MIT 2.853/2.854

Introduction to Manufacturing Systems

Manufacturing Systems Overview

Stanley B. Gershwin
Laboratory for Manufacturing and Productivity
Massachusetts Institute of Technology
gershwin@mit.edu

HP Printer Case

Background

- In 1993, the ink-jet printer market was taking off explosively, and manufacturers were competing intensively for market share.
- Manufacturers could sell all they could produce. Demand was much greater than production capacity.
- Hewlett Packard was designing and producing its printers in Vancouver, Washington (near Portland, Oregon).

HP Printer Case

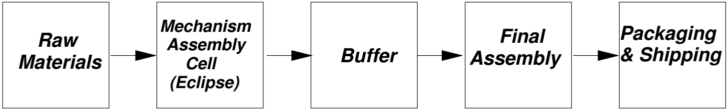
HP's needs

- Maintain quality.
- Meet increased demand *and* increase market share.
  - *Target: 300,000 printers/month.*
  - *Capacity with existing manual assembly: 200,000 printers/month.*
- Meet profit and revenue targets.
- Keep employment stable.

HP Printer Case

Printer Production

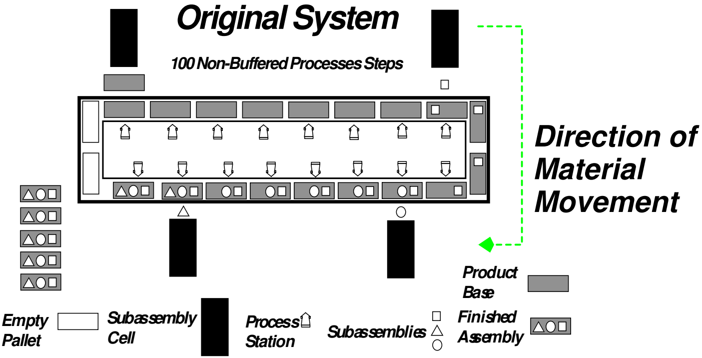
HP invested $25,000,000 in "Eclipse," a new system for automated assembly of the print engine.



Two Eclipses were installed.

HP Printer Case

Printer Production



Design philosophy: minimal — essentially zero — buffer space.

HP Printer Case

The Problem

- Machine efficiencies[1] were estimated to be about .99.
- Operation times were estimated to be 9 seconds, and constant.
    - ⋆ Consequently, the total production rate was estimated to be about 370,000 units/month.
- BUT data was collected when the first two machines were installed:
    - ⋆ Efficiency was less than .99.
    - ⋆ Operation times were variable, often greater than 9 seconds.

Actual production rate would be about 125,000 units/month.

[1] *(to be defined)*

HP Printer Case

The Problem

- HP tried to analyze the system by simulation. They consulted a vendor, but the project appeared to be too large and complex to produce useful results in time to affect the system design.

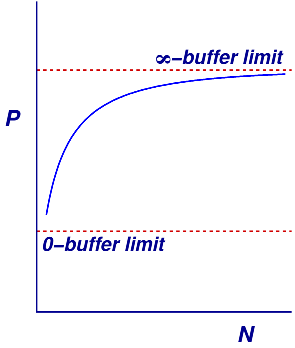
    ⋆ *This was because they tried to include too much detail.*

- Infeasible changes: adding labor, redesigning machines.

HP Printer Case

The Solution

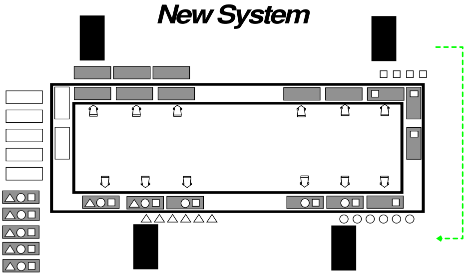
- Feasible change: visiting researcher proposed adding *a small amount* of buffer space within Eclipse.



- Design and analysis tools: *described in the second part of this course.*

HP Printer Case

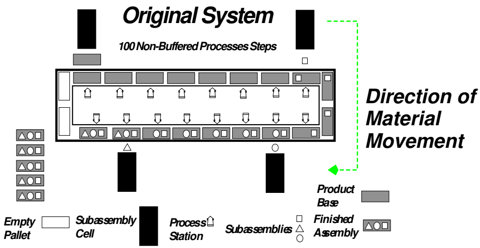
The Solution

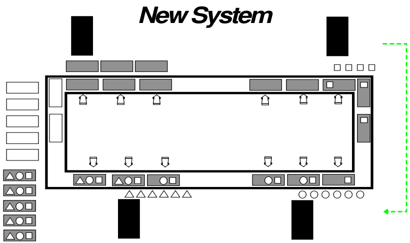


- Empty pallet buffer.
- WIP *(work in process)* space between subassembly lines and main line.
- WIP space on main line.
- Buffer sizes were large enough to hold about 30 minutes worth of material. This is a small multiple of the mean time to repair (MTTR) of the machines.

HP Printer Case

Comparison





Images from "Hewlett-Packard Uses Operations Research to Improve the Design of a Printer Production Line," Interfaces, Volume 28, Number 1, January-February, 1998. pp. 24-26; (c) Copyright 2018 INFORMS. All Rights Reserved.

HP Printer Case

Consequences

- Increased factory capacity — to over 250,000 units/month.
- Capital cost of changes was about $1,400,000.
- Incremental revenues of about $280,000,000.
- Labor productivity increased by about 50%.

HP Printer Case

Reasons for Success

- Early intervention (before too much of the system was built).
- Rapid response by visiting researcher was possible because much research work had already been done.
- HP managers' flexibility.
- The new analysis tool was fast, easy to use, and was at the right level of detail.
    - ⋆ It only dealt with important features of the system.
    - ⋆ It did not require much data.

Course Overview

Message

- Manufacturing systems can be understood like any complex engineered system.
- Engineers must have intuition about these systems in order to design and operate them most effectively.
- Such intuition can be developed by studying the elements of the system and their interactions.
- Using intuition and appropriate design tools can have a big payoff.

Course Overview

Goals

- To explain important measures of system performance.
- To show the importance of random, potentially disruptive events in factories.
- To give some intuition about behavior of these systems.
- To describe and justify some quantitative tools and methods.
- But *not* to describe all current common-sense approaches.

Approach

- To focus on important factory phenomena that can be analyzed quantitatively.
- To develop or describe mathematical models of these phenomena.
- To study the required mathematics, only as deeply as needed.

Problems

- Manufacturing System Engineering (MSE) is not as advanced as other branches of engineering.
- Practitioners are sometimes encouraged to rely on slogans or black boxes.
- A gap exists between theoreticians and practitioners.

Problems

- The research literature is incomplete,
  - ... but practitioners are often unaware of what does exist.
- Terminology, notation, basic assumptions are not standardized.
- There is typically a separation of product, process, and system design.
  - They should be done simultaneously or iteratively, *not* sequentially.

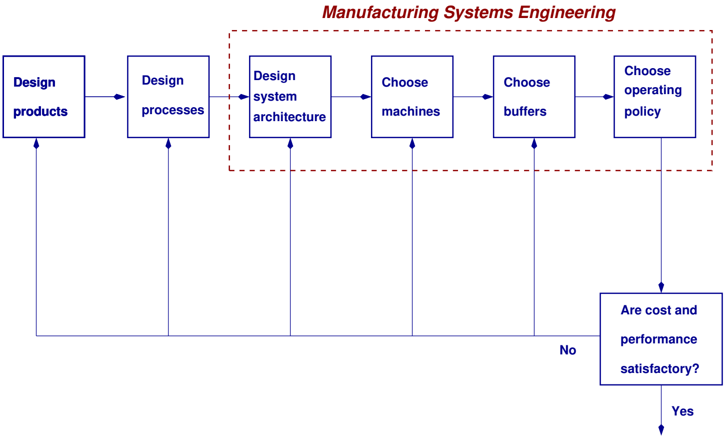
Problems

- Confusion about objectives:

  - ⋆ *maximize capacity?*
  - ⋆ *minimize capacity variability?*
  - ⋆ *maximize capacity utilization?*
  - ⋆ *minimize lead time?*
  - ⋆ *minimize lead time variability?*
  - ⋆ *maximize profit?*

Product Realization

Products, Processes, Machines, Buffers,and Operating Policy



Rule proliferation

- *When a system is not well understood, rules proliferate.*
- This is because rules are developed to regulate behavior.
- But the rules lead to unexpected, undesirable behavior. *(Why?)*
- New rules are developed to regulate the new behavior.
- Et cetera.

Rule proliferation

Example

- A factory starts with one rule: *do the latest jobs first* .
- Over time, more and more jobs are later and later.
- A new rule is added: *treat the highest priority customers' orders as though their due dates are two weeks earlier than they are.*
- The low priority customers find other suppliers, but the factory is still late.
- *Why?*

Rule proliferation

Why?

- There are significant setup times from part family to part family. If setup times are not considered, changeovers will occur too often, and waste capacity.

- Any rules that do not consider setup times in this factory will perform poorly.

Definitions

- *Manufacturing:* the transformation of material into something useful and portable.
- *Manufacturing System:* A manufacturing system is a set of machines, transportation elements, computers, storage buffers, people, and other items that are used together for manufacturing. These items are *resources.*
  - ⋆ Alternate terms:
    - *Factory*
    - *Production system*
    - *Fabrication facility*
- Subsets of manufacturing systems, which are themselves systems, are sometimes called *cells, work centers,* or *work stations* .

Basic Issues

- Frequent new product introductions.
- Product lifetimes often short.
- Process lifetimes often short.

This leads to frequent building and rebuilding of factories.

There is little time for improving the factory after it is built; it must be built right.

Basic Issues

Consequent Needs

- Tools to predict the performance of proposed factory designs.
- Tools for optimal factory design.
- Tools for optimal real-time management (control) of factories.
- Manufacturing Systems Engineering professionals who understand factories as complex systems.

- Design Quality – the design of products that give customers what they want or would like to have *(features).*
  - Examples: Fuel economy in cars. Advanced electronics, attractive styling in cell phones.

- Manufacturing Quality – the manufacturing of products to *avoid* giving customers what they *don't* want or *would not* like to have *(bugs).*
  - Examples: Exploding airbags in cars. Exploding batteries in cell phones.

This course is about manufacturing, *not* product design.

Basic Issues

Quantity, Quality, and Variability

- Quantity – *how much* is produced and *when* it is produced.
- Quality – *how well* it is produced.

In this course, we focus *mostly* on *quantity*.

General Statement: Variability is the enemy of manufacturing.

Basic Issues

Styles for Demand Satisfaction

- Make to Stock (Off the Shelf):
  - items available when a customer arrives
  - appropriate for large volumes, limited product variety, cheap raw materials
- Make to Order:
  - production started only after order arrives
  - appropriate for custom products, low volumes, expensive raw materials

Basic Issues

Conflicting Objectives

- Make to Stock:
  - ⋆ large finished goods inventories needed to prevent stockouts
  - ⋆ small finished goods inventories needed to keep costs low

Basic Issues

Conflicting Objectives

- Make to Order:
  - ⋆ excess production capacity *(low utilization)* needed to allow early, reliable delivery promises
  - ⋆ minimal production capacity *(high utilization)* needed to to keep costs low

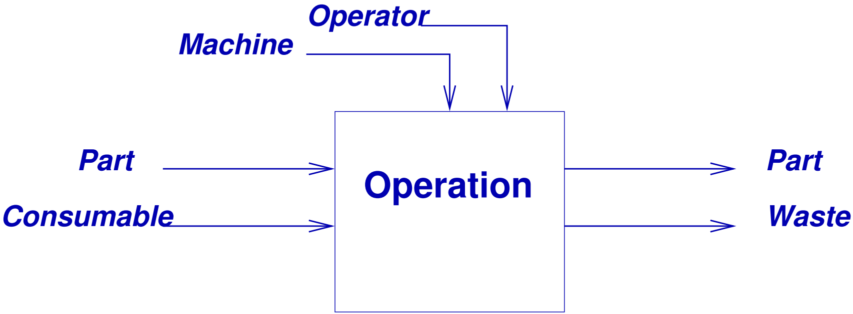
Basic Issues
Concepts

- *Complexity:* collections of things have properties that are non-obvious functions of the properties of the things collected.

- *Non-synchronism (especially randomness) and its consequences:* Factories do not run like clockwork.

Basic Issues

Operation



Nothing happens until everything is present.

Basic Issues

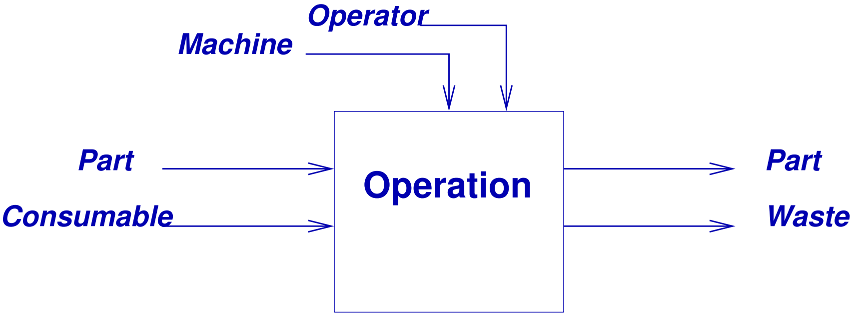
Waiting

Whatever does not arrive last must wait.

- *Inventory:* parts waiting.
- *Under-utilization:* machines waiting.
- *Idle work force:* operators waiting.

Basic Issues

Waiting



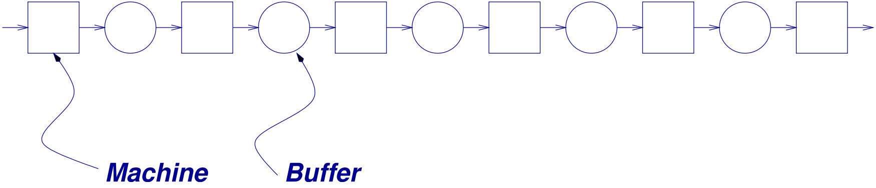
- *Reductions* in the availability, or ...
- *Increased variability* in the availability ...

... of any one of these items increases waiting in the rest of them and reduces performance of the system.

Kinds of Systems

Flow shop

... or *Flow line* , *Transfer line* , or *Production line.*

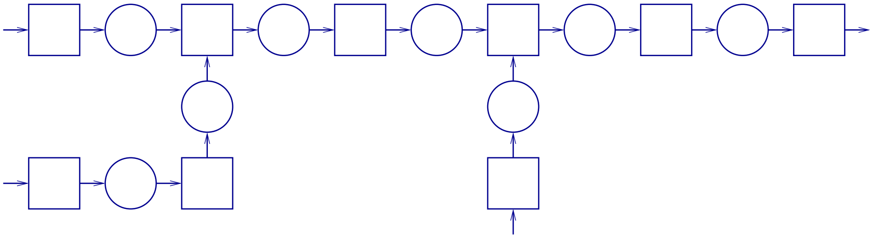


Traditionally used for high volume, low variety production.

What are the buffers for?

Kinds of Systems

Assembly system

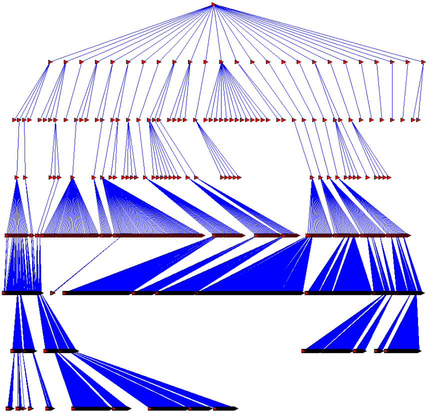


Assembly systems are *trees* , and may involve *thousands* of parts.

Kinds of Systems

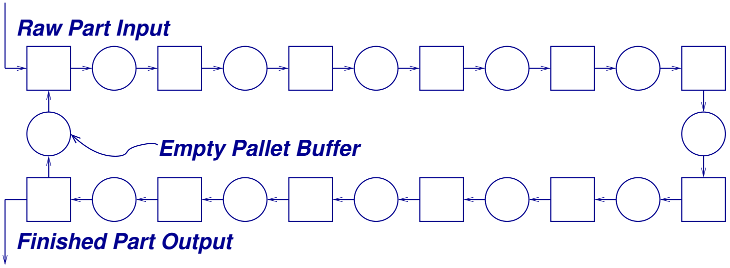
Assembly system

Bill of Materials of a large electronic product



Kinds of Systems — Loops

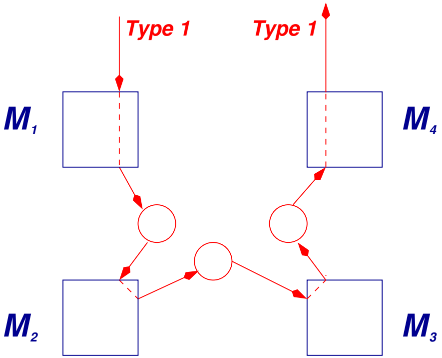
Closed loop *(1)*



Pallets or fixtures travel in a closed loop. Routes are determined. The number of pallets in the loop is constant.

Kinds of Systems — Loops

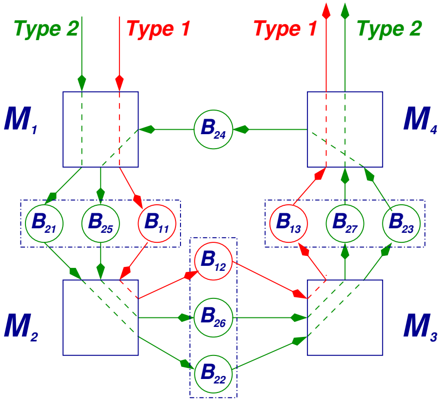
Reentrant loops *(2)*



Kinds of Systems — Loops

Reentrant loops *(2)*

System with reentrant flow and two part types

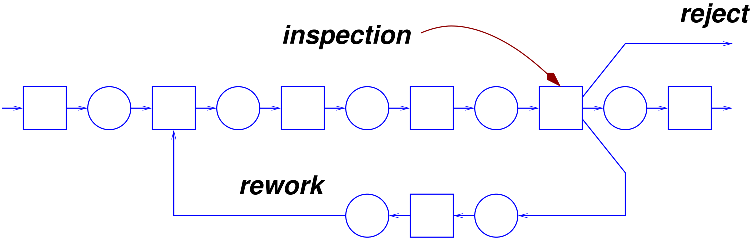


Routes are determined. The number of parts in the loop varies.

Semiconductor fabrication is highly reentrant.

Kinds of Systems — Loops

Rework loop *(3)*



Routes are random. The number of parts in the loop varies.

Kinds of Systems
Job shop

- Machines not organized according to process flow.
- Often, machines grouped by department:
  - mill department
  - lathe department
  - etc.
- Great variety of products.
- Different products follow different paths.
- Complex management.

Time

- Many factory performance measures are about time, such as
  - *production rate:* how much is made in a given time.
  - *lead time:* how much time before delivery.
  - *cycle time:* how much time a part spends in the factory.
  - *delivery reliability:* how often a factory delivers on time.
  - *capital pay-back period:* the time before the company get its investment back.

Time

Even inventory can be described in time units:

"we are holding x weeks of inventory"

means

"customer demand could consume all our inventory in x weeks."

Time

- Time appears in two forms:
  - ⋆ delay
  - ⋆ capacity utilization
- Every action has impact on both.

Time

Delay

- An operation that takes 10 minutes adds 10 minutes to the *delay* that
  - ⋆ a workpiece experiences while undergoing that operation;
  - ⋆ every other workpiece experiences that is waiting while the first is being processed.
- A machine stoppage that lasts 10 minutes adds 10 minutes to the delay that
  - ⋆ every workpiece that is waiting to be processed at that machine experiences.
  - ⋆ Machine stoppages are caused by failures, maintenance, blocking, starvation, set-up changes and other causes.
- The sum of all the delays that a part experiences during production is the extra time that a part spends in a factory beyond the time required for its operations.
  - ⋆ That is often between 10 and 100 times the total operation time.

Time

Capacity Utilization

- An operation that takes 10 minutes takes up 10 minutes of the available time of
  - ⋆ a machine,
  - ⋆ an operator,
  - ⋆ or other resources.
- Similarly for machine stoppages.
- Since there are a limited number of minutes of each resource that are available in a day, there are a limited number of operations that can be done in a day.
- In other words, this is the limit on the factory's production rate.

Time

Production Rate

- *Operation Time:* the time that a machine takes to do an operation.

- *Production Rate:* the average number of parts produced in a time unit. (Also called *throughput.*)

If nothing interesting ever happens (no failures, etc.),

$$\text{Production rate} = \frac{1}{\text{operation time}}$$

... but something interesting *always* happens.

- *Capacity:* the maximum possible production rate of a manufacturing system, for systems that are making only one part type.
  - ⋆ *Short term capacity:* determined by the resources available right now.
  - ⋆ *Long term capacity:* determined by the average resource availability.
- Capacity is harder to define for systems making more than one part type. Since it is hard to define, it is *very* hard to calculate.

MIT 2.853/2.854

Introduction to Manufacturing Systems

Probability

Stanley B. Gershwin
Laboratory for Manufacturing and Productivity
Massachusetts Institute of Technology
gershwin@mit.edu

I flip a coin 100 times, and it shows heads every time.

Question: What is the probability that it will show heads on the next flip?

Probability and Statistics

Another Trick Question

I flip a coin 100 times, and it shows heads every time.

Question: *How much would you bet* that it will show heads on the next flip?

Still Another Trick Question

I flip a coin 100 times, and it shows heads every time.

Question: *What odds would you demand before you bet* that it will show heads on the next flip?

Probability and Statistics

Probability $\neq$ Statistics

Probability: mathematical theory that describes uncertainty.

Statistics: set of techniques for extracting useful information from data.

Interpretations of probability

Frequency

The probability that the outcome of an experiment is A is P...

if the experiment is performed a large number of times and the fraction of times that the observed outcome is A is P.

Interpretations of probability
State of belief

The probability that the outcome of an experiment is A is P...

if that is the opinion (ie, belief or state of mind) of an observer *before* the experiment is performed.

Example of State of Belief: Betting odds

The probability that the outcome of an experiment is A is P . . .

if *before the experiment is performed* a risk-neutral observer would be willing to bet \$1 against more than $\$\frac{1-P}{P}$.

The expected value (slide **??**) of the bet is greater than

$$(1-P)\times(-1)+(P)\times\left(\frac{1-P}{P}\right)=0$$

Interpretations of probability

Abstract measure

The probability that the outcome of an experiment is A is $P(A)$...

if $P()$ satisfies a certain set of conditions: *the axioms of probability.*

Interpretations of probability

Axioms of probability

Let U be a set of *samples* . Let $\mathcal{E}_1$, $\mathcal{E}_2$, ... be subsets of U.

Let $\emptyset$ be the *null* (or *empty*) *set* , the set that has no elements.

- $0 \leq P(\mathcal{E}_i) \leq 1$
- $P(U) = 1$
- $P(\emptyset) = 0$
- If $\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$, then $P(\mathcal{E}_i \cup \mathcal{E}_j) = P(\mathcal{E}_i) + P(\mathcal{E}_j)$

Probability Basics

Discrete Sample Space

Notation, terminology:

- ω is often used as the symbol for a generic sample.
- Subsets of U are called *events.*
- $P(\mathcal{E})$ is the *probability* of $\mathcal{E}$.

Probability Basics

Discrete Sample Space

- *Example:* Throw a single die. The possible outcomes are $\{1, 2, 3, 4, 5, 6\}$. ω can be any one of those values.

