 10 parts / hr.**

Other questions:

- **What is probability that the machine is idle?**

- **What is the probability that there are more than 3 customers in the system?**

Avg. # of customers

- **L: = avg. number of customers *in the system***

$$L = E[n] = \sum_{n=0}^{\infty} nP_n$$

L = $l / (m - l) = r / (1 - r)$

Little's Law

- **W = mean time spent in system**
- ***<u>Little's law</u>*:**

  L = $l_{\mathbf{eff}}$ **W**

  mean number of customers in system =
  effective arrival rate x mean wait time

  – ***Very general result -- not just for M/M/1 queues***

  mean # in "system" = arrival rate **x** ***mean wait in "system"***
  mean # in queue = arrival rate **x** ***mean wait in queue***

  Why the "effective arrival rate"?

- **Nice graphical proof of Little's Law found in Leon-Garcia, 2nd edition, pages 502-503.**

Mean Wait Time in System

- **W = mean wait time in system**

L = $\frac{l}{(m - l)} = \frac{r}{(1 - r)}$

- **Little's law:** **L =** l **W**

- **W = L /** l **=** 1 **/** $(m - l)$

- **Examples:**

Effect of many small vs. single machine

55

* **Example: We have demand rate of 1 part per hour. We want average lead time per part under 2 hours. How fast a workstation do we need?**

56

Summary of results so far

M/M/1 queue:

- **exponential service times** **(service rate =** m **)**
- **Poisson arrivals** **(arrival rate =** l **)**
- **1 server, infinite buffer capacity**

- $P_0 = 1 - l/m = 1 - r$
- $P_n = (l/m)^n (1 - l/m) = r^n (1 - r) = r^n P_0$ **(for n>0)**

- **L: = mean number of customers *in the system***

$$L = E[n] = \sum_{n=0}^{\infty} nP_n = \frac{l}{(m-l)} = \frac{r}{1-r}$$

- **W:= mean wait time in the system**

$$W = \frac{L}{l} = \frac{1}{(m-l)} = \frac{1}{m(1-r)}$$

Waiting time in Queue and # in Queue

- **W_Q : mean waiting time in just the queue**
  = mean time in system -- mean time in service

  $W_Q = W - 1/m$

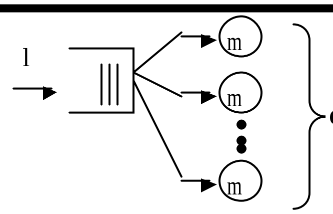
- **L_Q : mean number of customers in the queue**

- **Use Little's Law:**

  $L_Q = l\, W_Q$

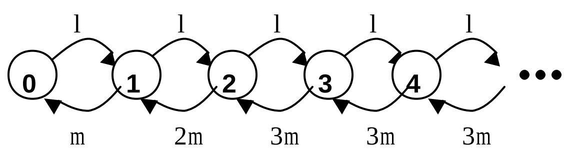
Examples:

Multiple servers: M/M/c



- **Utilization:** $r = \frac{l}{cm} < 1$

- **Basic intuition from birth-death graph (for c=3):**



Basic results for M/M/c

- **(table 11.1 from Askin Standridge)**

Table 11.1 *M/M/c* **Queueing Results**

	$M/M/1$	$M/M/c$
$p(0)$	$1-\rho$	$\left[\frac{(c\rho)^c}{c!(1-\rho)} + \sum_{n=0}^{c-1}\frac{(c\rho)^n}{n!}\right]^{-1}$
L_q	$\frac{\rho^2}{1-\rho}$	$\frac{\rho(c\rho)^c p(0)}{c!(1-\rho)^2}$
L	$\frac{\rho}{1-\rho}$	$L_q + \frac{\lambda}{\mu}$
W_q	$\frac{\rho}{\mu(1-\rho)}$	$\frac{(c\rho)^c p(0)}{c!c\mu(1-\rho)^2}$
W	$\frac{1}{\mu(1-\rho)}$	$W_q + \mu^{-1}$