- *Example:* Consider $n(t)$, the number of parts in inventory at time t. Then

$$\omega = \{n(1), n(2), ..., n(t), ...\}$$

is a *sample path.*

Probability Basics

Discrete Sample Space

- An event can often be defined by a statement. For example,

$$\mathcal{E} = \{\text{There are 6 parts in the buffer at time } t = 12\}$$

Formally, this can be written

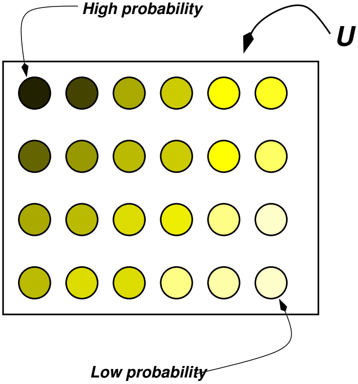
$$\mathcal{E} = \text{the set of all } \omega \text{ such that } n(12) = 6$$

or,

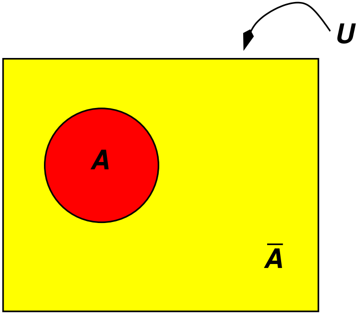
$$\mathcal{E} = \{\omega | n(12) = 6\}$$

Probability Basics

Discrete Sample Space



Venn diagrams

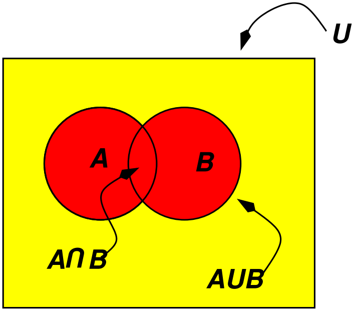


$$P(\bar{A}) = 1 - P(A)$$

Probability Basics

Set Theory

Venn diagrams



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Probability Basics

Independence

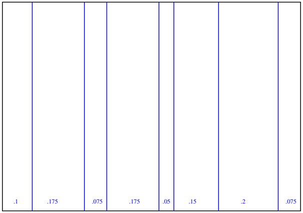
A and B are *independent* if

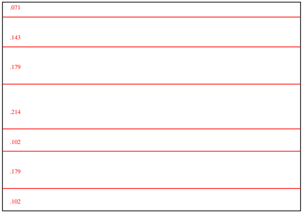
$$P(A \cap B) = P(A)P(B).$$

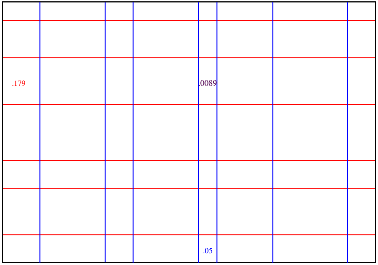
grid figure to illustrate independence

Probability Basics

Independence





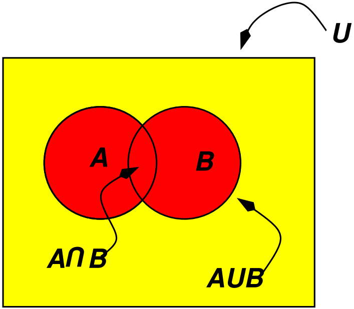


Probability Basics

Conditional Probability

If $P(B) \neq 0$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



We can also write $P(A \cap B) = P(A|B)P(B)$.

Probability Basics
Conditional Probability

$$P(A|B) = P(A \cap B)/P(B)$$

Example: Throw a die.Let

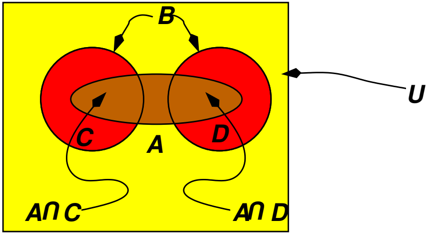
- A is the event of getting an odd number (1, 3, 5).
- B is the event of getting a number less than or equal to 3 (1, 2, 3).

Then $P(A) = P(B) = 1/2, P(A \cap B) = P(1, 3) = 1/3$.

Also, $P(A|B) = P(A \cap B)/P(B) = 2/3$.

Probability Basics

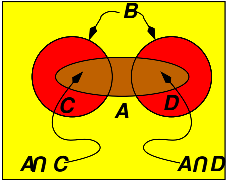
Law of Total Probability



- Let $B = C \cup D$ and assume $C \cap D = \emptyset$. Then

$$P(A|C) = \frac{P(A \cap C)}{P(C)} \text{ and } P(A|D) = \frac{P(A \cap D)}{P(D)}.$$

Probability Basics



Also,

- $P(C|B) = \dfrac{P(C \cap B)}{P(B)} = \dfrac{P(C)}{P(B)}$ because $C \cap B = C$.

  Similarly, $P(D|B) = \dfrac{P(D)}{P(B)}$

- $A \cap B = A \cap (C \cup D) = (A \cap C) \cup (A \cap D)$

- Therefore

  $P(A \cap B) = P(A \cap (C \cup D))$

  $\qquad = P(A \cap C) + P(A \cap D)$ because $(A \cap C)$ and $(A \cap D)$ are disjoint.

Probability Basics
Law of Total Probability

- Or, from the definition of conditional probability,

  $P(A|B)P(B) = P(A|C)P(C) + P(A|D)P(D)$

  or,

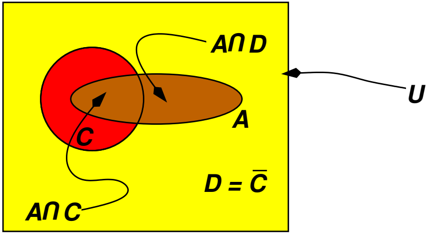
  $$\frac{P(A|B)P(B)}{P(B)} = \frac{P(A|C)P(C)}{P(B)} + \frac{P(A|D)P(D)}{P(B)}$$

  or,

  $$P(A|B) = P(A|C)P(C|B) + P(A|D)P(D|B)$$

Probability Basics

Law of Total Probability

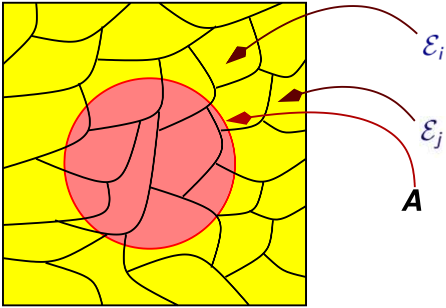


An important case is when $C \cup D = B = U$, so that $A \cap B = A$. Then $P(A) = P(A \cap C) + P(A \cap D)$ or

$$\boxed{P(A) = P(A|C)P(C) + P(A|D)P(D)}$$

Probability Basics

Law of Total Probability



More generally, if A and $\mathcal{E}_1, \ldots \mathcal{E}_k$ are events and

$$\mathcal{E}_i \text{ and } \mathcal{E}_j = \emptyset, \text{ for all } i \neq j$$

and

$$\bigcup_j \mathcal{E}_j = \text{ the universal set}$$

(ie, the set of $\mathcal{E}_j$ sets is *mutually exclusive* and *collectively exhaustive*) then ...

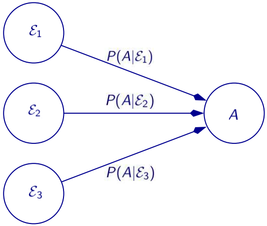
Probability Basics

Law of Total Probability

$$\sum_j P(\mathcal{E}_j) = 1$$

and

$$\boxed{P(A) = \textstyle\sum_j P(A|\mathcal{E}_j)P(\mathcal{E}_j).}$$



Example

$A = \{$I will have a cold tomorrow.$\}$
$\mathcal{E}_1 = \{$It is raining today.$\}$
$\mathcal{E}_2 = \{$It is snowing today.$\}$
$\mathcal{E}_3 = \{$It is sunny today.$\}$

(Assume $\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 = U$ and $\mathcal{E}_1 \cap \mathcal{E}_2 = \mathcal{E}_1 \cap \mathcal{E}_3 = \mathcal{E}_2 \cap \mathcal{E}_3 = \emptyset$.)

Then $A \cap \mathcal{E}_1 = \{$I will have a cold tomorrow *and* it is raining today$\}$.
And $P(A|\mathcal{E}_1)$ is the probability I will have a cold tomorrow *given* that it is raining today.

etc.

Probability Basics
Law of Total Probability

Then

{I will have a cold tomorrow.}=
{I will have a cold tomorrow *and* it is raining today} $\cup$
{I will have a cold tomorrow *and* it is snowing today} $\cup$
{I will have a cold tomorrow *and* it is sunny today}

so

P({I will have a cold tomorrow.})=
P({I will have a cold tomorrow *and* it is raining today}) $+$
P({I will have a cold tomorrow *and* it is snowing today}) $+$
P({I will have a cold tomorrow *and* it is sunny today})

Probability Basics

Law of Total Probability

$P(\{\text{I will have a cold tomorrow.}\}) =$

$P(\{\text{I will have a cold tomorrow} \mid \text{it is raining today}\}) P(\{\text{it is raining today}\}) +$

$P(\{\text{I will have a cold tomorrow} \mid \text{it is snowing today}\}) P(\{\text{it is snowing today}\}) +$

$P(\{\text{I will have a cold tomorrow} \mid \text{it is sunny today}\})\ P(\{\text{it is sunny today}\})$

or

$$P(A) = P(A|\mathcal{E}_1)P(\mathcal{E}_1) + P(A|\mathcal{E}_2)P(\mathcal{E}_2) + P(A|\mathcal{E}_3)P(\mathcal{E}_3)$$

Probability Basics

Random Variables

Let V be a vector space. Then a *random variable* X is a mapping (a function) from U to V.

If $\omega \in U$ and $x = X(\omega) \in V$, then X is a random variable.

Example: V could be the real number line.

Typical notation :

- Upper case letters (X) are usually used for random variables and corresponding lower case letters (x) are usually used for possible values of random variables.
- Random variables ($X(\omega)$) are usually not written as functions; the argument (ω) of the random variable is usually not written. *This sometimes causes confusion.*

Probability Basics

Random Variables

Flip of a Coin

Let U={H,T}. Let $\omega =$ H if we flip a coin and get heads; $\omega =$ T if we flip a coin and get tails.

Let V be the real number line.

$X(T) = 0$
$X(H) = 1$

Assume the coin is fair. *(No tricks this time!)* Then
$P(\omega = \text{ T }) = P(X = 0) = 1/2$
$P(\omega = \text{ H }) = P(X = 1) = 1/2$

Probability Basics

Random Variables

Flip of Three Coins

Let $U = \{$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$\}$.

Let $\omega =$ HHH if we flip 3 coins and get 3 heads; $\omega =$HHT if we flip 3 coins and get 2 heads and *then* one tail, etc. *The order matters!* There are 8 samples.

- $P(\omega) = 1/8$ for all ω.

Let X be the *number* of heads. Then $X = 0$, 1, 2, or 3.

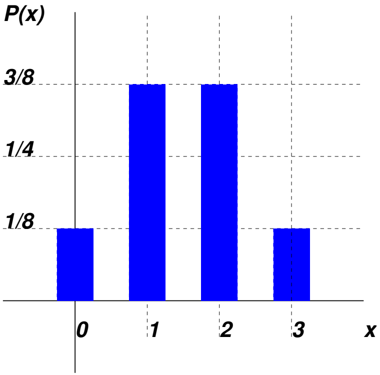
- $P(X = 0)$=1/8; $P(X = 1)$=3/8; $P(X = 2)$=3/8; $P(X = 3)$=1/8.

There are 4 distinct values of X.

Probability Basics

Probability Distributions

Let $X(\omega)$ be a random variable. Then $P(X(\omega) = x)$ is the *probability distribution* of X (usually written $P(x)$). For three coin flips:



Probability Basics

Probability Distributions

Shorthand:

- Instead of writing $P(X(\omega) = x)$, people often write $P(x)$ if the meaning is unambiguous.

Mean and Variance:

- *Mean (average):* $\bar{x} = \mu_x = E(X) = \sum_x xP(x)$
- *Variance:* $V_x = \sigma_x^2 = E(X - \mu_x)^2 = \sum_x (x - \mu_x)^2 P(x)$
- *Standard deviation (sd):* $\sigma_x = \sqrt{V_x}$
- *Coefficient of variation (cv):* σ_x / μ_x

Probability Basics

Probability Distributions

For three coin flips:

$$\bar{x} = 1.5$$

$$V_x = 0.75$$

$$\sigma_x = 0.866$$

$$\text{cv} = 0.577$$

- A function of a random variable is a random variable.

- *Special case: linear function*

  For every ω, let $Y(\omega) = aX(\omega) + b$. Then

  - $\bar{Y} = a\bar{X} + b$.
  - $V_Y = a^2 V_X$; $\sigma_Y = |a|\sigma_X$.

Discrete Random Variables

1. Bernoulli

Flip a biased coin.

X^B is 1 if outcome is heads; 0 if tails.

Let p be a real number, $0 \leq p \leq 1$.

$P(X^B = 1) = p$.

$P(X^B = 0) = 1 - p$.

X^B is a *Bernoulli random variable.*

Discrete Random Variables

2. Binomial

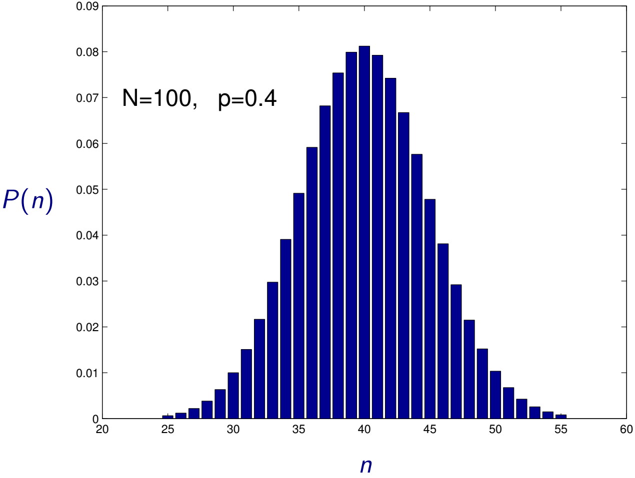
The sum of N independent Bernoulli random variables X_i^B with the same parameter p is a *binomial* random variable X^b.

$$X^b = \sum_{i=0}^{N} X_i^B$$

$$P(X^b = x) = \frac{N!}{x!(N-x)!} p^x (1-p)^{(N-x)}$$

Discrete Random Variables

2. Binomial probability distribution



Discrete Random Variables

3. Geometric

The number of independent Bernoulli random variables X_i^B with the same parameter p tested *until the first 1 appears* is a *geometrically distributed* random variable X^g.

$$
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & \ldots & k-4 & k-3 & k-2 & k-1 & k \\
0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \mathbf{1} \\
\multicolumn{10}{c}{\longleftarrow \qquad k \qquad \longrightarrow}
\end{array}
$$

$$
\boxed{X^g = k \text{ if } X_1^B = 0,\ X_2^B = 0,\ \ldots,\ X_{k-1}^B = 0,\ X_k^B = 1}
$$

Discrete Random Variables

3. Geometric

To calculate $P(X^g = k)$, observe that $P(X^g = 1) = p$, so $P(X^g > 1) = 1 - p$. Also, observe that $\{X^g > k\}$ is a subset of $\{X^g > k-1\}$.

Then

$$\begin{aligned} P(X^g > k) &= P(X^g > k | X^g > k-1) P(X^g > k-1) \\ &= (1-p) P(X^g > k-1), \end{aligned}$$

because

$$P(X^g > k | X^g > k-1) = P(X_1^B = 0, \ldots, X_k^B = 0 \ | X_1^B = 0, \ldots, X_{k-1}^B = 0)$$
$$= 1 - p$$

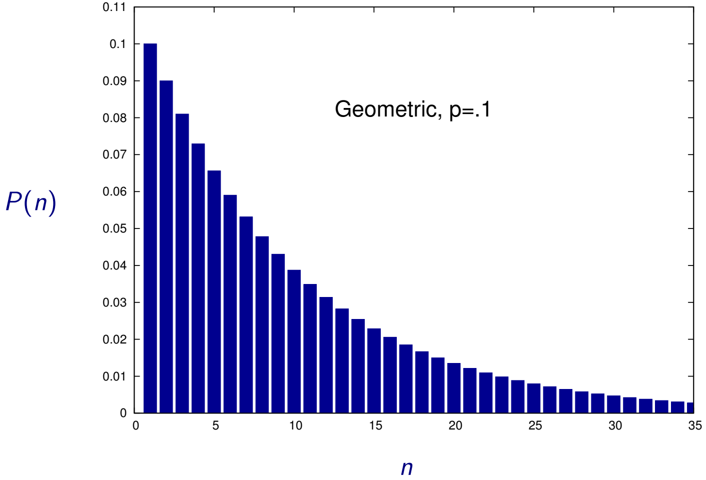
so

$$P(X^g > 1) = 1 - p, \ P(X^g > 2) = (1-p)^2, \ \ldots \ P(X^g > k-1) = (1-p)^{k-1}$$

and $P(X^g = k) = P(\{X^g > k-1\}$ and $\{X_k^B = 1\}) = (1-p)^{k-1} p$.

Discrete Random Variables

3. Geometric probability distribution



Discrete Random Variables

4. Poisson Distribution

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

Discussion later.

Continuous Random Variables

Philosophical Issues

1. *Mathematically* , continuous and discrete random variables are very different.
2. *Quantitatively* , however, some continuous models are very close to some discrete models.
3. Therefore, which kind of model to use for a given system is a matter of *convenience* .

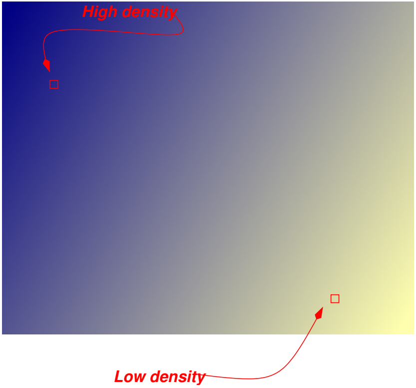
Continuous Random Variables

Philosophical Issues

Example: The production process for small metal parts (nuts, bolts, washers, etc.) might better be modeled as a continuous flow than as a large number of discrete parts.

Continuous Random Variables

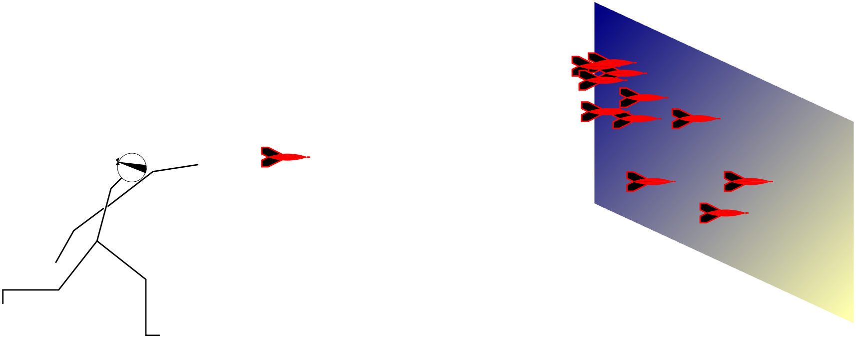
Philosophical Issues



The probability of a two-dimensional random variable being in a small square is the *probability density* times the area of the square. (The definition is similar in higher-dimensional spaces.)

Compare with slide **??**.

Continuous Random Variables
Philosophical Issues



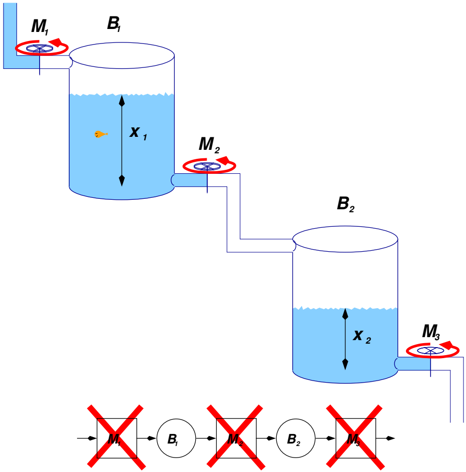
Continuous Random Variables
Spaces

Dimensionality

- Continuous random variables can be defined
  - ⋆ in one, two, three, ..., infinite dimensional spaces;
  - ⋆ in finite or infinite regions of the spaces.
- Continuous random variables can have
  - ⋆ probability measures with the same dimensionality as the space;
  - ⋆ lower dimensionality than the space;
  - ⋆ a mix of dimensions.

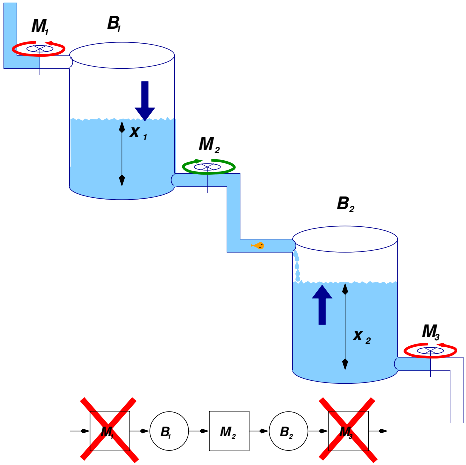
Continuous Random Variables

No change in water levels



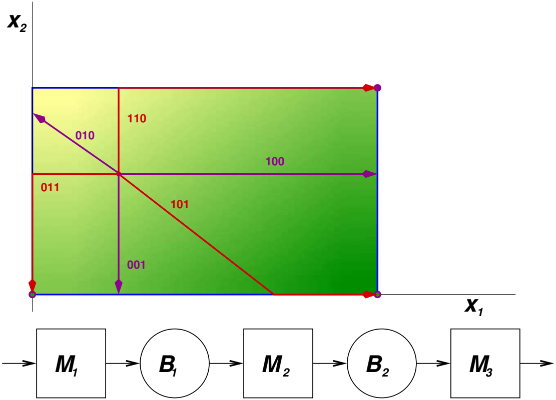
Continuous Random Variables

One kind of change in water levels



Continuous Random Variables

Trajectories

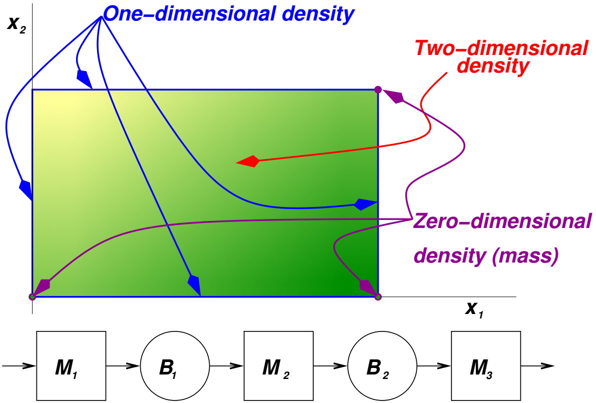


Trajectories of buffer levels in the three-machine line if the machine states stay constant for a long enough time period.

Notation: 110 means M_1 and M_2 are operational and M_3 is down, 100 means M_1 is operational, M_2 and M_3 are down, etc.

Continuous Random Variables

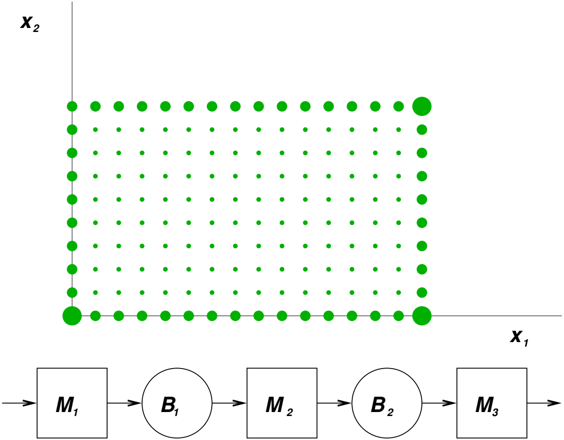
Two-dimensional probability distribution



Probability distribution of the amount of material in each of the two buffers.

Continuous Random Variables

Discrete approximation of the probability distribution



Probability distribution of the amount of material in each of the two buffers.

Continuous Random Variables

Densities and Distributions

In one dimension, $F()$ is the *cumulative probability distribution of X* if

$$F(x) = P(X \leq x)$$

$f()$ is the *density function of X* if

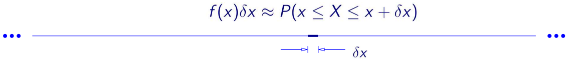
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Therefore,

$$f(x) = \frac{dF}{dx}$$

wherever F is differentiable.

Fact: $f(x)\delta x \approx P(x \leq X \leq x + \delta x)$ for sufficiently small δx.



Fact: $F(b) - F(a) = \int_a^b f(t)dt$

Definition: Expected value of $x = \bar{x} = \int_{-\infty}^{\infty} tf(t)dt$

Continuous Random Variables

Standard Normal Distribution

The density function of the *normal* (or *gaussian*) distribution with mean 0 and variance 1 (the *standard normal*) is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

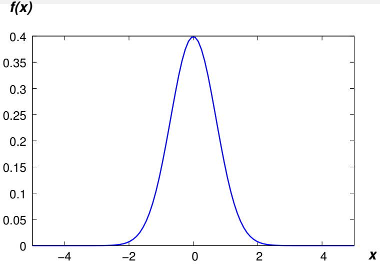
The *normal distribution function* is

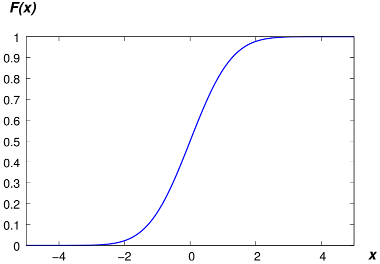
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

(There is no closed form expression for $F(x)$.)

Continuous Random Variables

Standard Normal Distribution





Continuous Random Variables

Normal Distribution

Notation: $N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

Note: *Some people write $N(\mu, \sigma)$ for the normal distribution with mean μ and variance σ^2.*

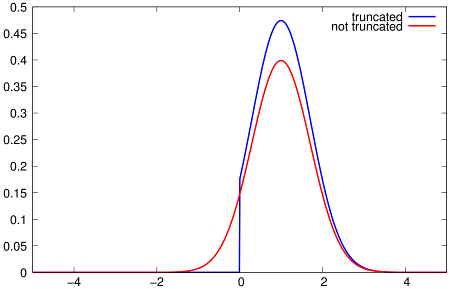
Fact: If X and Y are normal, then $aX + bY + c$ is normal.

Fact: If X is $N(\mu, \sigma)$, then $\frac{X-\mu}{\sigma}$ is $N(0, 1)$, the standard normal.

Consequently, $N(\mu, \sigma)$ easy to compute from $N(0, 1)$. This is why $N(0, 1)$ is tabulated in books.

Continuous Random Variables

Truncated Normal Density (1)

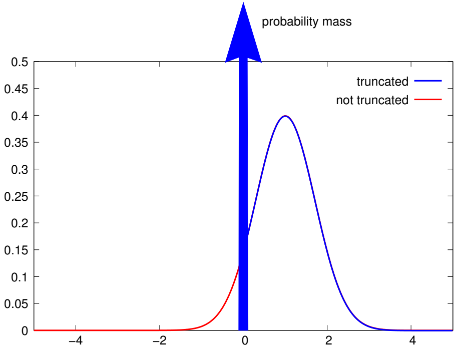


$f_T(x)\delta x = P(x \leq X \leq x + \delta x) = \dfrac{f(x)}{1 - F(0)}\delta x$ where $F()$ and $f()$ are the normal distribution and density functions with parameters μ and σ.

Note: μ and σ are the parameters of $f(x)$, *not* $f_T(x)$.

Continuous Random Variables

Truncated Normal Density (2)



$f_{T'}(x)\delta x = P(x \leq X \leq x + \delta x) = f(x)\delta x$ for $x > 0$ and $P(X = 0) = F(0)$ where $F()$ and $f()$ are the normal distribution and density functions with parameters μ and σ.

Here again, μ and σ are the parameters of $f(x)$, *not* $f_{T'}(x)$.

For *both* kinds of truncation, $f_T(x)$ and $f_{T'}(x)$ are close to $f(x)$ when $\mu \gg \sigma$, and not otherwise.

Continuous Random Variables

Law of Large Numbers

Let $\{X_k\}$ be a sequence of independent identically distributed *(i.i.d.)* random variables that have finite mean μ. Let S_n be the sum of the first n X_ks, so

$$S_n = X_1 + ... + X_n$$

Then for every $\epsilon > 0$,

$$\lim_{n \to \infty} P\left(\left| \frac{S_n}{n} - \mu \right| > \epsilon \right) = 0$$

That is, *the average approaches the mean.*

Continuous Random Variables

Central Limit Theorem

Let $\{X_k\}$ be a sequence of i.i.d. random variables with finite mean μ and finite variance σ^2.

Then as $n \to \infty$, $P(\frac{S_n - n\mu}{\sqrt{n}\sigma}) \to N(0,1)$.

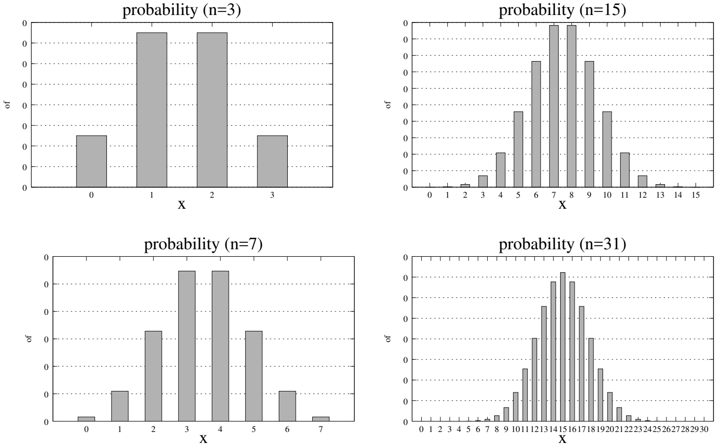
If we define A_n as S_n/n, the average of the first n X_ks, then this is equivalent to:

As $n \to \infty$, $P(A_n) \to N(\mu, \sigma/\sqrt{n})$.

Continuous Random Variables

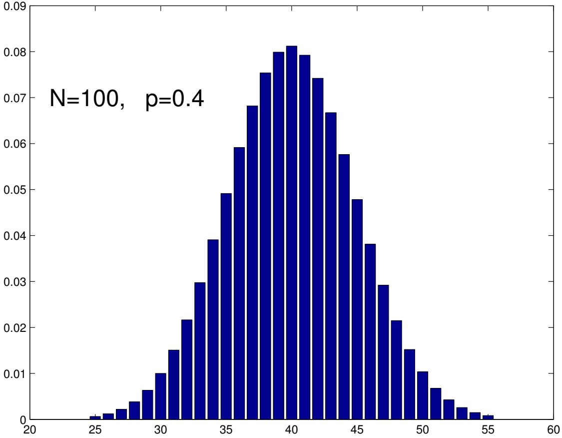
Coin flip examples

Probability of x heads in n flips of a fair coin



Continuous Random Variables

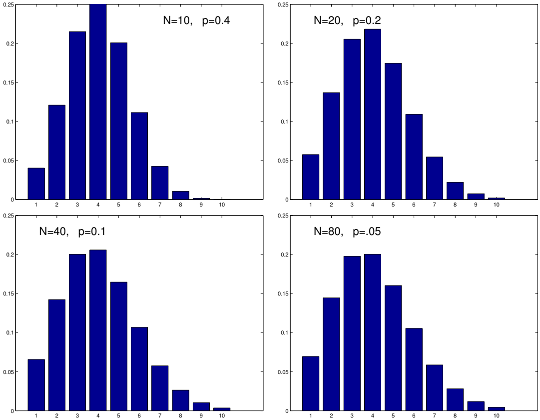
Binomial probability distribution approaches normal for large N.



Continuous Random Variables

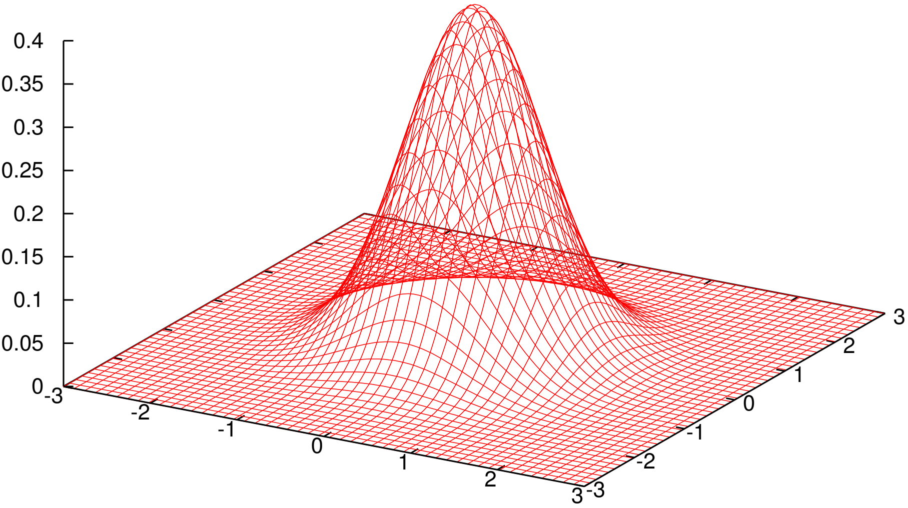
Binomial distributions

Note the resemblance to a *truncated* normal in these examples.



Normal Density Function

... in Two Dimensions

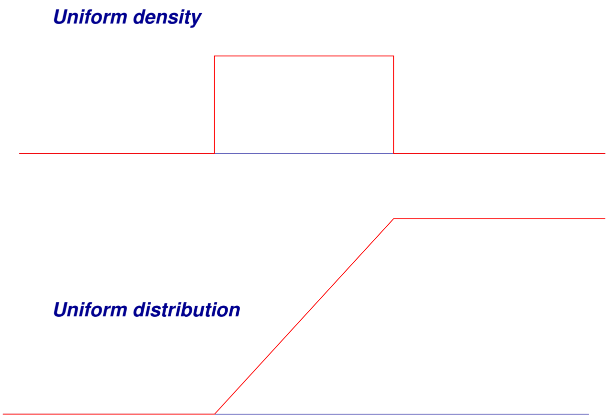


More Continuous Distributions

Uniform

$$f(x) = \frac{1}{b-a} \quad \text{for } a \le x \le b$$

$$f(x) = 0 \qquad\qquad \text{otherwise}$$

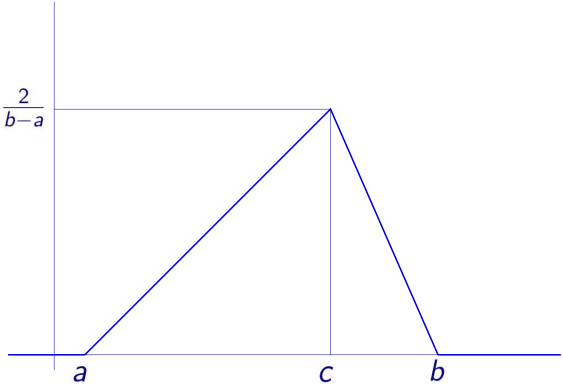


Uniform density
Uniform distribution

More Continuous Distributions

Triangular

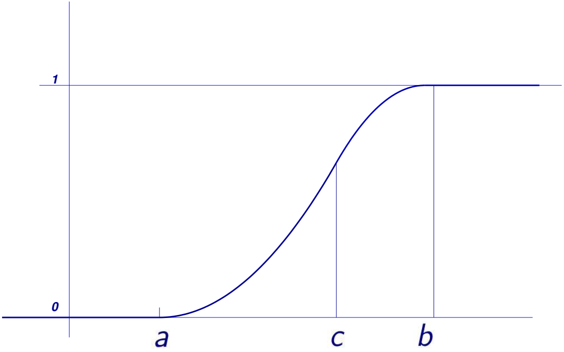
Probability density function



More Continuous Distributions

Triangular

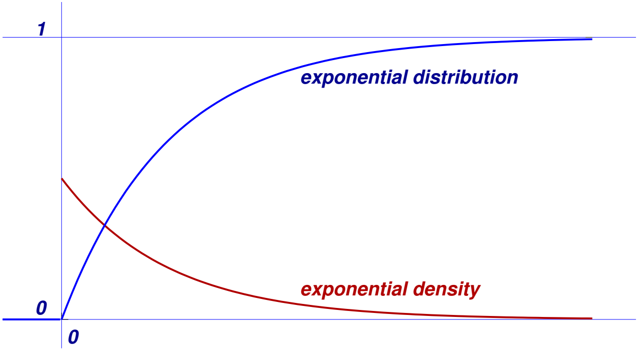
Cumulative distribution function



More Continuous Distributions

Exponential

- Very often used for the time until a specified event occurs.
- **Density:** $f(t) = \lambda e^{-\lambda t}$ **for** $t \geq 0$; $f(t) = 0$ **otherwise;**
- Distribution: $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$ for $t \geq 0$; $F(t) = 0$ otherwise.



More Continuous Distributions

Exponential

- Close to the geometric distribution but for continuous time.
- *Very* mathematically convenient.
- Memorylessness:

  $$P(T > t + x | T > x) = P(T > t)$$

  Suppose an exponentially distributed process is started at time 0 and the event of interest has not occurred yet at time x. Then the probability distribution of the time after x at which it occurs is the same as the original exponential distribution. The process has no "memory" of when it was actually started.

Another Discrete Random Variable

Poisson Distribution

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

is the probability that x events happen in $[0, t]$ if the events are independent and the times between them are exponentially distributed with parameter λ.

Typical examples: arrivals and services at queues. *(Next lecture!)*

NOT Random

...but almost

- A *pseudo-random number generator* is a set of numbers $X_0, X_1, \ldots$ where there is a function F such that

$$X_{n+1} = F(X_n)$$

  and F is such that the sequence of X_n satisfies certain conditions.

- For example,
  - ⋆ there is a known finite maximum $X^{\max}$,
  - ⋆ $0 \leq X_n \leq X^{\max}$,
  - ⋆ and the sequence $U_0, U_1, \ldots$ (where $U_i = X_i / X^{\max}$) *looks like* a set of uniformly distributed, independent random variables.
    - ▶ That is, statistical tests say that the probability of the sequence *not* being independent uniform random variables is very small.

NOT Random

...but almost

- The sequence is deterministic: it is determined by X_0, the *seed* of the random number generator.
- If you use the same seed twice, you get the same sequence both times. This can be convenient, especially in development of software.
- If you use different seeds, you get completely different sequences, even if the seeds are close to one another.
- Pseudo-random number generators are used extensively in *simulation*.

MIT 2.853/2.854

Introduction to Manufacturing Systems

Markov Processes and Queues

Stanley B. Gershwin
Laboratory for Manufacturing and Productivity
Massachusetts Institute of Technology
gershwin@mit.edu

Stochastic processes

- t is time.
- $X()$ is a *stochastic process* if $X(t)$ is a random variable for every t.
- t is a scalar — it can be discrete or continuous.
- $X(t)$ can be discrete or continuous, scalar or vector.

Stochastic processes
Markov processes

- A *Markov process* is a stochastic process in which the probability of finding X at some value at time $t + \delta t$ depends only on the value of X at time t.

- Or, let $x(s), s \leq t$, be the history of the values of X before time t and let A be a possible value of X.

$$P\{X(t+\delta t) = A | X(s) = x(s), s \leq t\} =$$
$$P\{X(t+\delta t) = A | X(t) = x(t)\}$$

Stochastic processes

Markov processes

- In words: if we know what X was at time t, we don't gain any more useful information about $X(t + \delta t)$ by *also* knowing what X was at any time earlier than t.

- *This is **ONLY** the definition of a class of mathematical models. It is NOT a statement about reality!!* That is, not everything is a Markov process.

Markov processes
Example

Example:

- I have \$100 at time $t = 0$.
- At every time $t \geq 1$, I have \$$N(t)$.
  - ⋆ A (possibly biased) coin is flipped.
  - ⋆ If it lands with H showing, $N(t+1) = N(t) + 1$.
  - ⋆ If it lands with T showing, $N(t+1) = N(t) - 1$.

$N(t)$ is a Markov process. *Why?*

Discrete state, discrete time

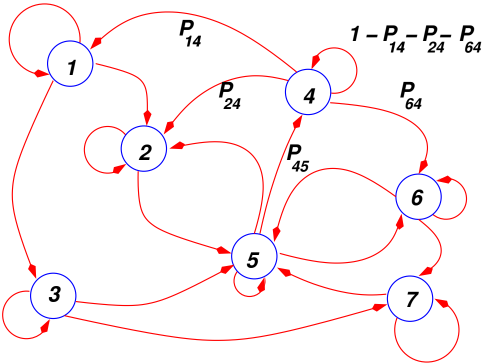
States and transitions

- States can be numbered 0, 1, 2, 3, ... *(or with multiple indices if that is more convenient)*.
- Time can be numbered 0, 1, 2, 3, ... *(or 0, Δ, 2Δ, 3Δ, ... if more convenient)*.
- The probability of a transition from j to i in one time unit is often written P_{ij}, where

  $$P_{ij} = P\{X(t+1) = i | X(t) = j\}$$

States and transitions
Transition graph

Transition graph

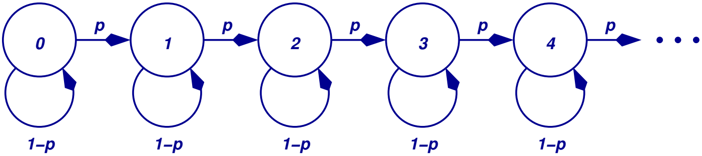


P_{ij} is a probability. Note that $P_{ii} = 1 - \sum_{m, m \neq i} P_{mi}$. This is the *self-loop* probability.

States and transitions
Transition graph

Example : $H(t)$ is the number of Hs after t coin flips.

Assume probability of H is p.

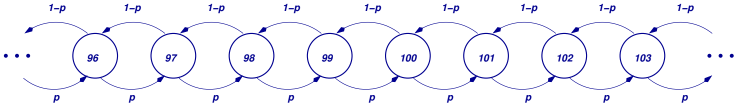


This is a system with an infinite state space.

States and transitions
Transition graph

Example : Coin flip bets on Slide 5.

Assume probability of H is p.



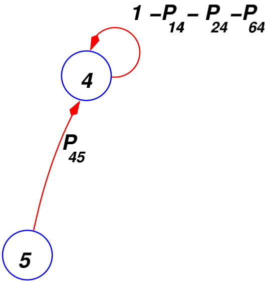
Markov processes

Notation

- $\{X(t) = i\}$ is the event that random quantity $X(t)$ has value i.
  - ⋆ *Example:* $X(t)$ is any state in the graph on Slide 7. i is a *particular* state.
- Define $\pi_i(t) = P\{X(t) = i\}$.
- Normalization equation: $\sum_i \pi_i(t) = 1$.

Markov processes
Transition equations

Transition equations: application of the law of total probability.



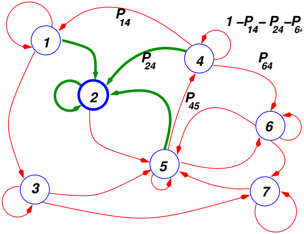
(Detail of graph on slide 7.)

$$\pi_4(t+1) = \pi_5(t)P_{45} + \pi_4(t)(1 - P_{14} - P_{24} - P_{64})$$

(Remember that
$$P_{45} = P\{X(t+1) = 4 | X(t) = 5\},$$
$$P_{44} = P\{X(t+1) = 4 | X(t) = 4\} = 1 - P_{14} - P_{24} - P_{64}$$
)

Markov processes

Transition equations



$$P\{X(t+1)=2\}$$

$$= P\{X(t+1)=2|X(t)=1\}P\{X(t)=1\}$$

$$+P\{X(t+1)=2|X(t)=2\}P\{X(t)=2\}$$

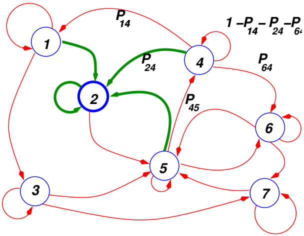
$$+P\{X(t+1)=2|X(t)=4\}P\{X(t)=4\}$$

$$+P\{X(t+1)=2|X(t)=5\}P\{X(t)=5\}$$

Markov processes
Transition equations

- Define $P_{ij} = P\{X(t+1) = i | X(t) = j\}$
- Transition equations: $\pi_i(t+1) = \sum_j P_{ij}\pi_j(t)$.
  An application of the *(Law of Total Probability)*
- Normalization equation: $\sum_i \pi_i(t) = 1$.

Markov processes
Transition equations



Therefore, since

$$P_{ij} = P\{X(t+1) = i | X(t) = j\} \text{ and}$$

$$\pi_i(t) = P\{X(t) = i\},$$

we can write

$$\pi_2(t+1) = P_{21}\pi_1(t) + P_{22}\pi_2(t) + P_{24}\pi_4(t) + P_{25}\pi_5(t).$$

Note that $P_{22} = 1 - P_{52}$.

Markov processes

Transition equations — Matrix-Vector Form

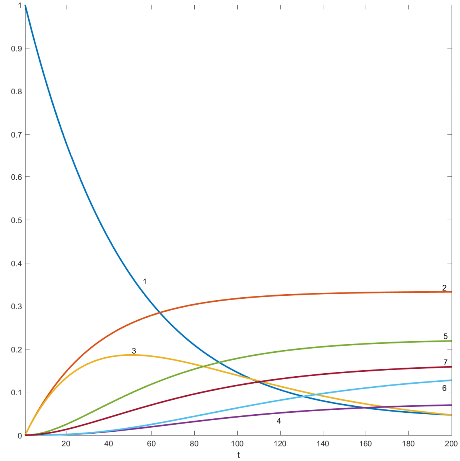
For an n-state system,

- Define

$$\pi(t) = \begin{bmatrix} \pi_1(t) \\ \pi_2(t) \\ \dots \\ \pi_n(t) \end{bmatrix}, \quad P = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ & & \dots & \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}, \quad \nu = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

- Transition equations: $\pi(t+1) = P\pi(t)$
- Normalization equation: $\nu^T \pi(t) = 1$
- Other facts:
  - $\nu^T P = \nu^T$ *(Each column of P sums to 1.)*
  - $\pi(t) = P^t \pi(0)$

Markov processes
Steady state



◊ * ⊛

State probabilities vs. t for system in Slide 7

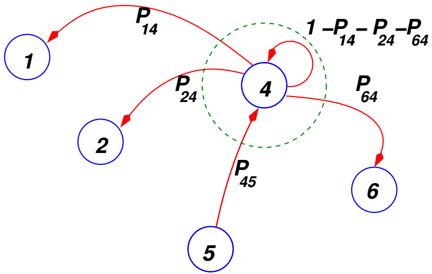
Markov processes
Steady state

* ⋆ ▲

- Steady state: $\pi_i = \lim_{t \to \infty} \pi_i(t) = \lim_{t \to \infty} P^t \pi(0)$, if it exists.
- Steady-state transition equations: $\pi_i = \sum_j P_{ij}\pi_j$.
- *Alternatively,* steady-state balance equations: $\pi_i \sum_{m, m \neq i} P_{mi} = \sum_{j, j \neq i} P_{ij}\pi_j$
- Normalization equation: $\sum_i \pi_i = 1$.

Markov processes

Balance equations



Balance equation:

$$(P_{14} + P_{24} + P_{64})\pi_4$$
$$= P_{45}\pi_5$$

in steady state *only*.

Intuitive meaning: The average number of transitions *into* the circle per unit time equals the average number of transitions *out of* the circle per unit time.

Markov processes

Steady state

How to calculate the steady-state probability distribution π

- Assume that the system has N states, where N is finite.
- Assume that there is a unique steady-state probability distribution.
- The transition equations form a set of N linear equations in N unknowns.
- The normalization equation is also a linear equation.
- *Problem?* We have $N+1$ equations in N unknowns.
- *No problem:* there is one redundant equation because each column sums to 1.
- Delete one transition equation and replace it with the normalization equation.
- Solve the system of N linear equations in N unknowns.