Example of M/M/c

- **Suppose c=2, m = 5, l = 8**

$$r = \frac{l}{cm} = \frac{8}{10}$$

$$P_0 = \left(\frac{(cr)^c}{c!(1-r)} + \sum_{n=0}^{c-1} \frac{(cr)^n}{n!} \right)^{-1} = \left(\frac{(1.6)^2}{2(.2)} + \left(1 + \frac{1.6}{1}\right)\right)^{-1} = (9)^{-1} = 0.1111$$

$$L_Q = \frac{r(cr)^c P_0}{c!(1-r)^2} = \frac{0.8(1.6)^2(0.111)}{2(.2)^2} = 2.84$$

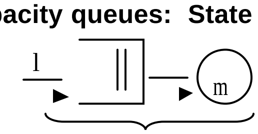
$$L = L_Q + \frac{l}{m} = 2.84 + 1.6 = 4.44$$

$$W = \frac{L}{l} = \frac{4.44}{8} = 0.555$$

$$W_Q = \frac{L_Q}{l} = \frac{4.44}{8} = 0.355$$

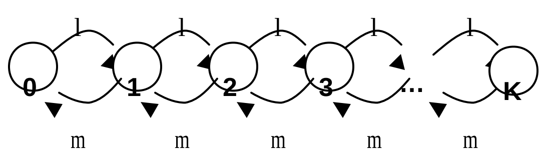
M/M/1/K

- **Finite capacity queues: State probabilities and # in system**



System capacity of K

- **Birth-death model**



$$P_i = r^i P_0$$

$$1 = \sum_{i=0}^{K} P_i = \sum_{i=0}^{K} r^i P_0 \longrightarrow P_0 = \frac{1-r}{1-r^{K+1}}$$

$$L = E[n] = \sum_{n=0}^{K} nP_n = \begin{cases} \frac{r}{1-r} - \frac{(K+1)r^{K+1}}{1-r^{K+1}} & for\ r \neq 1 \\ \frac{K}{2} & for\quad r = 1 \end{cases}$$

- **M/M/1/K queues: mean wait in system:**
  - **Note some customers get turned away.**
  - **For Little's Law, we need *effective arrival rate***

    <u>*offered load*</u> is measure of demand on system: λ/μ

    <u>*carried load*</u> is actual demand met by system: λ_a/μ

$$l_a = effective\ arrival\ rate = l\ (1 - P_K)$$

$$W = \frac{L}{l_a}$$

$$W_Q = W - \frac{1}{m}$$

$$L_Q = W_Q l_a \qquad note\ it\ can\ be\ shown\ this\ is\ L_Q = L - (1 - P_0)$$

M/M/1/K example

- **Suppose K=6, m = 1/8, l = 1/10**

  P_i = ?

M/M/1/6	M/M/1/infinity
P0 = .25307	P0=.2
P1=.20246	P1=.16
P2=.16197	P2=.128
P3=.12957	P3=.1024
P4=.10366	P4=.08192
P5=.08293	P5=.065536
P6=.06634	P6=.0524288
Total P0 to P6: 1.0	Total P0 to P6: .790284

- **Key point: the limited size system turns away parts sometimes since the M/M/1/infinity system has probability (1-.79)= 21% of being in state beyond 6.**

M/G/1 Queues

- **Insert table and equations from page 367**

Example from text: λ = 10/week. All distributions with same mean
Key points: the exponential distribution gives the worst case out the distributions with comparable means. It is thus useful as a worst case (maximum randomness) estimate.

Review:

- **M/M/1 queue:**
  $L = \lambda / (\mu - \lambda) = \rho / (1 - \rho)$
  $W = L / \lambda = 1 / (\mu - \lambda)$
  $W_Q = W - 1/\mu$
  $L_Q = \lambda W_Q$

- **M/M/1/K**
  $L =$

  $\lambda_a = \lambda (1 - P_K)$
  $W = L / \lambda_a$
  $W_Q = W - 1/\mu$
  $L_Q = \lambda_a W_Q$