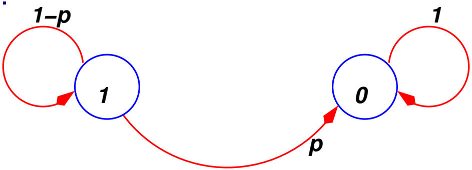
Markov processes

Steady state

- A system that has a unique steady-state solutions is called *ergodic* . The probability distribution approaches that limit no matter the initial probability distribution was.
- For systems that have more than one steady-state solution, the limiting distribution depends on the initial probability.
- The balance equations can be used to find the limiting distribution instead of the transition equations. As before, one equation has to be replaced by the normalization equation.
- If a system has an infinite number of states and it has a steady state probability distribution, there are two possibilties for finding it:
  - ⋆ It might be possible to solve the equations analytically. We will see an example of that.
  - ⋆ Truncate the system. That is, solve a system with a large but finite subset of the states. If you understand the system, you can guess which are the highest probability states. Keep those. This provides an approximate solution.

Markov processes
Geometric distribution

Consider a two-state system. The system can go from 1 to 0, but not from 0 to 1.



Let p be the conditional probability that the system is in state 0 at time $t+1$, given that it is in state 1 at time t. Then

$$p = P\left[\alpha(t+1) = 0 \middle| \alpha(t) = 1\right].$$

Markov processes

Geometric distribution — Transition equations

Let $\pi(\alpha, t)$ be the probability of being in state α at time t.

$$\begin{aligned}\pi(0, t+1) \quad &= P\left[\alpha(t+1) = 0 \middle| \alpha(t) = 1\right] P[\alpha(t) = 1] \\ &+ P\left[\alpha(t+1) = 0 \middle| \alpha(t) = 0\right] P[\alpha(t) = 0],\end{aligned}$$

we have

$$\begin{aligned}\pi(0, t+1) \quad &= p\pi(1, t) + \pi(0, t), \\ \pi(1, t+1) \quad &= (1-p)\pi(1, t),\end{aligned}$$

and the normalization equation

$$\pi(1, t) + \pi(0, t) = 1.$$

Markov processes

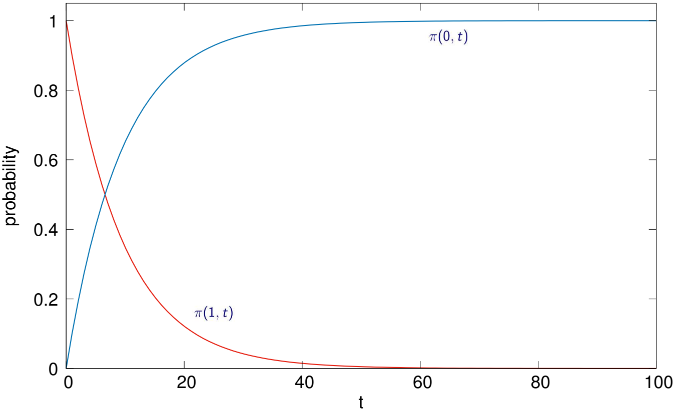
Geometric distribution — transient probability distribution

Assume that $\pi(1, 0) = 1$. Then the solution is

$$\begin{aligned} \pi(0, t) &= 1 - (1 - p)^t, \\ \pi(1, t) &= (1 - p)^t. \end{aligned}$$

Markov processes

Geometric distribution — transient probability distribution



Markov processes
Geometric distribution

We have shown that the probability that the state goes from 1 to 0 at time t is

$$P(t) = (1-p)^{t-1}p$$

The mean time for the state to go from 1 to 0 is then

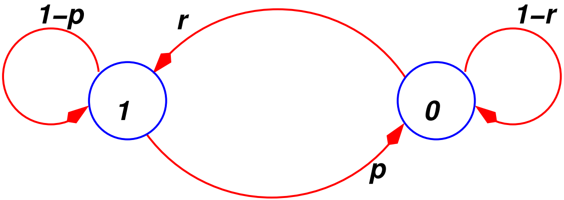
$$\bar{t} = \sum_{t=1}^{\infty} tP(t) = \sum_{t=1}^{\infty} t(1-p)^{t-1}p$$

It is not hard to show that

$$\bar{t} = \frac{1}{p}$$

Markov processes
Unreliable machine

1=up; 0=down.



Mean up time $=$ Mean time to fail $=$ MTTF$= 1/p$

Mean down time $=$ Mean time to repair $=$ MTTR$= 1/r$

Markov processes
Unreliable machine — transient probability distribution

The probability distribution satisfies

$$\pi(0, t+1) = \pi(0, t)(1-r) + \pi(1, t)p,$$

$$\pi(1, t+1) = \pi(0, t)r + \pi(1, t)(1-p).$$

Markov processes

Unreliable machine — transient probability distribution

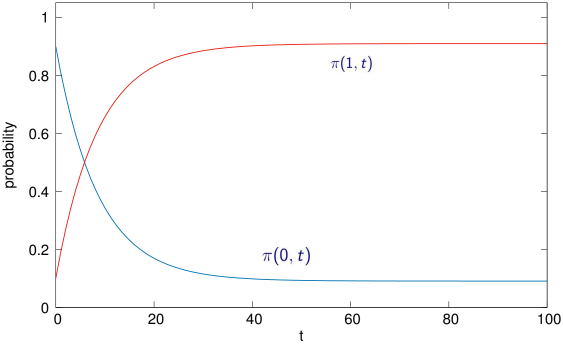
It is not hard to show that

$$\begin{aligned}\pi(0,t) &= \pi(0,0)(1-p-r)^t \\ &\quad + \frac{p}{r+p}\left[1-(1-p-r)^t\right],\end{aligned}$$

$$\begin{aligned}\pi(1,t) &= \pi(1,0)(1-p-r)^t \\ &\quad + \frac{r}{r+p}\left[1-(1-p-r)^t\right].\end{aligned}$$

Markov processes

Unreliable machine — transient probability distribution



Markov processes

Unreliable machine — steady-state probability distribution

As $t \to \infty$,

$$\begin{aligned}\pi(0,t) &\to \frac{p}{r+p},\\ \pi(1,t) &\to \frac{r}{r+p}\end{aligned}$$

which is the solution of

$$\begin{aligned}\pi(0) &= \pi(0)(1-r) + \pi(1)p,\\ \pi(1) &= \pi(0)r + \pi(1)(1-p).\end{aligned}$$

Markov processes
Unreliable machine — efficiency

If a machine makes one part per time unit when it is operational, its average production rate is

$$\pi(1) = \frac{r}{r+p}$$

This quantity is the *efficiency* (e) of the machine. If the machine makes one part per τ time units when it is operational, its average production rate is

$$P = \frac{1}{\tau}\left(\frac{r}{r+p}\right)$$

Note that we can also write

$$e = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}$$

Discrete state, continuous time

States and transitions

- States can be numbered 0, 1, 2, 3, ... *(or with multiple indices if that is more convenient)*.
- Time is a real number, defined on $(-\infty, \infty)$ or a smaller interval.
- The probability of a transition from j to i during $[t, t + \delta t]$ is approximately $\lambda_{ij}\delta t$, where δt is small, and

  $$\lambda_{ij}\delta t \approx P\{X(t + \delta t) = i | X(t) = j\} \text{ for } i \neq j$$

Discrete state, continuous time

States and transitions

More precisely,

$$\lambda_{ij}\delta t = P\{X(t+\delta t) = i | X(t) = j\} + o(\delta t)$$
$$\text{for } i \neq j$$

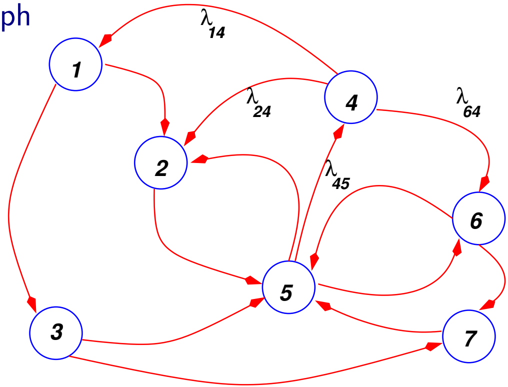
where $o(\delta t)$ is a function that satisfies $\lim_{\delta t \to 0} \frac{o(\delta t)}{\delta t} = 0$

This implies that for small δt, $o(\delta t) \ll \delta t$.

Discrete state, continuous time

States and transitions

Transition graph



λ_{ij} is a probability *rate.* $\lambda_{ij}\delta t$ is a probability.

Compare with the discrete-time graph.

* □

Discrete state, continuous time

States and transitions

One of the transition equations:

Define $\pi_i(t) = P\{X(t) = i\}$.

$$\pi_5(t + \delta t) \approx$$

$$(1 - \lambda_{25}\delta t - \lambda_{45}\delta t - \lambda_{65}\delta t)\pi_5(t) +$$

$$\lambda_{52}\delta t \pi_2(t) + \lambda_{53}\delta t \pi_3(t) + \lambda_{56}\delta t \pi_6(t) + \lambda_{57}\delta t \pi_7(t) +$$

Discrete state, continuous time

States and transitions

Or,

$$\pi_5(t + \delta t) \approx$$

$$\pi_5(t) - (\lambda_{25} + \lambda_{45} + \lambda_{65})\pi_5(t)\delta t$$

$$+(\lambda_{52}\pi_2(t) + \lambda_{53}\pi_3(t) + \lambda_{56}\pi_6(t) + \lambda_{57}\pi_7(t))\delta t$$

Discrete state, continuous time

States and transitions

Or,

$$\lim_{\delta t \to 0} \frac{\pi_5(t+\delta t) - \pi_5(t)}{\delta t} =$$

$$\frac{d\pi_5}{dt}(t) = -(\lambda_{25} + \lambda_{45} + \lambda_{65})\pi_5(t)$$

$$+\lambda_{52}\pi_2(t) + \lambda_{53}\pi_3(t) + \lambda_{56}\pi_6(t) + \lambda_{57}\pi_7(t)$$

Discrete state, continuous time

States and transitions

Define $\boxed{\textit{for convenience}}$

$$\lambda_{55} = -(\lambda_{25} + \lambda_{45} + \lambda_{65})$$

Then

$$\frac{d\pi_5}{dt}(t) = \lambda_{55}\pi_5(t) +$$

$$\lambda_{52}\pi_2(t) + \lambda_{53}\pi_3(t) + \lambda_{56}\pi_6(t) + \lambda_{57}\pi_7(t)$$

Discrete state, continuous time

States and transitions

- Define $\pi_i(t) = P\{X(t) = i\}$
- It is $\boxed{\textit{convenient}}$ to define $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$ $***$
- Transition equations: $\dfrac{d\pi_i(t)}{dt} = \sum_j \lambda_{ij}\pi_j(t).$
- Normalization equation: $\sum_i \pi_i(t) = 1.$

$***$ *Often confusing!!!*

Discrete state, continuous time

Transition equations — Matrix-Vector Form

- Define $\pi(t), \nu$ as before *.

  Define $\Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2n} \\ & & \dots & \\ \lambda_{n1} & \lambda_{n2} & \dots & \lambda_{nn} \end{bmatrix}$

- Transition equations: $\dfrac{d\pi(t)}{dt} = \Lambda \pi(t)$.

- Normalization equation: $\nu^T \pi = 1$.

- Other facts:

  - ⋆ $\nu^T P = 0$ *(Each column of P sums to 0.)*
  - ⋆ $\pi(t) = e^{\Lambda t} \pi(0)$

Discrete state, continuous time

Steady State

- *Steady state:* $\pi_i = \lim_{t\to\infty} \pi_i(t)$, if it exists.
- Steady-state transition equations: $0 = \sum_j \lambda_{ij}\pi_j$.
- *Alternatively,* steady-state balance equations:
  $\pi_i \sum_{m,m\neq i} \lambda_{mi} = \sum_{j,j\neq i} \lambda_{ij}\pi_j$
- Normalization equation: $\sum_i \pi_i = 1$.

Discrete state, continuous time

Steady State — Matrix-Vector Form

- *Steady state:* $\pi = \lim_{t \to \infty} \pi(t)$, if it exists.

- Steady-state transition equations: $0 = \Lambda \pi$.

- Normalization equation: $\nu^T \pi = 1$.

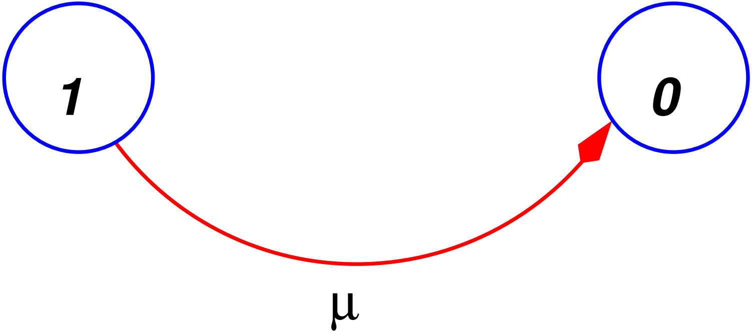
Discrete state, continuous time

Sources of confusion in continuous time models

- $\boxed{\textit{Never}}$ Draw self-loops in continuous time Markov process graphs.
- $\boxed{\textit{Never}}$ write $1 - \lambda_{14} - \lambda_{24} - \lambda_{64}$. Write
  - ⋆ $1 - (\lambda_{14} + \lambda_{24} + \lambda_{64})\delta t$, or
  - ⋆ $-(\lambda_{14} + \lambda_{24} + \lambda_{64})$
- $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ji}$ is $\boxed{\textit{NOT}}$ a rate and $\boxed{\textit{NOT}}$ a probability. It is $\boxed{\textit{ONLY}}$ a convenient notation.

Discrete state, continuous time
Exponential distribution

Exponential random variable T: the time to move from state 1 to state 0.



Discrete state, continuous time

Exponential distribution

$\pi(0, t + \delta t) =$

$$P\left[\alpha(t+\delta t)=0 \middle| \alpha(t)=1\right] P[\alpha(t)=1]+$$

$$P\left[\alpha(t+\delta t)=0 \middle| \alpha(t)=0\right] P[\alpha(t)=0].$$

or

$$\pi(0, t+\delta t) = \mu \delta t \pi(1, t) + \pi(0, t) + o(\delta t)$$

or

$$\frac{d\pi(0, t)}{dt} = \mu \pi(1, t).$$

Discrete state, continuous time

Exponential distribution

$\pi(1, t + \delta t) =$

$$P\left[\alpha(t + \delta t) = 1 \middle| \alpha(t) = 1\right] P[\alpha(t) = 1] +$$

$$P\left[\alpha(t + \delta t) = 1 \middle| \alpha(t) = 0\right] P[\alpha(t) = 0].$$

or

$$\pi(1, t + \delta t) = (1 - \mu \delta t)\pi(1, t) + (0)\pi(0, t) + o(\delta t)$$

or

$$\frac{d\pi(1, t)}{dt} = -\mu \pi(1, t).$$

Discrete state, continuous time

Exponential distribution

Transition equations
$$\begin{cases} \dfrac{d\pi(0,t)}{dt} &= \quad \mu\pi(1,t) \\ \dfrac{d\pi(1,t)}{dt} &= -\mu\pi(1,t) \end{cases}$$

If $\pi(0,0) = 0,\ \pi(1,0) = 1$, then

$$\pi(1,t) = e^{-\mu t}$$

and

$$\pi(0,t) = 1 - e^{-\mu t}$$

Discrete state, continuous time

Exponential distribution

The probability that the transition takes place at some $T \in [t, t + \delta t]$ is

$$f(t)\delta t = P[\alpha(t + \delta t) = 0 \text{ and } \alpha(t) = 1]$$

$$\approx P[\alpha(t + \delta t) = 0 | \alpha(t) = 1] P[\alpha(t) = 1]$$

$$= (\mu \delta t)(e^{-\mu t})$$

The exponential density function is therefore $f(t) = \mu e^{-\mu t}$ for $t \geq 0$ and 0 for $t < 0$.

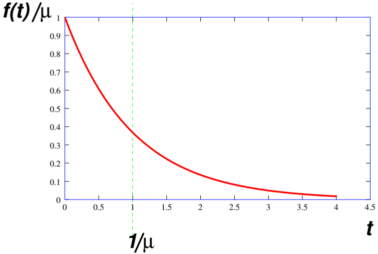
The time of the transition from 1 to 0 is said to be *exponentially distributed* with rate μ.

The expected transition time is $\dfrac{1}{\mu} = \displaystyle\int_0^\infty t e^{-\mu t}$.

Discrete state, continuous time

Exponential distribution

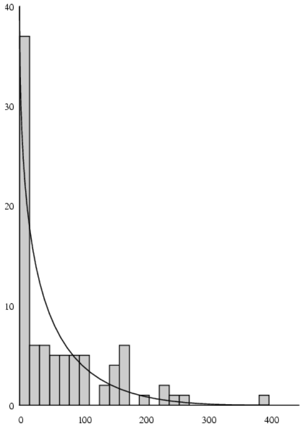
- $f(t) = \mu e^{-\mu t}$ for $t \geq 0$; $f(t) = 0$ otherwise;
  $F(t) = 1 - e^{-\mu t}$ for $t \geq 0$; $F(t) = 0$ otherwise.
- $ET = 1/\mu, V_T = 1/\mu^2$. Therefore, $\sigma = ET$ so cv=1.



Markov processes
Exponential

Density function

Exponential density function and a small number of samples.



Discrete state, continuous time

Exponential distribution: some properties

- Memorylessness:

  $P(T > t + x | T > x) = P(T > t)$

- $P(t \leq T \leq t + \delta t | T \geq t) \approx \mu \delta t$ for small δt.

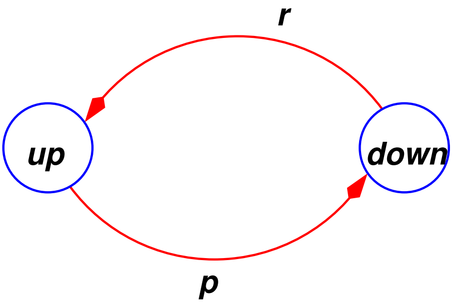
Discrete state, continuous time

Exponential distribution: some properties

- If $T_1, ..., T_n$ are independent exponentially distributed random variables with parameters $\mu_1 ..., \mu_n$, and
- $T = \min(T_1, ..., T_n)$, then
- T is an exponentially distributed random variable with parameter $\mu = \mu_1 + ... + \mu_n$.
- Consequently, the time that the system stays in any state is exponentially distributed.

Discrete state, continuous time
Unreliable machine

Continuous time unreliable machine.



Discrete state, continuous time

Unreliable machine

From the *Law of Total Probability:*

$P(\{\text{the machine is up at time } t+\delta t\})=$

$P(\{\text{the machine is up at time } t+\delta t \mid \text{the machine was up at time } t\}) \times$
$P(\{\text{the machine was up at time } t\}) +$

$P(\{\text{the machine is up at time } t+\delta t \mid \text{the machine was down at time } t\}) \times$
$P(\{\text{the machine was down at time } t\})$

$+o(\delta t)$

and similarly for $P(\{\text{the machine is down at time } t+\delta t\})$.

Discrete state, continuous time

Unreliable machine

Probability distribution notation and dynamics:

$\pi(1, t) =$ the probability that the machine is up at time t.

$\pi(0, t) =$ the probability that the machine is down at time t.

$$P(\text{the machine is up at time } t + \delta t | \text{ the machine was up at time } t) \\ = 1 - p\delta t$$

$$P(\text{the machine is up at time } t + \delta t | \text{ the machine was down at time } t) \\ = r\delta t$$

Discrete state, continuous time

Unreliable machine

Therefore

$$\pi(1, t + \delta t) = (1 - p\delta t)\pi(1, t) + r\delta t\pi(0, t) + o(\delta t)$$

Similarly,

$$\pi(0, t + \delta t) = p\delta t\pi(1, t) + (1 - r\delta t)\pi(0, t) + o(\delta t)$$

Discrete state, continuous time

Unreliable machine

or,

$$\pi(1, t + \delta t) - \pi(1, t) = -p\delta t \pi(1, t) + r\delta t \pi(0, t) + o(\delta t)$$

or,

$$\frac{\pi(1, t + \delta t) - \pi(1, t)}{\delta t} = -p\pi(1, t) + r\pi(0, t) + \frac{o(\delta t)}{\delta t}$$

Discrete state, continuous time

or,

$$\frac{d\pi(1,t)}{dt} = \pi(0,t)r - \pi(1,t)p$$

$$\frac{d\pi(0,t)}{dt} = -\pi(0,t)r + \pi(1,t)p$$

Markov processes
Unreliable machine

Solution

$$\pi(0,t) = \frac{p}{r+p} + \left[\pi(0,0) - \frac{p}{r+p}\right] e^{-(r+p)t}$$

$$\pi(1,t) = 1 - \pi(0,t).$$

As $t \to \infty$,

$$\pi(0) \to \frac{p}{r+p},$$
$$\pi(1) \to \frac{r}{r+p}$$

Markov processes
Unreliable machine

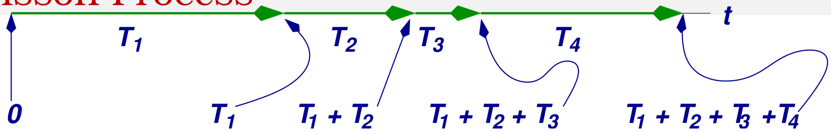
Note that MTTF$=1/p$; MTTR$=1/r$. Units are natural time units, not operation times.

If the machine makes μ parts per time unit on the average when it is operational, the steady-state average production rate is

$$\mu\pi(1) = \mu\frac{r}{r+p} = \mu\frac{\text{MTTF}}{\text{MTTF}+\text{MTTR}} = \mu e$$

Discrete state, continuous time

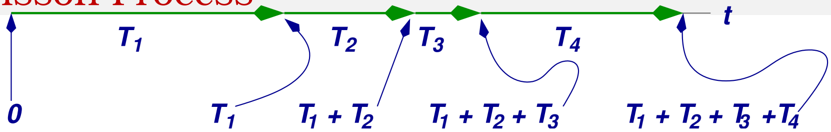
Poisson Process



- Let $T_i, i = 1, \ldots$ be a set of independent exponentially distributed random variables with parameter λ. Each random variable may represent the time between occurrences of a repeating event.
  - ⋆ Examples: customer arrivals, clicks of a Geiger counter
- Then $\sum_{i=1}^{n} T_i$ is the time required for n such events.

Discrete state, continuous time

Poisson Process

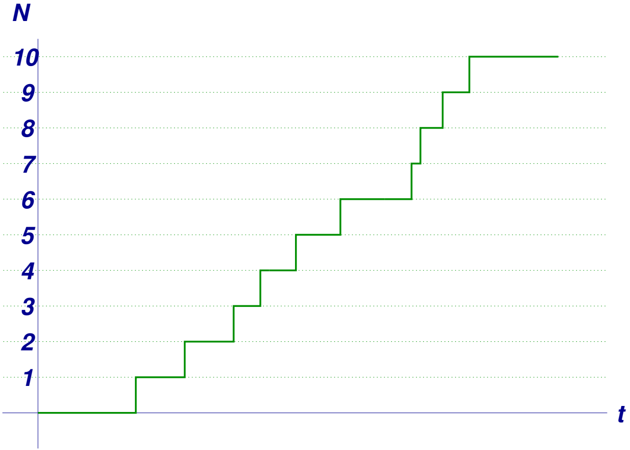


- *Informally:* $N(t)$ is the number of events that occur between 0 and t.
- *Formally:* Define
$$N(t) = \begin{cases} 0 \text{ if } T_1 > t \\ n \text{ such that } \sum_{i=1}^{n} T_i \leq t, \quad \sum_{i=1}^{n+1} T_i > t \end{cases}$$
- Then $N(t)$ is a *Poisson process* with parameter λ.

Queueing theory

$M/M/1$ Queue

Number of events $N(t)$ during $[0, t]$

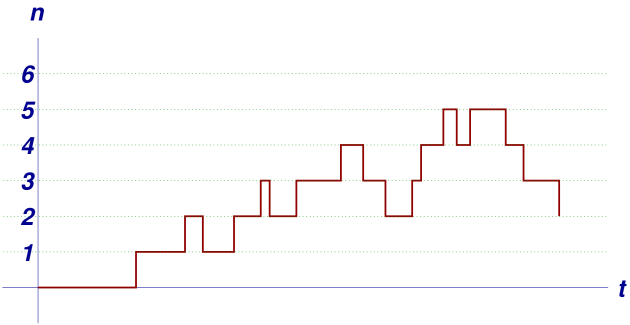


Queueing theory
$M/M/1$ Queue

- Simplest model is the $M/M/1$ queue:
  - ⋆ Exponentially distributed inter-arrival times — mean is $1/\lambda$; λ is *arrival rate* (customers/time). *(Poisson arrival process.)*
  - ⋆ Exponentially distributed service times — mean is $1/\mu$; μ is *service rate* (customers/time).
  - ⋆ The arrival and service processes are independent.
  - ⋆ 1 server.
  - ⋆ Infinite waiting area.
- Define the *utilization* $\rho = \lambda/\mu$.

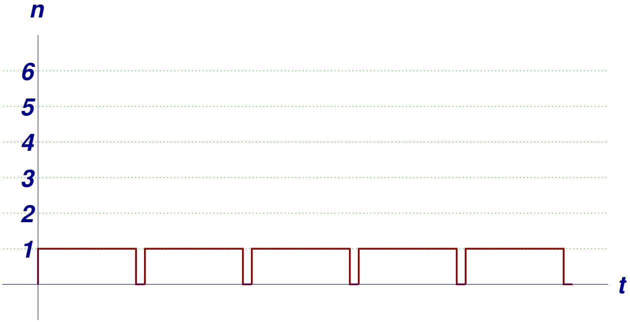
Queueing theory
$M/M/1$ Queue

Number of customers in the system as a function of time for a $M/M/1$ queue.



Queueing theory
$D/D/1$ Queue

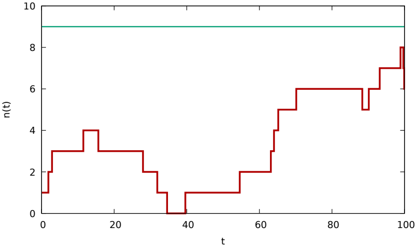
Number of customers in the system as a function of time for a $D/D/1$ queue.



Queueing theory
Sample path

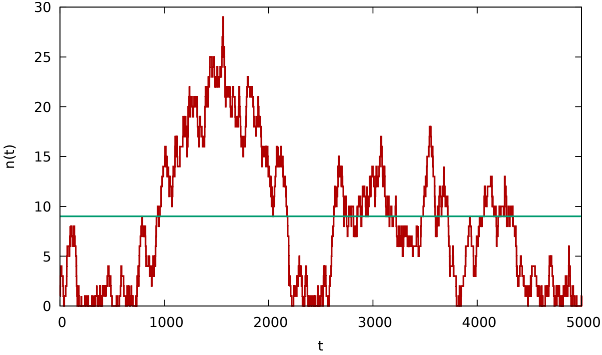
- Suppose customers arrive in a Poisson process with *average* inter-arrival time $1/\lambda = 1$ minute; and that service time is exponentially distributed with *average* service time $1/\mu = 54$ seconds.

    ⋆ The average number of customers in the system is *9.*



Queue behavior over a short time interval — initial transient

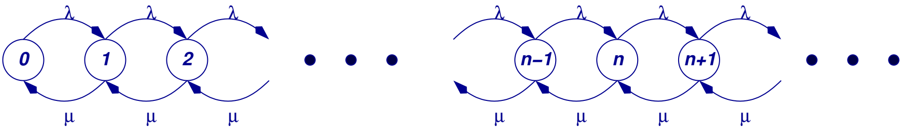
Queueing theory
Sample path



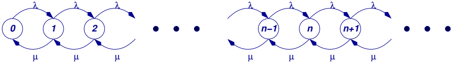
Queue behavior over a long time interval

Queueing theory
$M/M/1$ Queue

State space



Queueing theory
$M/M/1$ Queue



Let $\pi(n, t)$ be the probability that there are n parts in the system at time t. Then,

For $n > 0$,

$$\begin{aligned}\pi(n, t + \delta t) &= \pi(n-1, t)\lambda\delta t + \pi(n+1, t)\mu\delta t + \\ &\quad \pi(n, t)(1 - (\lambda\delta t + \mu\delta t)) + o(\delta t)\end{aligned}$$

and

$$\pi(0, t + \delta t) = \pi(1, t)\mu\delta t + \pi(0, t)(1 - \lambda\delta t) + o(\delta t).$$

Queueing theory
$M/M/1$ Queue

Or,

$$\begin{aligned}\frac{d\pi(n,t)}{dt} &= \pi(n-1,t)\lambda + \pi(n+1,t)\mu - \pi(n,t)(\lambda+\mu), \quad n>0 \\ \frac{d\pi(0,t)}{dt} &= \pi(1,t)\mu - \pi(0,t)\lambda.\end{aligned}$$

If a steady state distribution exists, it satisfies

$$\begin{aligned}0 &= \pi(n-1)\lambda + \pi(n+1)\mu - \pi(n)(\lambda+\mu), n>0 \\ 0 &= \pi(1)\mu - \pi(0)\lambda.\end{aligned}$$

Why "if"?

Queueing theory
$M/M/1$ Queue – Steady State

Let $\rho = \lambda/\mu$. These equations are satisfied by

$$\pi(n) = (1 - \rho)\rho^n, n \geq 0$$

if $\rho < 1$.

The average number of parts in the system is

$$\bar{n} = \sum_{n=0}^{\infty} n\pi(n) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}.$$

Queueing theory
Little's Law

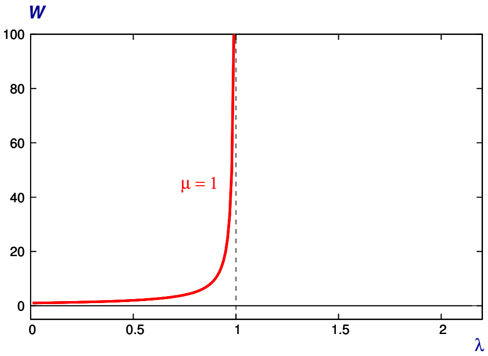
- True for most systems of practical interest *(not just $M/M/1$)*.
- Steady state only.
- $L =$ the average number of customers in a system.
- $W =$ the average delay experienced by a customer in the system.

$$L = \lambda W$$

In the $M/M/1$ queue, $L = \bar{n}$ and

$$W = \frac{1}{\mu - \lambda}.$$

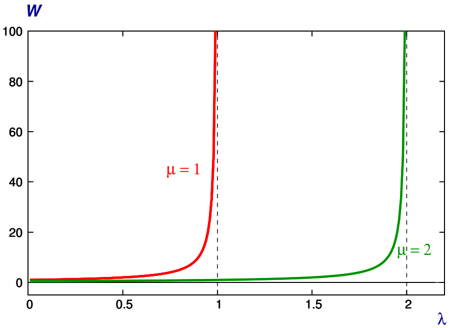
Queueing theory
$M/M/1$ Queue capacity



- μ is the *capacity* of the system.
- If $\lambda < \mu$, system is stable and waiting time remains bounded.
- If $\lambda > \mu$, waiting time grows over time.

Queueing theory

$M/M/1$ Queue capacity



- To increase capacity, increase μ.
- To decrease delay for a given λ, increase μ.

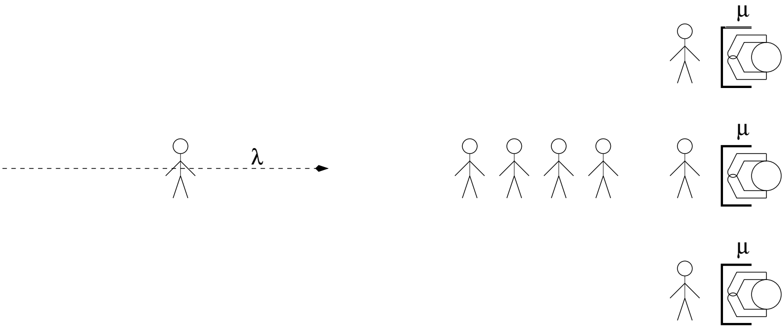
Queueing theory
Other Single-Stage Models

Things get more complicated when:

- There are multiple servers.
- There is finite space for queueing.
- The arrival process is not Poisson.
- The service process is not exponential.

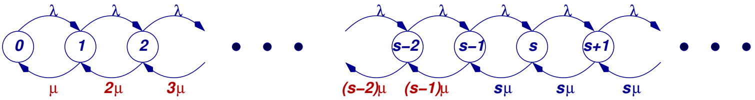
Closed formulas and approximations exist for some, but not all, cases.

Queueing theory
$M/M/s$ Queue



s-Server Queue, $s = 3$

Queueing theory
$M/M/s$ Queue



- The departure rate when there are $k > s$ customers in the system is $s\mu$ since all s servers are always busy.

- The departure rate when there are $k \leq s$ customers in the system is $k\mu$ since only k of the servers are busy.

Queueing theory
$M/M/s$ Queue

$$P(k) = \begin{cases} \pi(0)\dfrac{s^k \rho^k}{k!}, & k \leq s \\ \\ \pi(0)\dfrac{s^s \rho^k}{s!}, & k > s \end{cases}$$

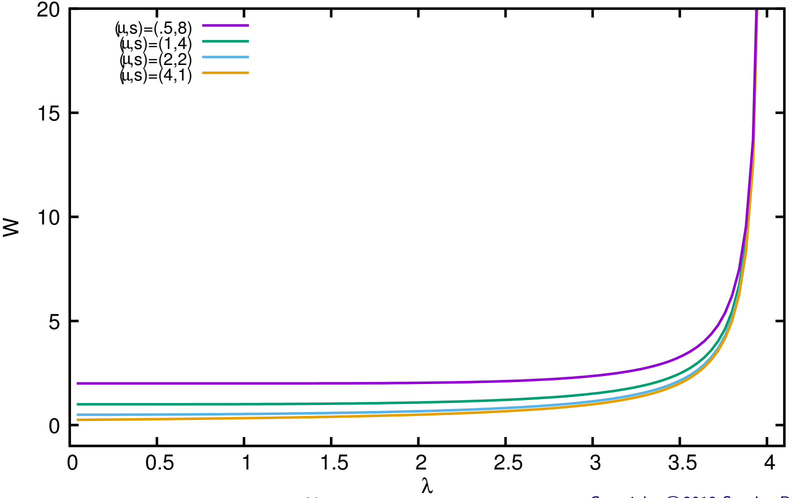
where

$$\rho = \frac{\lambda}{s\mu} < 1; \qquad \pi(0) \text{ chosen so that} \sum_k P(k) = 1$$

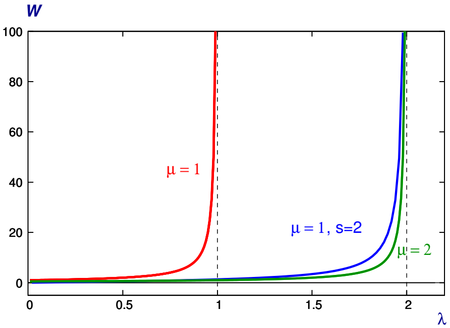
Queueing theory

$M/M/s$ Queue

W vs. λ; $s\mu = 4$



Queueing theory
$M/M/1$ Queue capacity



To increase capacity or reduce delay,

- increase μ, or
- add servers in parallel
  ... but that will not reduce delay as much.

Queueing theory

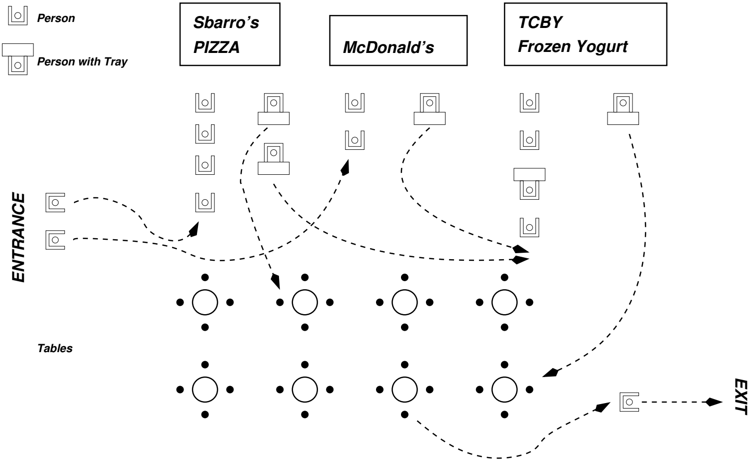
Networks of Queues

- Set of queues where customers can go to another queue after completing service at a queue.
- *Open network:* where customers enter and leave the system. λ is known and we must find L and W.
- *Closed network:* where the population of the system is constant. L is known and we must find λ and W.

Queueing theory
Networks of Queues

Examples of open networks

- internet traffic
- emergency room (*arrive*, triage, waiting room, treatment, tests, *exit* or *hospital admission*)
- food court
- airport (*arrive*, ticket counter, security, passport control, gate, *board plane*)
- factory with no *centralized* material flow control after material enters



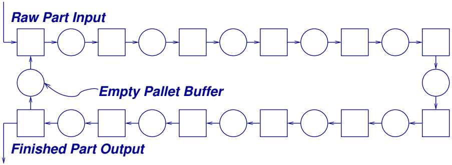
Person
Person with Tray
Sbarro's PIZZA
McDonald's
TCBY Frozen Yogurt
ENTRANCE
Tables
EXIT

Queueing theory

Networks of Queues

Examples of closed networks

- factory with limited fixtures or pallets



- factory with material controlled by keeping the number of items constant (CONWIP)

Queueing theory
Jackson Networks

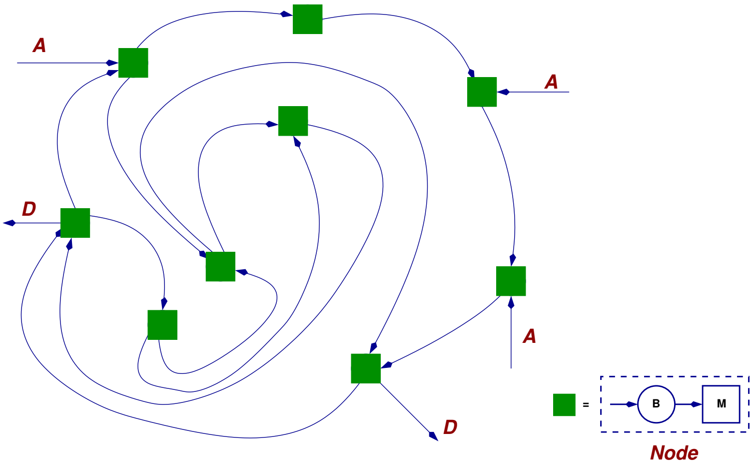
Queueing networks are often modeled as *Jackson networks.*

- Relatively easy to compute performance measures (capacity, average time in system, average queue lengths).
- Easily provides intuition.
- Easy to optimize and to use for design.
- Valid (or good approximation) for a large class of systems ...

Queueing theory
Jackson Networks

- ... but not all. Storage areas must be assumed to be infinite (which means blocking is assumed not to happen).
    - ⋆ This assumption leads to bad results for systems with bottlenecks at locations other than the first station.

Queueing theory
Open Jackson Networks



Queueing theory
Open Jackson Networks

- Items *arrive* from outside the system to node i according to a Poisson process with rate α_i.
- $\alpha_i > 0$ for at least one i.
- When an item's service at node i is finished, it goes to node j next with probability p_{ij}.
- If $p_{i0} = 1 - \sum_j p_{ij} > 0$, items *depart* from the network from node i.
- $p_{i0} > 0$ for at least one i.
- We will focus on the special case in which each node has a single server with exponential processing time. The service rate of node i is μ_i.

Queueing theory

Open Jackson Networks

Goals of analysis:

- to determine if the system is feasible
- to determine how much inventory is in this system (on the average) and how it is distributed
- to determine the average waiting time at each node and the average time a part spends in the system.

Queueing theory
Open Jackson Networks

- Define λ_i as the total arrival rate of items to node i. This includes items entering the network at i and items coming from all other nodes.
- $p_{ji}\lambda_j$ is the portion of the flow arriving at node j that goes to node i.
- Then $\lambda_i = \alpha_i + \sum_j p_{ji}\lambda_j$
- In matrix form, let λ be the vector of λ_i, α be the vector of α_i, and P be the matrix of p_{ij}. Then
$$\lambda = \alpha + \mathsf{P}^T\lambda$$
- Solving for λ,
$$\lambda = (I - \mathsf{P}^T)^{-1}\alpha$$

Queueing theory
Open Jackson Networks

Probability distribution:

- If $\lambda_i < \mu_i$ for each i, define $\rho_i = \lambda_i / \mu_i$ and
$$\pi_i(n_i) = (1 - \rho_i)\rho_i^{n_i}$$
- This is the solution of an M/M/1 queue with arrival rate λ_i calculated on the previous slide and service rate μ_i specified by the given problem data.
- If $\lambda_i \geq \mu_i$ for some i, *the demand is not feasible.* The system cannot handle the demand placed on it.

Queueing theory
Open Jackson Networks

Solution:

- Define $\pi(n_1, n_2, ..., n_k)$ to be the steady-state probability that there are n_i items at node i, $i = 1, ..., k$.
- Then the probability distribution for the entire system is
$$\pi(n_1, n_2, ..., n_k) = \prod_i \pi_i(n_i)$$
- At each node i

$$\bar{n}_i = En_i = \frac{\rho_i}{1 - \rho_i}$$

Queueing theory
Open Jackson Networks

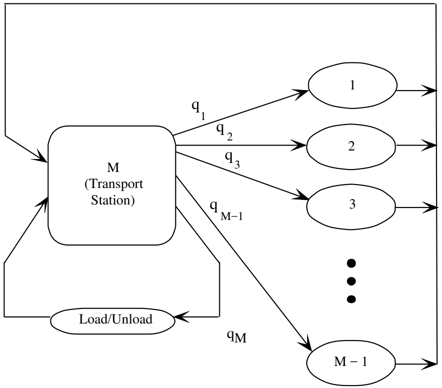
- The solution is product form. It says that the probability of the system being in a given state is the product of the probabilities of the queue at each node being in the corresponding state.
- This exact analytic formula is the reason that the Jackson network model is of interest. It is relatively easy to use to calculate the performance of a complex system.
- The product form solution holds for some more general cases.
- However, it is restricted to models of systems with unlimited storage space. *Consequently, it cannot model blocking.*
  - ⋆ It is a good approximation for systems where blocking is rare, for example when the arrival rate of material is much less than the capacity of the system.
  - ⋆ It will not work so well where blocking occurs often.

Queueing theory
Closed Jackson Networks

- Consider an extension in which
  - ⋆ $\alpha_i = 0$ for all nodes i.
  - ⋆ $p_{i0} = 1 - \sum_j p_{ij} = 0$ for all nodes i.
- Then
  - ⋆ Since nothing is entering and nothing is departing from the network, the number of items in the network is *constant* . That is, $\sum_i n_i(t) = N$ for all t.
  - ⋆ $\lambda_i = \sum_j p_{ji}\lambda_j$ does not have a unique solution.
  - ⋆ This means that a different solution approach is needed to analyze the system. It is used in the example that follows.

Queueing theory
Closed Jackson Network model of an FMS

Solberg's "CANQ" model.



Let $\{p_{ij}\}$ be the set of routing probabilities, as defined on Slide 89.

$p_{iM} = 1$ if $i \neq M$

$p_{Mj} = q_j$ if $j \neq M$

$p_{ij} = 0$ otherwise

Service rate at Station i is μ_i.

Queueing theory
Closed Jackson Network model of an FMS

- Input data: $M, N, q_j, \mu_j, s_j \quad (j = 1, ..., M)$
  - ⋆ M = number of stations, including transportation system
  - ⋆ N = number of pallets
  - ⋆ q_j = fraction of parts going from the transportation system to Station j
  - ⋆ μ_j = processing rate of machines at Station j
  - ⋆ s_j = number of machines at Station j

- Output data: $P, W, \rho_j \quad (j = 1, ..., M)$
  - ⋆ P = production rate
  - ⋆ W = average time a part spends in the system
  - ⋆ ρ_j = utilization per machine of Station j

Queueing theory
Closed Jackson Network model of an FMS

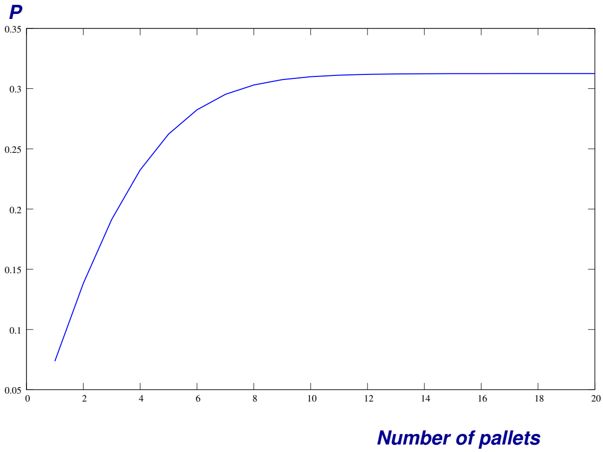
For the following graphs,

- Base input data: $M, N, q_j, \mu_j, s_j \quad (j = 1, ..., M)$
  - $\star$ $M = 5$
  - $\star$ $N = 10$
  - $\star$ $q_j = .1,\ .2,\ .2,\ .25,\ .25$
  - $\star$ $1/\mu_j = 3.,\ 4.,\ 3.44,\ 1.41,\ 5.$
  - $\star$ $s_j = 2,\ 1,\ 2,\ 1,\ 15$

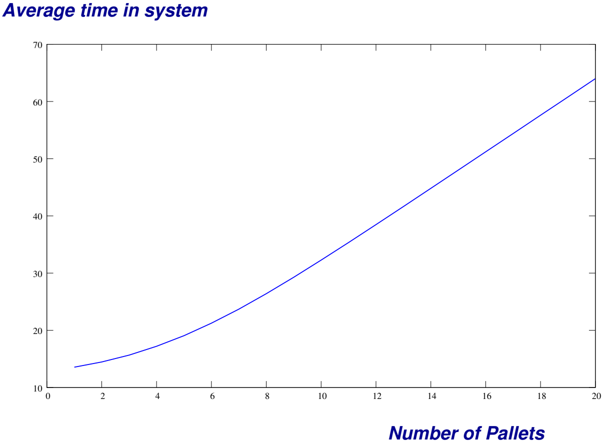
We see the effect of one of the variables on the performance measures in the following graphs.

Queueing theory

Closed Jackson Network model of an FMS

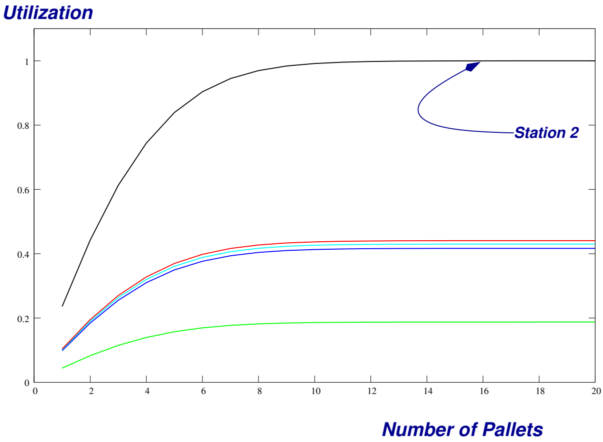


Queueing theory
Closed Jackson Network model of an FMS

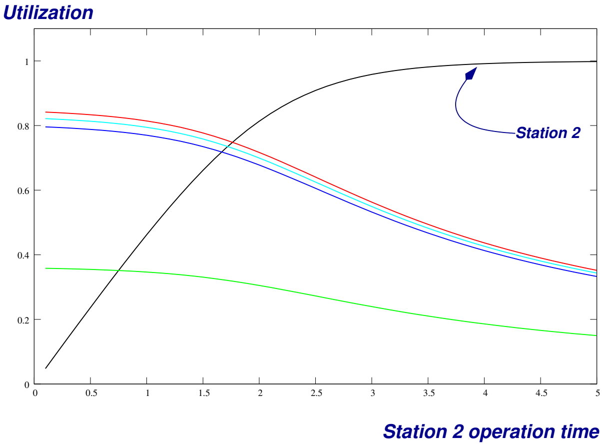


Queueing theory

Closed Jackson Network model of an FMS

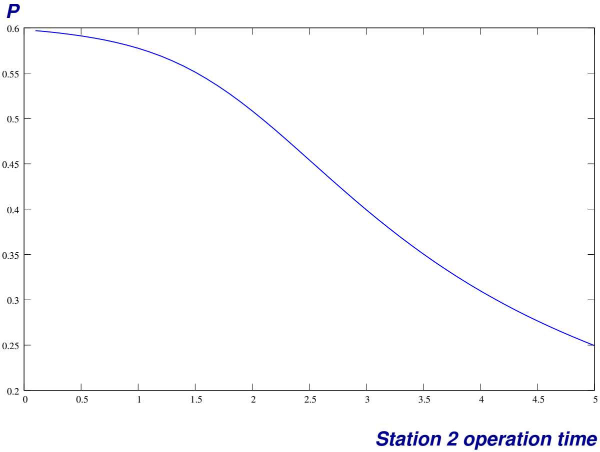


Queueing theory
Closed Jackson Network model of an FMS



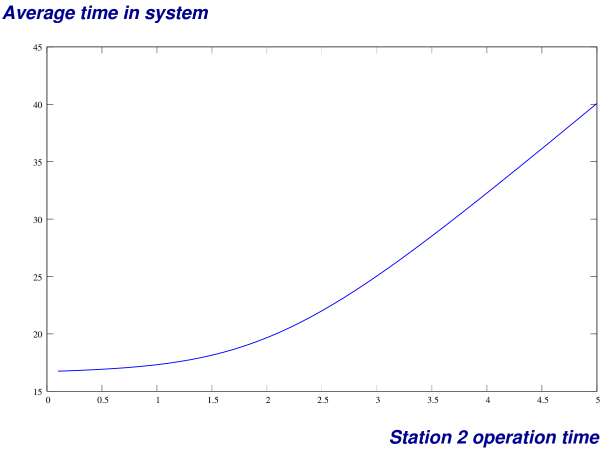
Queueing theory

Closed Jackson Network model of an FMS



Queueing theory

Closed Jackson Network model of an FMS



MIT 2.853/2.854

Introduction to Manufacturing Systems

Inventory

Stanley B. Gershwin
Laboratory for Manufacturing and Productivity
Massachusetts Institute of Technology
gershwin@mit.edu

Storage

Storage is fundamental!

- Storage is *fundamental* in nature, management, and engineering.
  - ⋆ In nature, *energy* is stored. Life can only exist if the acquisition of energy can occur at a different time from the the expenditure of energy.
  - ⋆ In management, *products* are stored.
  - ⋆ In engineering, *energy* is stored (springs, batteries, reservoirs); etc.

Storage

Purpose of storage

- *The purpose of storage is to allow systems to survive even when important events are unsynchronized.* For example,
  - ⋆ the separation in time from the acquisition or production of something and its consumption; and
  - ⋆ the occurrence of an event at one location (such as a machine failure or a power surge) which can prevent desired performance or do damage at another.
- Storage improves system performance by *decoupling* parts of the system from one another.

Storage
Purpose of storage

- It allows production systems (of energy or goods) to be built with capacity less than the peak demand.
  - ⋆ Helps manage seasonality.
- It reduces the propagation of disturbances, and thus reduces instability and the fragility of complex, expensive systems.

Manufacturing Inventory

Motives/benefits

Storage

- Allows economies of scale:
  - ⋆ volume purchasing
  - ⋆ reduced set-ups
- Helps manage uncertainty:
  - ⋆ Random arrivals of customers or orders.
  - ⋆ Unreliable deliveries of raw materials.

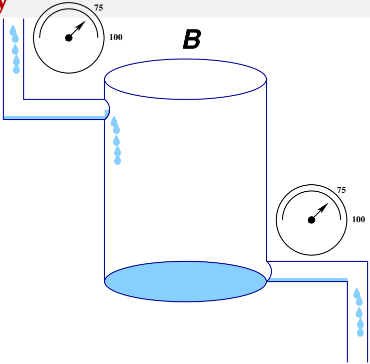
Manufacturing Inventory

Costs

- Financial: raw materials are paid for, but no revenue comes in until the item is sold.
- Demand risk: item loses value or is unsold due to obsolescence, for example
  - ⋆ newspapers
  - ⋆ technology
  - ⋆ fashion
- Holding cost (warehouse space)
- "Shrinkage" = damage/theft/spoilage/loss

Variability and Storage

No variability

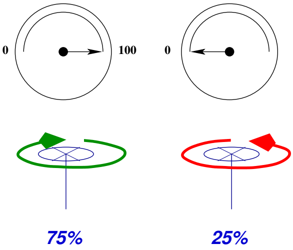


75 gal/sec in, 75 gal/sec out constantly

If we start with an empty tank, the tank is always empty (except for splashing at the bottom).

Variability and Storage
Variability from random valves

Consider a random valve:



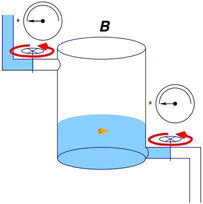
- The average period when the valve is open is 15 minutes.
- The average period when the valve is closed is 5 minutes.
- Consequently, the *average* flow rate through it is 75 gal/sec.

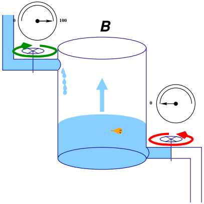
Variability and Storage

Variability from random valves

Four possibilities for two *unsynchronized* valves:

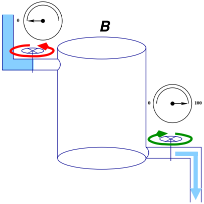
0 in, 0 out

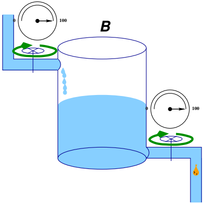




100 gal/sec in, 0 out

0 in, 100 gal/sec out





100 gal/sec in, 100 gal/sec out

Variability and Inventory

Observation:

- There is *never* any water in the tank when the flow is constant.
- There is *sometimes* water in the tank when the flow is variable.

Conclusions:

1. You can't always replace random variables with their averages.
2. *Variability causes inventory!!*

Variability and Inventory

- To be more precise, *non-synchronization causes inventory.*
  - ⋆ Living things do not acquire energy at the same time they expend it. Therefore, they must store energy in the form of fat or sugar.
  - ⋆ Rivers are dammed and reservoirs are created to control the flow of water — to reduce the variability of the water supply.
  - ⋆ For solar and wind power to be successful, energy storage is required for when the sun doesn't shine and the wind doesn't blow.

Inventory Topics

- Newsvendor Problem — pure demand risk (no inventory dynamics)
- Economic Order Quantity (EOQ) — pure economies of scale (deterministic demand)
- Q, R policy — supplier lead time and random demand
- Base Stock Policy — manage a simple factory to avoid stockouts
- Other topics

Newsvendor Problem
Demand risk

- Newsvendor buys x newspapers at c dollars per paper (*cost*).

- Newspapers are sold at *customer price* $r > c$.
- Unsold newspapers are redeemed at *salvage price* $s < c$.

$$r > c > s$$

Newsvendor Problem
Demand risk

Assume demand W is a continuous random variable with distribution function

$$F(w) = \text{prob}(W \leq w);$$

Assume $f(w) = dF(w)/dw$ exists for all $w > 0$.

Problem: Choose x to maximize expected profit.

Note that $w \geq 0$ so $F(w) = 0$ and $f(w) = 0$ for $w \leq 0$.

Newsvendor Problem
Model

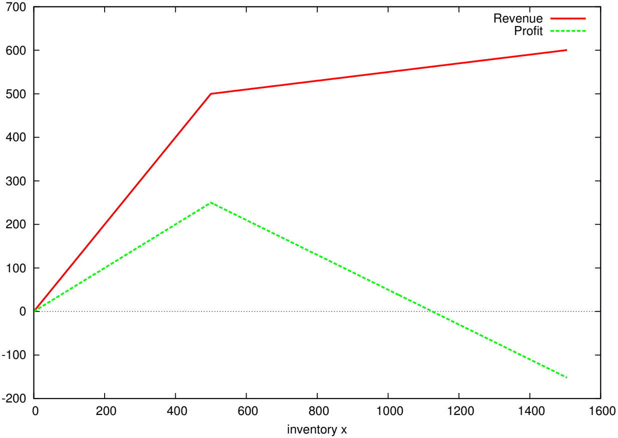
Revenue =

$$R(x,w) = \begin{cases} rx & \text{if } x \leq w \\ rw + s(x-w) & \text{if } x > w \end{cases}$$

Profit =

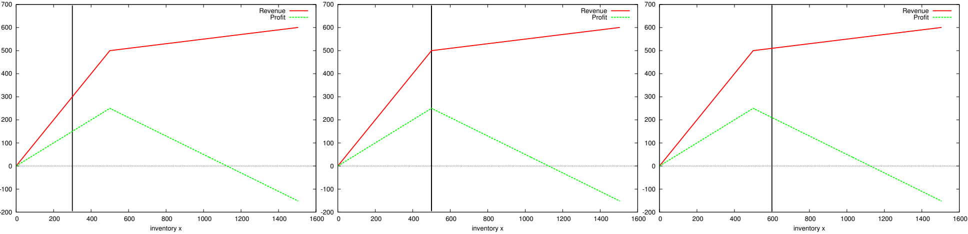
$$P(x,w) = \begin{cases} (r-c)x & \text{if } x \leq w \\ rw + s(x-w) - cx & \text{if } x > w \\ \quad = (r-s)w - (c-s)x & \end{cases}$$

Newsvendor Problem
Model



$$r = 1.; c = .5; s = .1; w = 500.$$

Newsvendor Problem
Model

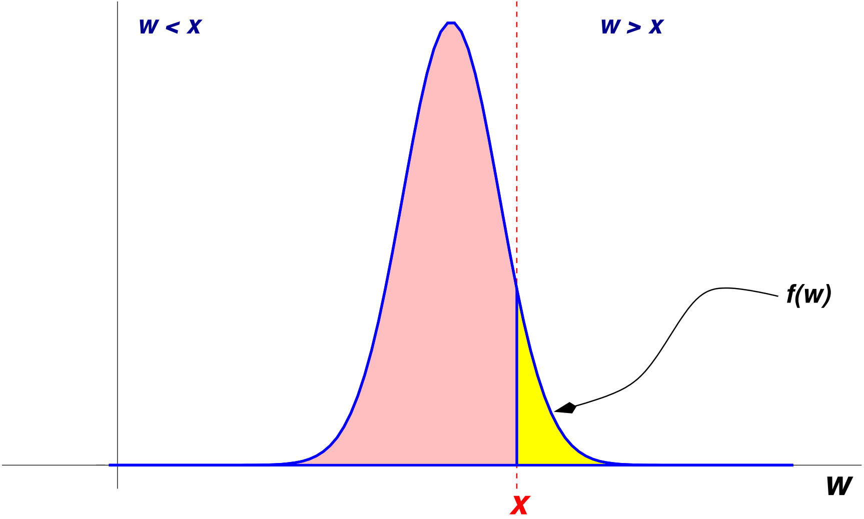


x too little x just right x too much

$$r = 1.; c = .5; s = .1; w = 500.$$

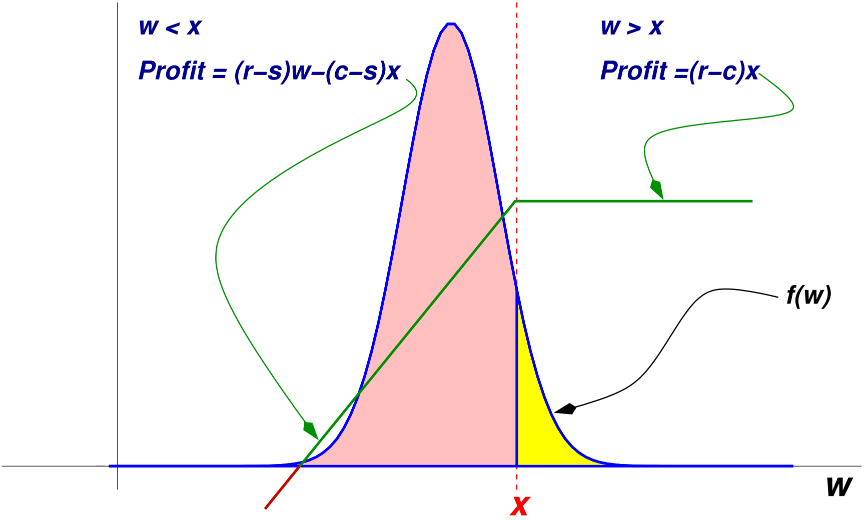
Note that if x is too large, the profit could be negative.

Newsvendor Problem
Normally distributed demand



Newsvendor Problem

Normally distributed demand



Newsvendor Problem
Expected profit

Expected Profit $= EP(x) = E_w P(x,w) =$

$$\int_{-\infty}^{\infty} P(x,w) f(w) dw =$$

$$\int_{-\infty}^{x} [(r-s)w - (c-s)x] f(w) dw +$$

$$\int_{x}^{\infty} (r-c) x f(w) dw$$

Newsvendor Problem

Expected profit

or $EP(x)$

$$= (r-s)\int_{-\infty}^{x} wf(w)dw - (c-s)x\int_{-\infty}^{x} f(w)dw$$

$$+(r-c)x\int_{x}^{\infty} f(w)dw$$

$$= (r-s)\int_{-\infty}^{x} wf(w)dw$$

$$-(c-s)xF(x)+(r-c)x(1-F(x))$$

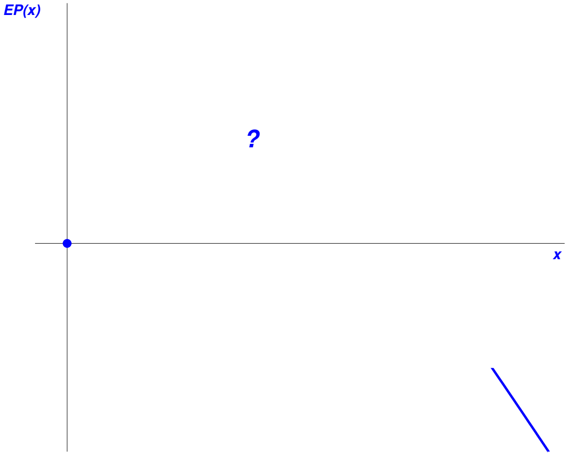
Newsvendor Problem
Expected profit

- When $x = 0$, EP is 0.
- When $x \rightarrow \infty$,
  - ⋆ the first term goes to a finite constant, $(r - s)Ew$;
  - ⋆ the middle term, $-(c - s)xF(x), \rightarrow -\infty$.
  - ⋆ the last term goes to 0;

Therefore when $x \rightarrow \infty$, $EP \rightarrow -\infty$.

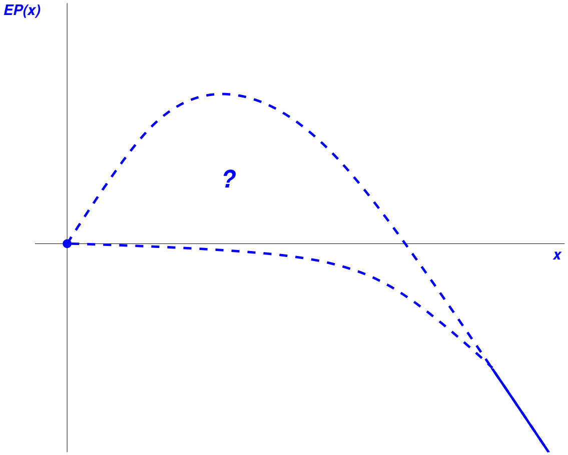
Newsvendor Problem

Expected profit



Newsvendor Problem

Expected profit



Newsvendor Problem

Expected profit

$$EP(x) = (r - s) \int_{-\infty}^{x} w f(w) dw - (c - s) x F(x) + (r - c) x (1 - F(x))$$

$$\frac{dEP}{dx} = (r - s) x f(x) - (c - s) F(x) - (c - s) x f(x)$$
$$+ (r - c)(1 - F(x)) - (r - c) x f(x)$$

$$= x f(x)(r - s + s - c - r + c)$$
$$+ r - c + (s - c - r + c) F(x)$$

$$= r - c - (r - s) F(x)$$

Note also that $\dfrac{d^2 EP}{dx^2} = -(r - s) f(x)$.

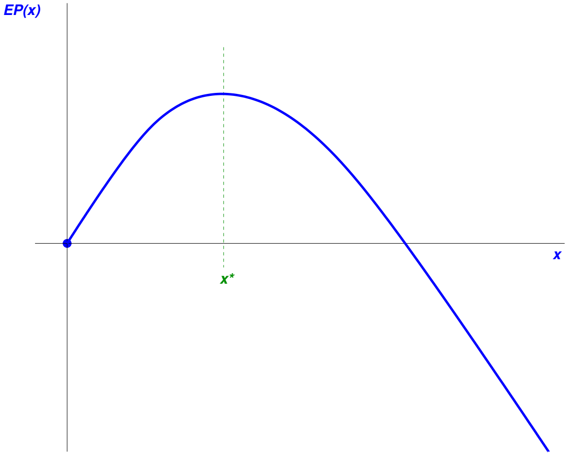
Newsvendor Problem

Expected profit

Note that

- $d^2EP/dx^2 = -(r-s)f(x) < 0$ for all $x > 0$. Therefore EP is concave and has a maximum.
- $dEP/dx = r - c > 0$ when $x = 0$. Therefore EP is increasing at $x = 0$, so $EP(x)$ is positive for some x.
- x^* satisfies $\dfrac{dEP}{dx}(x^*) = 0$.

Newsvendor Problem
Expected profit



Newsvendor Problem
Expected profit

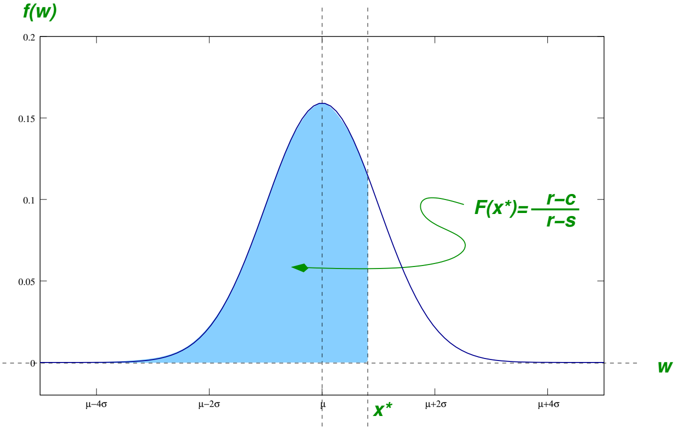
Therefore $$\boxed{F(x^*) = \frac{r-c}{r-s}.}$$

Note that $0 \leq \dfrac{r-c}{r-s} \leq 1$. Therefore there is an x^*.
Recall the definition of $F()$: $F(x^*)$ is the probability that $W \leq x^*$.
So the equation in the box means

Buy enough stock to satisfy demand $100K\%$ of the time, where

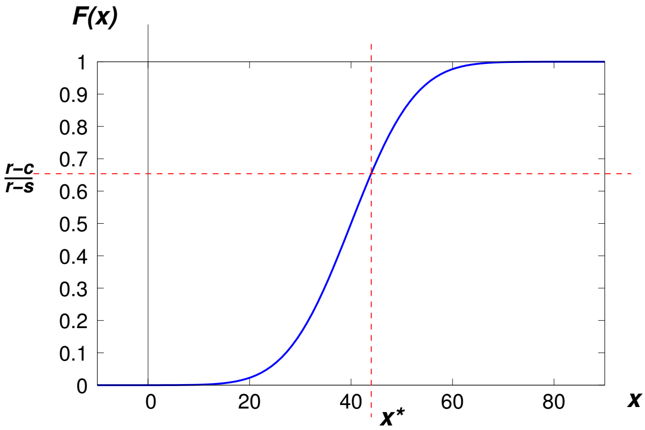
$$K = \frac{r-c}{r-s}.$$

Newsvendor Problem
Expected profit



Newsvendor Problem

Expected profit



Newsvendor Problem
Expected profit

Note that $F(x^*) = \dfrac{r-c}{r-s} = \dfrac{r-c}{(r-c)+(c-s)}$.

$r - c > 0$ is the marginal profit when $x < w$.

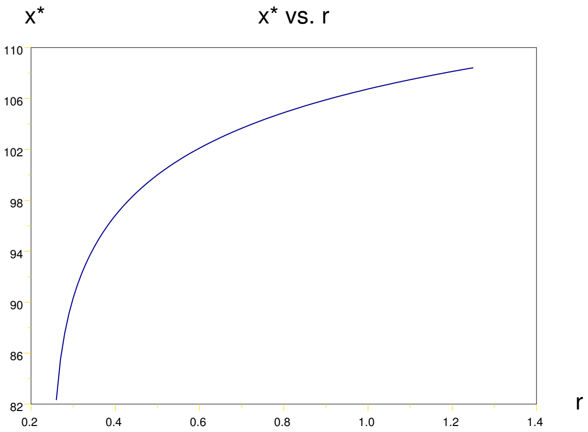
$c - s > 0$ is the marginal loss when $x > w$.

Choose x^* so that the fraction of time you do not buy too much is

$$\frac{\text{marginal profit}}{\text{marginal profit} + \text{marginal loss}}$$

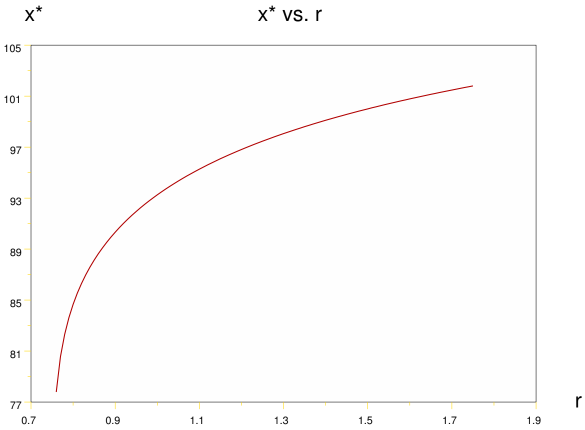
Newsvendor Problem

Example 1: r varying, $c = .25, s = 0, \mu_w = 100, \sigma_w = 10$



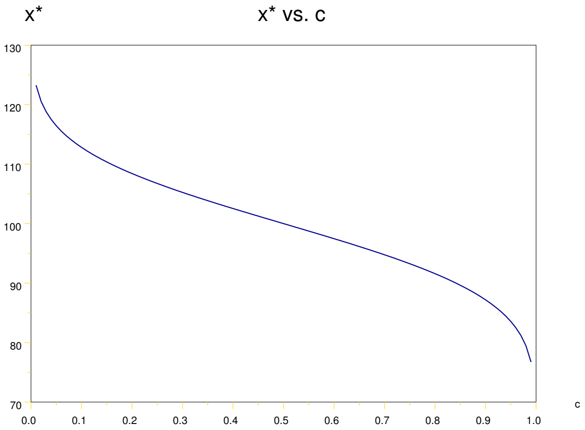
Newsvendor Problem

Example 2: r varying, $c = .75, s = 0, \mu_w = 100, \sigma_w = 10$



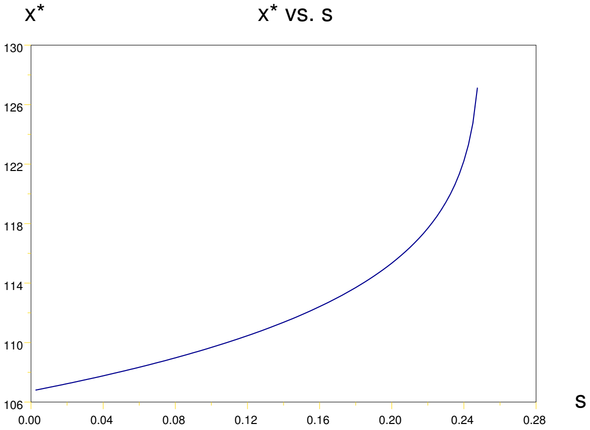
Newsvendor Problem

$r = 1$, c varying, $s = 0, \mu_w = 100, \sigma_w = 10$



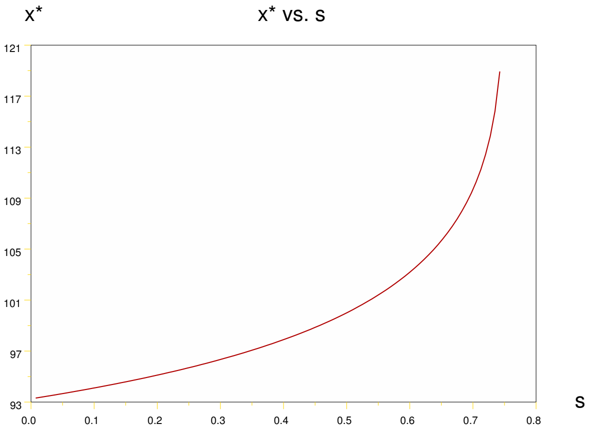
Newsvendor Problem

Example 1: $r = 1, c = .25$, s varying, $\mu_w = 100, \sigma_w = 10$



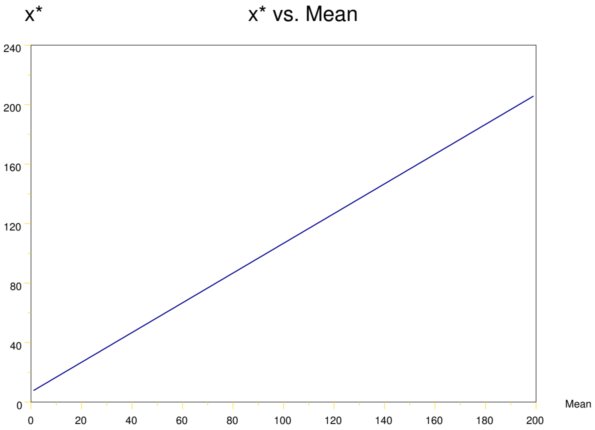
Newsvendor Problem

Example 2: $r = 1, c = .75$, s varying, $\mu_w = 100, \sigma_w = 10$



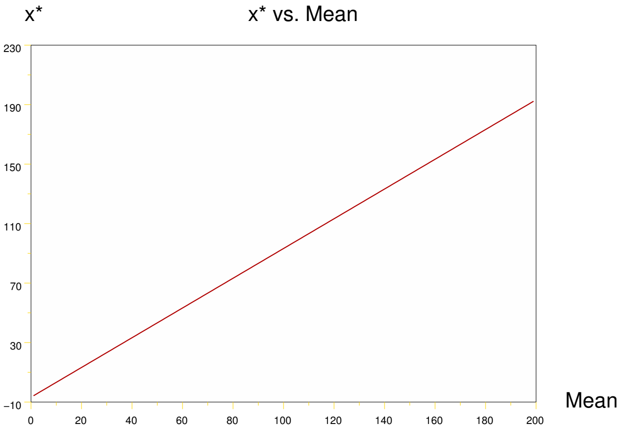
Newsvendor Problem

Example 1: $r = 1, c = .25, s = 0$, μ_w varying, $\sigma_w = 10$



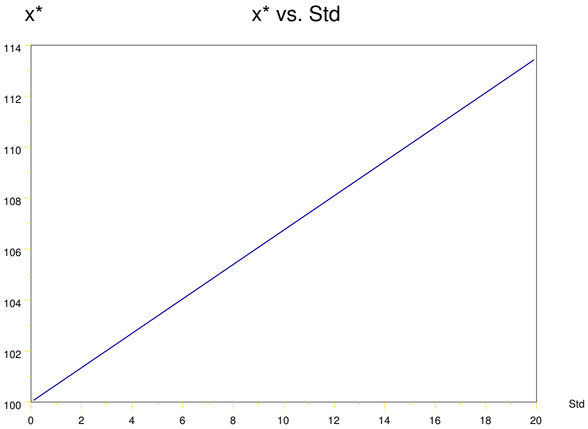
Newsvendor Problem

Example 2: $r = 1, c = .75, s = 0,$ μ_w varying, $\sigma_w = 10$



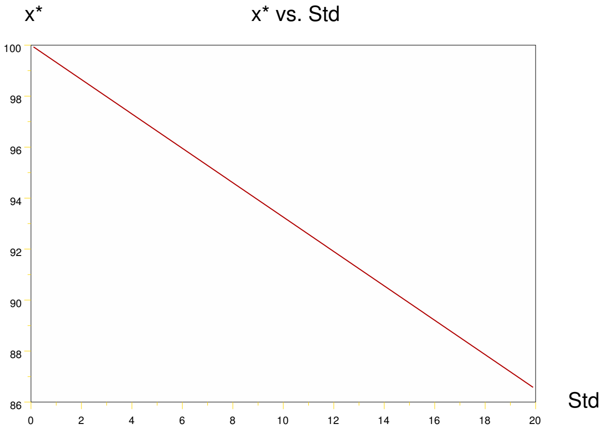
Newsvendor Problem

Example 1: $r = 1, c = .25, s = 0, \mu_w = 100$, σ_w varying



Newsvendor Problem

Example 2: $r = 1, c = .75, s = 0, \mu_w = 100,\ \sigma_w$ varying



Newsvendor Problem

Why does x^* look linear in μ_w and σ_w?

- x^* is the solution to $F(x^*) = \dfrac{r-c}{r-s}$.
- Assume demand w is $N(\mu_w, \sigma_w)$. Then, for any demand w,

$$F(w) = \Phi\left(\frac{w-\mu_w}{\sigma_w}\right)$$

where Φ is the standard normal cumulative distribution function.

Newsvendor Problem

Why does x^* look linear in μ_w and σ_w?

- Therefore

$$\Phi\left(\frac{x^* - \mu_w}{\sigma_w}\right) = \frac{r-c}{r-s}$$

- or

$$\frac{x^* - \mu_w}{\sigma_w} = \Phi^{-1}\left(\frac{r-c}{r-s}\right)$$

- or

$$\boxed{x^* = \mu_w + \sigma_w \Phi^{-1}\left(\frac{r-c}{r-s}\right)}$$

Newsvendor Problem

Why does x^* look linear in μ_w and σ_w? And should it?

Is x^* actually linear in μ_w and σ_w?

Answer: yes and no.

Yes if we accept the assumption that the demand is normal. But ...

No in reality because the demand cannot be normal.

For what values of μ_w and σ_w is x^* closest to linear in μ_w and σ_w?

Answer: When $\mu_w >> \sigma_w$. If so, the normal distribution predicts a very small probability for w being negative. Therefore the normal is a good approximation for the truncated normal.

Newsvendor Problem

x^* Increasing or Decreasing in σ_w

What determines whether x^* increases or decreases in σ_w?

- Note that
$$\Phi^{-1}(k) > 0 \text{ if } k > .5$$
and
$$\Phi^{-1}(k) < 0 \text{ if } k < .5$$
- Therefore, since $x^* = \mu_w + \sigma_w \Phi^{-1}\left(\frac{r-c}{r-s}\right)$,
    - ⋆ x^* increases with σ_w if $\frac{r-c}{r-s} > .5$, and
    - ⋆ x^* decreases with σ_w if $\frac{r-c}{r-s} < .5$.

Newsvendor Problem

A General Strategy

Question:

- Can we extend this strategy to manage inventory in other settings?
- In particular, suppose the stock remaining at the end of the day today can be used in the future. Does the result of the newsvendor problem suggest a possible *heuristic* extension?

Answer: Yes.

"Build enough inventory to satisfy demand $100K\%$ of the time, for some K."

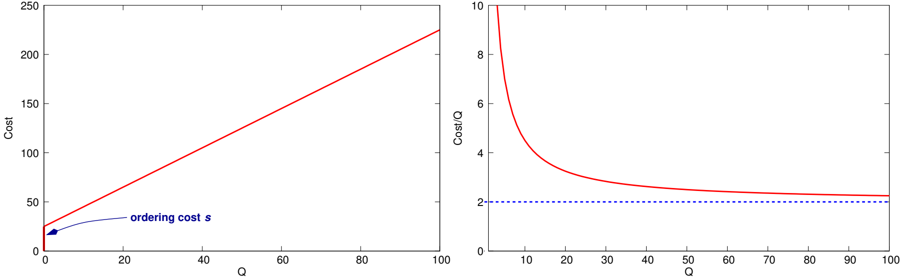
Economic Order Quantity
How much to order and when

- *Economic Order Quantity*
- Motivation: economy of scale in ordering.
- Tradeoff:
  - ⋆ Each time an order is placed, the company incurs a fixed cost in addition to the cost of the goods. *This motivates infrequent ordering.*
  - ⋆ The order must be large enough to satisfy demand until the next order is placed. This determines how much inventory must be stored.

Economic Order Quantity
Assumptions

- No randomness.
- No delay between ordering and arrival of goods.
- No backlogs.
- Goods are required at an annual rate of λ units per year. Inventory is therefore depleted at the rate of λ units per year.
- It costs h to store one unit for one year. h is the *holding cost* (\$/unit/year).
- If the company orders Q units, it must pay $s + cQ$ for the order. s is the *ordering cost* (\$). c is the unit cost (\$/unit).

Economic Order Quantity
Assumptions



Total cost and cost per unit of an order

$$s = 25, c = 2$$

Economic Order Quantity
Problem

- Find a strategy for ordering materials that will minimize the total cost *per year.*

- There are two costs to consider: the *annual* ordering cost and the *annual* holding cost.

Economic Order Quantity
Scenario

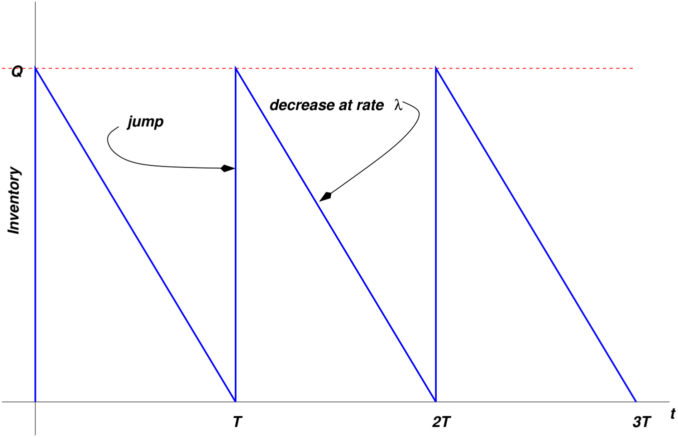
- At time 0, inventory level is 0.
- Q units are ordered and the inventory level jumps instantaneously to Q.
- Material is depleted at rate λ.
- Since the problem is totally deterministic, we can wait until the inventory goes to zero before we order next. There is no danger that the inventory will go to zero earlier than we expect it to.

Economic Order Quantity
Scenario

- Because of the very simple assumptions, we can assume that the optimal strategy does not change over time.
- Therefore the policy is to order Q each time the inventory goes to zero.
- We must determine the optimal Q.

Economic Order Quantity

Scenario



$$T = Q/\lambda \text{ (years)}$$

Economic Order Quantity
Formulation

- The number of orders in a year is $1/T = \lambda/Q$. Therefore, the ordering cost in a year is $s\lambda/Q$.
- The average inventory is $Q/2$. Therefore the average inventory holding cost is $hQ/2$.
- Therefore, we must minimize the annual cost

$$C = \frac{hQ}{2} + \frac{s\lambda}{Q}$$

over Q.

Economic Order Quantity
Formulation

Then

$$\frac{dC}{dQ} = \frac{h}{2} - \frac{s\lambda}{Q^2} = 0$$

or

$$Q^* = \sqrt{\frac{2s\lambda}{h}}$$

Economic Order Quantity
Examples

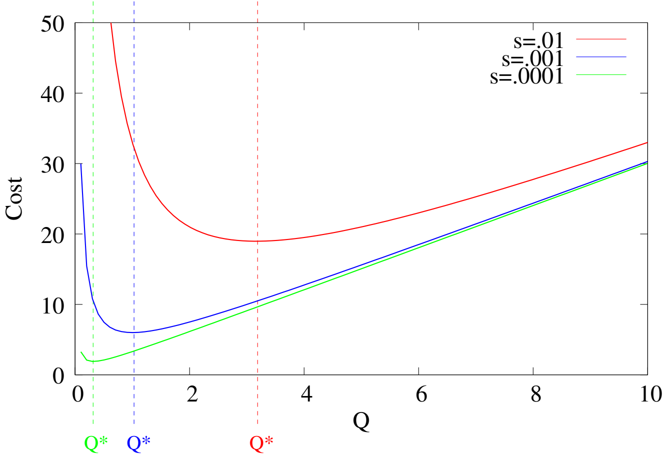
In the following graphs, the base case is

- $\lambda = 3000$
- $s = .001$
- $h = 6$

Note that

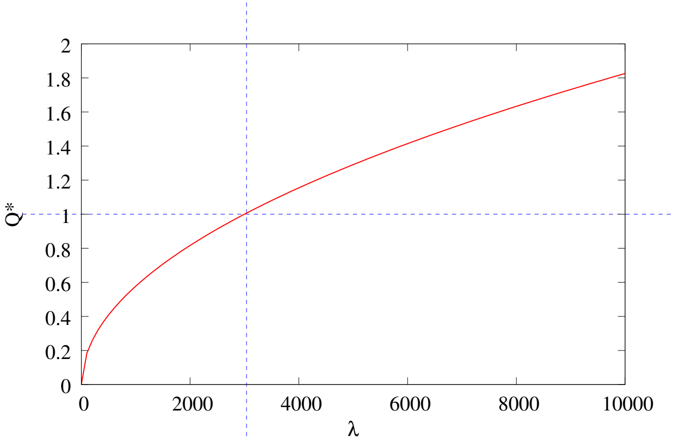
$$Q^* = 1$$

Economic Order Quantity
Examples

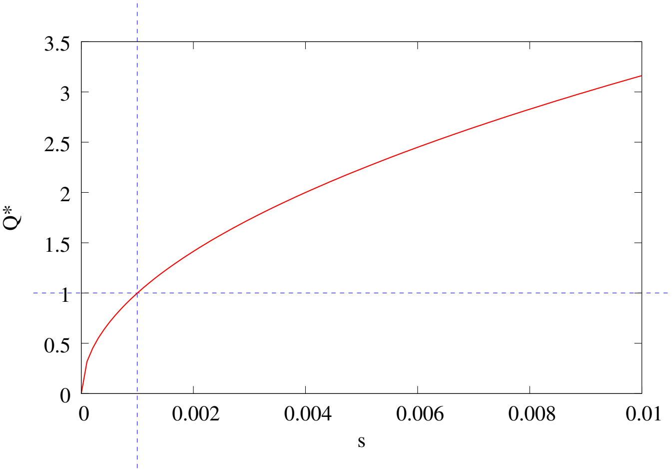


Economic Order Quantity

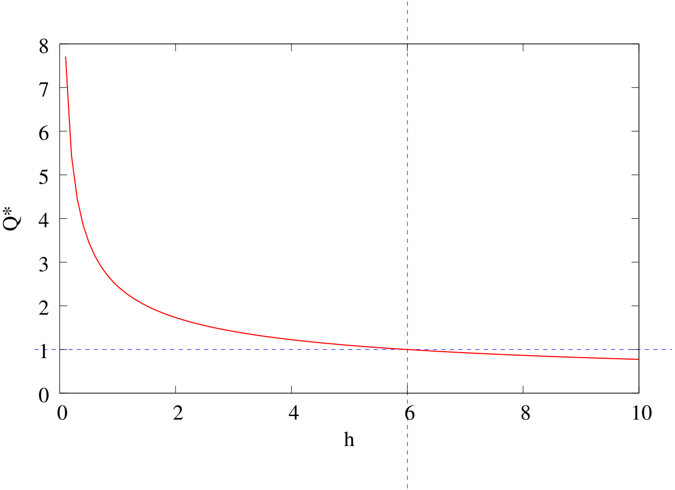
Examples



Economic Order Quantity
Examples



Economic Order Quantity
Examples



Supplier Lead Time + Random Demand

Issues

Issue: Supplies are not delivered instantaneously. The time between order and delivery, the lead time $L > 0$.

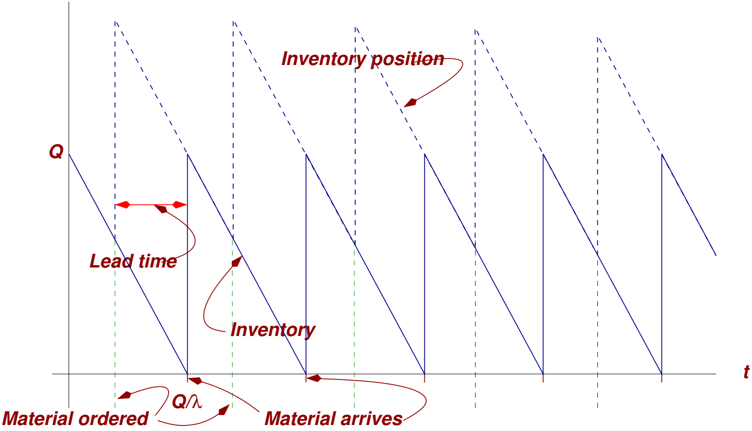
- If everything were deterministic, this would not be a problem. Just order earlier.

Issue: ... but demand is random.

- *Problem:* The demand that occurs between when the order is placed and and when the goods are delivered might be large enough to cause stockout.

Other possible issues: Lead time could be random, production quantity could be random, etc.

Supplier Lead Time + Random Demand
Inventory Position



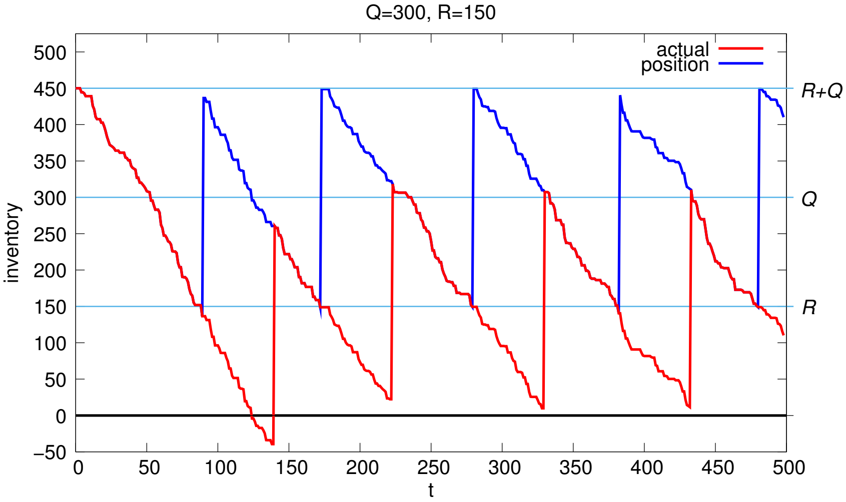
Supplier Lead Time + Random Demand

Q, R Policy

- Fixed ordering cost, like in EOQ problem.
- Random demand, like in newsvendor problem.
  - ⋆ *But the inventory retains its value. It is not scrapped.*
- Make to stock — no advance ordering by customers.
- *Policy:* when the inventory level goes below R, order a quantity Q.
- *Problem:* find Q and R to minimize ordering + holding costs while not losing too many sales due to stockout.

Random Demand

Q, R Policy



Random Demand

R

First we describe how to choose R.

- Let L be the number of days between when an order is placed and when the shipment arrives, the order lead time. L is constant.
- Let D be the total demand that arrives between when an order is placed and when the shipment arrives (L days). D is a random variable.
- The order is placed when the inventory level goes below R.
- The factory would like R to be greater than the demand D that arrives before the raw material arrives.
- But since the demand is random, the best we can do is to choose R such that

$$P(D > R) \leq \epsilon$$

where $P()$ is the probability and ϵ is a small number.

Random Demand

R

It is convenient to write the last equation as

$$P(D < R) \geq 1 - \epsilon$$

If L is large enough, the Central Limit Theorem says that D is approximately normal regardless of the distribution of the daily demand.

Assume that D is $N(\mu, \sigma)$, a normal random variable with mean μ and standard deviation σ.

That is, the probability distribution of the demand is

$$P(D < d) = \Phi\left(\frac{d - \mu}{\sigma}\right)$$

Random Demand

R

Then we will choose R to satisfy

$$\Phi\left(\frac{R-\mu}{\sigma}\right) \geq 1-\epsilon$$

The smallest value of R that satisfies this is

$$R^* = \mu + \sigma\Phi^{-1}(1-\epsilon)$$

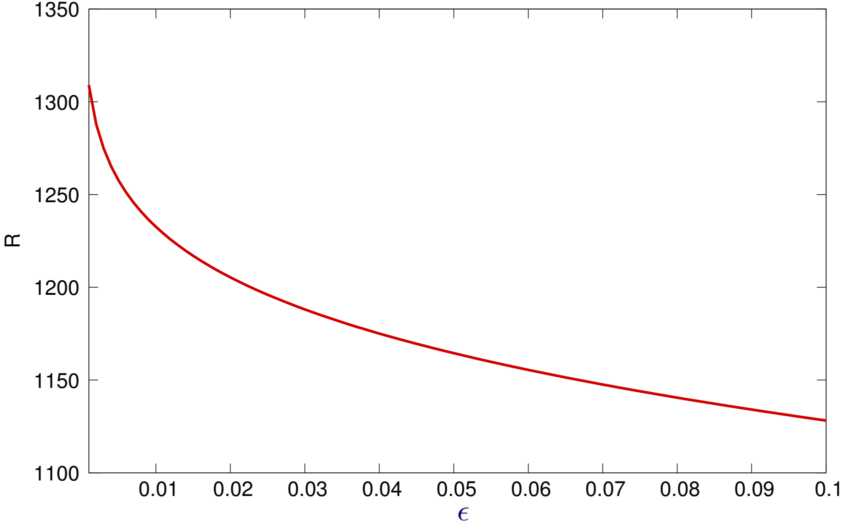
Recall that in the EOQ formulation, λ was the yearly demand. Here, μ is the mean demand for L days. Therefore,

$$\mu = \frac{L\lambda}{365}$$

A graph of R^* as a function of ϵ with $\mu = 1000$ and $\sigma = 100$ is on the next slide.

Random Demand

R



Random Demand

Q

Now we describe how to choose Q.

- The *expected* time between orders is $T = Q/\lambda$.
- The expected number of orders in a year is λ/Q. Therefore, the expected ordering cost in a year is $s\lambda/Q$.
- Assume that the holding cost for material in transit is the same as for inventory in the factory. The average inventory position is $(R + Q)/2$ and the average inventory holding cost is $h(R + Q)/2$.

Random Demand

Q

- Therefore, we choose Q to minimize the approximate expected annual cost

$$C = \frac{h}{2}(R + Q) + \frac{s\lambda}{Q}$$

But since R is constant, this is essentially the same as the EOQ problem for determining Q.

$$Q^* = \sqrt{\frac{2s\lambda}{h}}$$

MIT 2.853/2.854

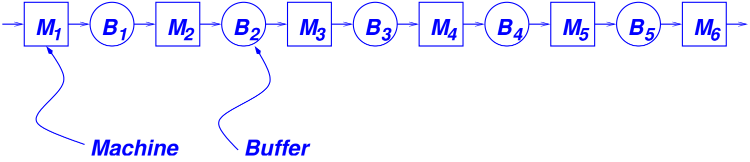
Introduction to Manufacturing Systems

Single-part-type, multiple stage systems

Stanley B. Gershwin
Laboratory for Manufacturing and Productivity
Massachusetts Institute of Technology
gershwin@mit.edu

Flow Lines

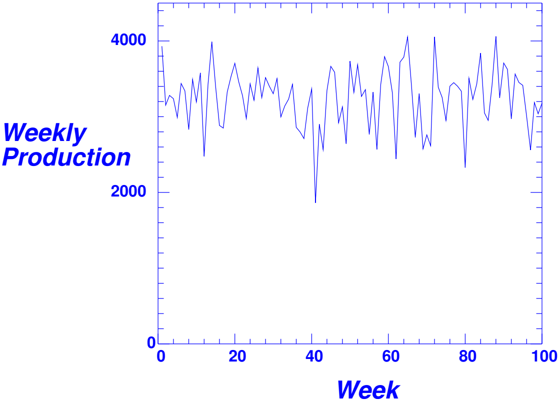
... also known as a Production or Transfer Lines.



- Machines are unreliable.
- Buffers are finite.
- In many cases, the operation times are constant and equal for all machines.

Flow Lines

Output Variability



Production output from a simulation of a transfer line.

Single Reliable Machine

- If the machine is perfectly reliable, and its average operation time is τ, then its maximum production rate is $\mu = 1/\tau$.

- *Note:*
  - ⋆ Sometimes *cycle time* is used instead of *operation time,* but *BEWARE:* cycle time has (at least) two meanings!
  - ⋆ The other meaning is the time a part spends in a system. If the system is a single, reliable machine, the two meanings are the same.

Single Unreliable Machine

Failures and Repairs

- A machine is either *up* (operational) or *down* (being repaired or maintained).
- MTTF = mean time to fail.
- MTTR = mean time to repair
- MTBF = MTTF + MTTR

Single Unreliable Machine

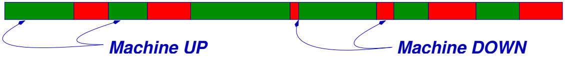
Production rate

- If the machine is unreliable, and
  - ⋆ its average operation time is τ,
  - ⋆ its mean time to fail is MTTF,
  - ⋆ its mean time to repair is MTTR,

  then its maximum production rate is

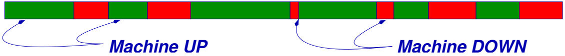
$$\frac{1}{\tau}\left(\frac{\text{MTTF}}{\text{MTTF}+\text{MTTR}}\right)$$

Single Unreliable Machine
Proof



- Average production rate, while machine is up, is $1/\tau$.
- Average *duration* of an up period is MTTF.
- Average *production* during an up period is MTTF$/\tau$.
- Average *duration* of up-down period: MTTF + MTTR.
- Average *production* during up-down period: MTTF$/\tau$.

Single Unreliable Machine
Proof

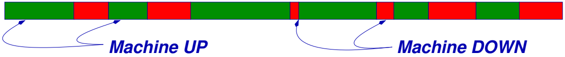


- Therefore, the average production rate is

$$\frac{\text{MTTF}/\tau}{\text{MTTF} + \text{MTTR}} = \frac{1}{\tau}\left(\frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}\right)$$

Single Unreliable Machine

Efficiency



$$\frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}} = e,$$ the *efficiency* of the machine.

- e is the fraction of time the machine is up.
- e can be thought of as the production rate in units of parts per operation time...
- ... or the actual production rate divided by what the production rate would be if the machine never failed.

Single Unreliable Machine
Geometric Up- and Down-Times

- *Assumptions:* Operation time is constant. Failure and repair times are *geometrically* distributed.

- Let p be the probability that a machine fails during any given operation. Then MTTF $= 1/p$.

Single Unreliable Machine

Geometric Up- and Down-Times

- Let r be the probability that M gets repaired during any operation time when it is down. Then MTTR $= 1/r$.

- Then the *average production rate* of M, in parts per operation time, is

$$e = \frac{r}{r+p}.$$

Single Unreliable Machine

Production Rates

- The *capacity* of a machine is its maximum possible production rate.
- The machine really has *three* capacities:
  - ⋆ 1 when it is up *(short-term capacity)* ,
  - ⋆ 0 when it is down *(short-term capacity)* ,
  - ⋆ $e = r/(r+p)$ on the average *(long-term capacity)* .

Multiple-Machine Lines

Failure Assumptions

- These assumptions are important for machines in systems, in which a machine can be made idle by another machine's failure.
- Operation-Dependent Failures (ODF)
  - ⋆ A machine can only fail while it is working — not idle.
  - ⋆ Idleness occurs due to starvation or blockage.
  - ⋆ This is the usual assumption.
- Time-Dependent Failures (TDF)
  - ⋆ A machine can fail even if it is idle.

Infinite-Buffer Lines

- Starvation: Machine M_i is starved at time t if Buffer B_{i-1} is empty at time t.

Assumptions:

- A machine is not idle if it is not starved.

- The first machine is never starved.

Infinite-Buffer Lines

Bottleneck

- The production rate of the line is the production rate of the *slowest* machine in the line — called the *bottleneck* .

- Slowest means *least average production rate.*

Infinite-Buffer Lines
Bottleneck

- The production rate is therefore

$$P = \min_i \frac{1}{\tau_i} \left(\frac{\text{MTTF}_i}{\text{MTTF}_i + \text{MTTR}_i} \right)$$

- and M_i is the bottleneck.

Infinite-Buffer Lines
Bottleneck

- Or, if all $\tau_i = 1$,

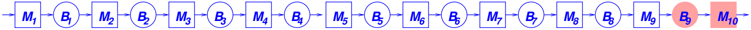
$$P = \min_i \left(\frac{\text{MTTF}_i}{\text{MTTF}_i + \text{MTTR}_i} \right)$$

- and M_i is the bottleneck.

Infinite-Buffer Lines
Bottleneck

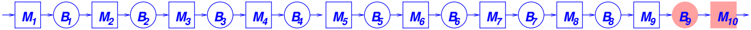
- The system is not in steady state.
- An increasing amount of inventory accumulates in the buffer upstream of the bottleneck.
- A finite amount of inventory appears downstream of the bottleneck.

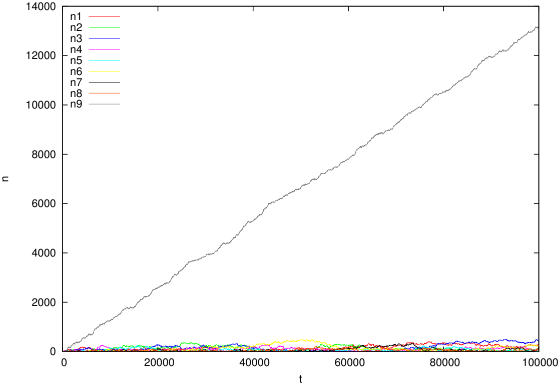
Infinite-Buffer Lines
Example 1



- Parameters:
  $r_i = .1, p_i = .01, i = 1, ..., 9; r_{10} = .1, p_{10} = .03.$
- Therefore, $e_i = .909, i = 1, ..., 9; e_{10} = .769.$

Infinite-Buffer Lines
Example 1





Infinite-Buffer Lines

Example 1

- *Question:* what is the the rate of growth of $n_9(t)$, the inventory in B_9?

- *Answer:*

  - ⋆ The rate that parts arrive at buffer 9 is .909.
  - ⋆ The rate that parts leave is .769.
  - ⋆ Therefore the rate of increase is $.909 - .769 =$ *0.14* parts per operation time.

Infinite-Buffer Lines

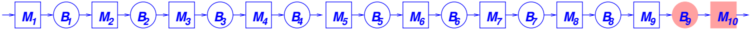
Example 1

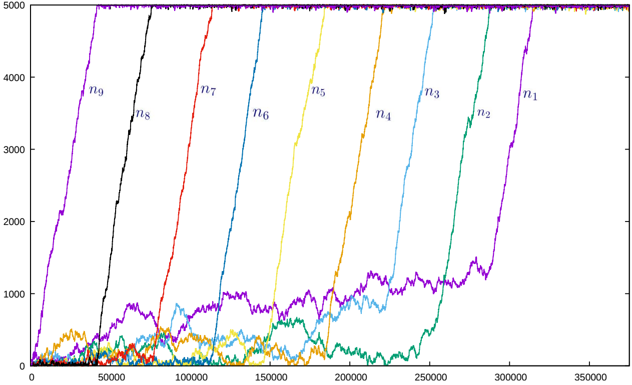
- *Question:* What happens when the buffers are large but finite?

- *Answer:* The last buffer gains material until it becomes full. Then the next to last buffer gains material until it becomes full. The process repeats until all buffers are full. See the graph on next slide.

Infinite-Buffer Lines

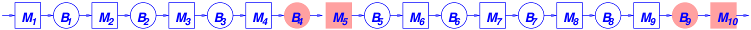
Example 1, but with finite buffers





Infinite-Buffer Lines

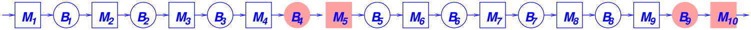
Second Bottleneck



- The *second bottleneck* is the slowest machine upstream of the bottleneck.
- An increasing amount of inventory accumulates just upstream of it.
- A finite amount of inventory appears between the second bottleneck and the machine upstream of the first bottleneck.
- An increasing amount of inventory accumulates in the buffer just upstream of the first bottleneck.
- A finite amount of inventory appears downstream of the first bottleneck.

Infinite-Buffer Lines

Example 2

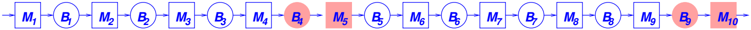


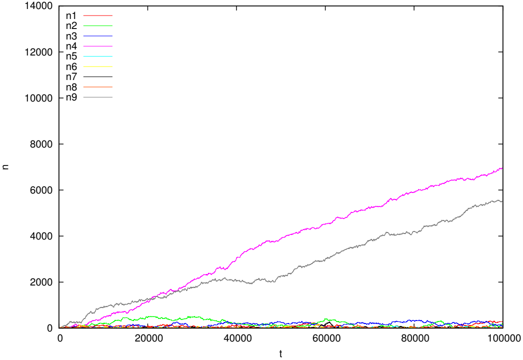
- Parameters: $r_i = .1, p_i = .01, i = 1, ..., 4, 6, ..., 9;$
  $r_5 = .1, p_5 = .02, r_{10} = .1, p_{10} = .03.$

- Therefore, $e_i = .909, i = 1, ..., 4, 6, ..., 9;$
  $e_5 = .833, e_{10} = .769.$

Infinite-Buffer Lines

Example 2





Infinite-Buffer Lines

Example 2

Rates of growth of $n_4(t)$ and $n_9(t)$

- The rate of growth of $n_4(t)$ is $.909 - .833 = .076$
- The rate of growth of $n_9(t)$ is $.833 - .769 = .064$
- Note that when t is large enough, $n_4(t) > n_9(t)$.
- It is tempting to believe that the easiest way to find the bottleneck of a line is to look for the greatest accumulation of inventory. *Is that correct?*

Infinite-Buffer Lines

Improvements

Questions:

- If we want to increase production rate, which machine should we improve?
- What would happen to production rate if we improved any other machine?

Simulation Note

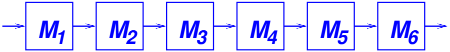
- The simulations shown here were *discrete-time* rather than *discrete-event.*
- Discrete-time simulations are easier to program, but less general, less accurate, and slower, than discrete-event simulations.
- Discrete-time simulations are easiest to write for systems where all event times are geometrically distributed.

Simulation Note

Discrete-time simulation

- Assume that some event occurs according to a geometric probability distribution and it has a mean time to occur of T time steps.
- Then the probability that it occurs in any given time step is $p = 1/T$.
- The discrete-time simulation logic is:
  - ⋆ At each time step, choose x, a U[0,1] random number.
  - ⋆ If $x \leq p$, the event has occurred. Change the state accordingly.
  - ⋆ If $x > p$, the event has not occurred. Change the state accordingly.

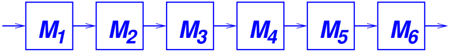
Zero-Buffer Lines



- If any one machine fails, or takes a very long time to do an operation, *all* the other machines must wait.

- Therefore the production rate is usually less — possibly much less – than the slowest machine.

Zero-Buffer Lines



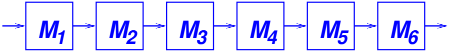
- *Example:* Constant, unequal operation times, perfectly reliable machines.

  - ⋆ The operation time of the line is equal to the maximum operation time of the of all the machines, ...
  - ⋆ ... so the production rate of the line is the inverse of the maximum operation time.

- *Line balancing:* the assignment of tasks to machines in a way that minimizes the maximum operation time.

Zero-Buffer Lines

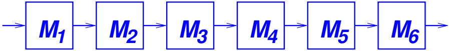
Constant, equal operation times, unreliable machines



- Assumption: Time-Dependent Failures (TDF).
- The operation time is the time unit.
- $e_i = \frac{\text{MTTF}_i}{\text{MTTR}_i + \text{MTTF}_i}$ is the probability that M_i is operational in any time unit.

Zero-Buffer Lines

Constant, equal operation times, unreliable machines



- A part is produced as often as *all* the machines are operational.
- Since machines fail independently of the states of the other machines, the probability that all machines are operational is $e_1 e_2 ... = \Pi_i e_i$.
- Therefore, the production rate of the line is

$$\boxed{E_{\text{TDF}} = \Pi_i e_i}$$

parts per operation time.

Zero-Buffer Lines

Constant, equal operation times, unreliable machines

$$\rightarrow \boxed{M_1} \rightarrow \boxed{M_2} \rightarrow \boxed{M_3} \rightarrow \boxed{M_4} \rightarrow \boxed{M_5} \rightarrow \boxed{M_6} \rightarrow$$

- *Assumption:* Operation-Dependent Failures (ODF).
- The operation time is the time unit.
- *Assumption:* Failure and repair times are *geometrically* distributed.
- $p_i = 1/\text{MTTF}_i =$ probability of failure during an operation.
- $r_i = 1/\text{MTTR}_i =$ probability of repair during an operation time when the machine is down.

Zero-Buffer Lines

Production Rate

$$\rightarrow \boxed{M_1} \rightarrow \boxed{M_2} \rightarrow \boxed{M_3} \rightarrow \boxed{M_4} \rightarrow \boxed{M_5} \rightarrow \boxed{M_6} \rightarrow$$

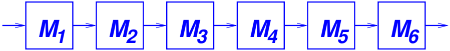
Buzacott's Zero-Buffer Line Formula:

Let k be the number of machines in the line. Then the production rate per operation time is

$$\boxed{E_{\text{ODF}} = \frac{1}{1 + \sum_{i=1}^{k} \frac{p_i}{r_i}}}$$

Zero-Buffer Lines

Production Rate



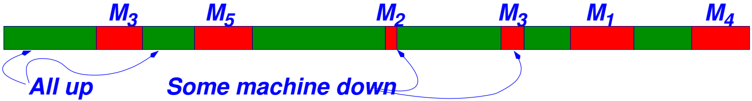
- This reduces to the earlier formula (Slides **??** and **??**) when $k = 1$.
- According to the new formula, the *isolated production rate* of a single machine M_1 is

$$\frac{1}{1+\frac{p_1}{r_1}} = \frac{r_1}{r_1+p_1}.$$

which is the same as the earlier formula when the operation time τ is 1.

Zero-Buffer Lines
Proof of formula

- *Approximation:* At most, only <u>one</u> machine can be down at any time.
- Consider a long time interval of length T operation times during which Machine M_i fails m_i times ($i = 1, \ldots k$).



- Without failures, the line would produce T parts.

Zero-Buffer Lines
Proof of formula

- The average repair time of M_i is $1/r_i$ each time it fails, so the total system down time is close to

$$D = \sum_{i=1}^{k} \frac{m_i}{r_i}$$

where D is the number of operation times in which a machine is down.

Zero-Buffer Lines

Proof of formula

- The total up time is approximately

$$U = T - \sum_{i=1}^{k} \frac{m_i}{r_i}$$

- where U is the number of operation times in which all machines are up.

Zero-Buffer Lines

Proof of formula

- Since the system produces one part per time unit while it is working, it produces U parts during the interval of length T.
- Note that, approximately,

$$m_i = p_i U$$

because M_i can only fail while it is operational.

Zero-Buffer Lines
Proof of formula

- Thus,

$$U = T - U \sum_{i=1}^{k} \frac{p_i}{r_i},$$

or,

$$\frac{U}{T} = E_{\text{ODF}} = \frac{1}{1 + \sum_{i=1}^{k} \frac{p_i}{r_i}}$$

Zero-Buffer Lines

p_i and r_i and p_i/r_i

$$P = E_{\mathsf{ODF}} = \frac{1}{1 + \sum_{i=1}^{k} \frac{p_i}{r_i}}$$

- Note that P is a function of the *ratio* p_i/r_i and not p_i or r_i separately.
- The same statement is true for the infinite-buffer line.
- However, the same statement is *not* true for a line with finite, non-zero buffers.

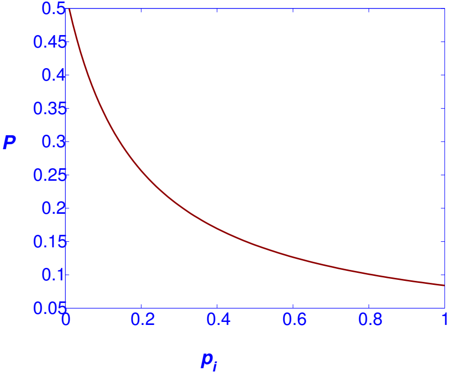
Zero-Buffer Lines
Improvements

Questions:

- If we want to increase production rate, which machine should we improve?
- What would happen to production rate if we improved any other machine?

Zero-Buffer Lines
P as a function of p_i

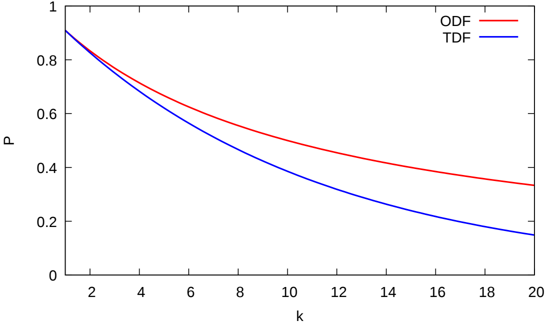
All machines are the same except M_i. As p_i increases, the production rate decreases.



Zero-Buffer Lines

P as a function of k

All machines are the same. This compares E_{TDF} and E_{ODF}.

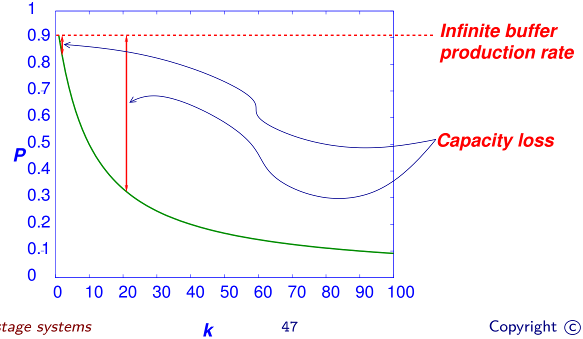


Zero-Buffer Lines

P as a function of k

This graph compares the production rates of an infinite-buffer line with that of a zero-buffer line.

All machines are the same. As k increases, the production rate of the zero-buffer line decreases. The production rate of the infinite-buffer line stays the same.



Finite-Buffer Lines

- Motivation for buffers: increase production rate.
- Cost
  - ⋆ in-process inventory/lead time
  - ⋆ floor space
  - ⋆ material handling mechanism

Finite-Buffer Lines

- Infinite buffers: delayed downstream propagation of disruptions *(starvation)* and *no* upstream propagation.

- Zero buffers: instantaneous propagation in both directions.

- Finite buffers: delayed propagation in both directions.
  - ⋆ New phenomenon: *blockage.*

- Blockage: Machine M_i is blocked at time t if Buffer B_i is full at time t.

Finite-Buffer Lines

- Difficulty:
  - ⋆ No simple formula for calculating production rate, inventory levels, or other performance measures.
- Solutions:
  - ⋆ Simulation
  - ⋆ Analytical approximation
  - ⋆ Exact numerical solution for short lines
  - ⋆ *Exact analytical solution for two-machine lines only*

Two-Machine, Finite-Buffer Lines
Parameters

- p_1 is the probability of M_1 failing while doing an operation.
- r_1 is the probability of M_1 getting repaired during an operation time when it is down.
- p_2 is the probability of M_2 failing while doing an operation.
- r_2 is the probability of M_2 getting repaired during an operation time when it is down.
- N is the number of parts the buffer can hold.

Two-Machine, Finite-Buffer Lines

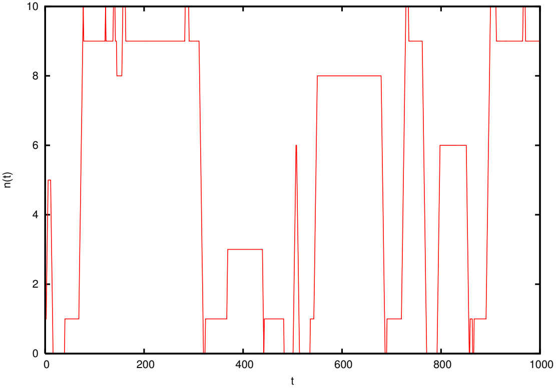
Simulations

Graphs of $n(t)$, the number of parts in the buffer of a two-machine line, are shown in the next four slides for four examples. The parameters of each example are chosen so that the lines have different behaviors.

- Example 1: $r_1 = .1, p_1 = .01, r_2 = .1, p_2 = .01, N = 10$
  Both machines are the same. The buffer is small.
- Example 2: $r_1 = .1, p_1 = .01, r_2 = .1, p_2 = .01, N = 100$
  Same as Example 1 except the buffer is large.
- Example 3: $r_i = .1, i = 1, 2, p_1 = .02, p_2 = .01, N = 100$
  Same as Example 2 except the first machine fails more often.
- Example 4: $r_i = .1, i = 1, 2, p_1 = .01, p_2 = .02, N = 100$
  Same as Example 2 except the second machine fails more often.

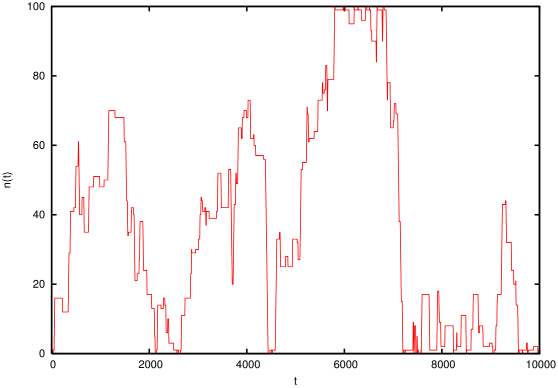
Two-Machine, Finite-Buffer Lines

Simulations



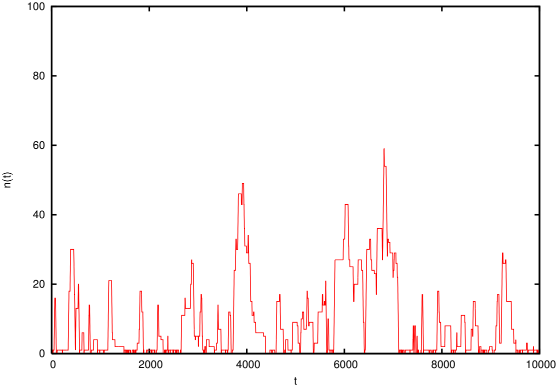
Example 1: $r_1 = .1, p_1 = .01, r_2 = .1, p_2 = .01, N = 10$

Two-Machine, Finite-Buffer Lines
Simulations



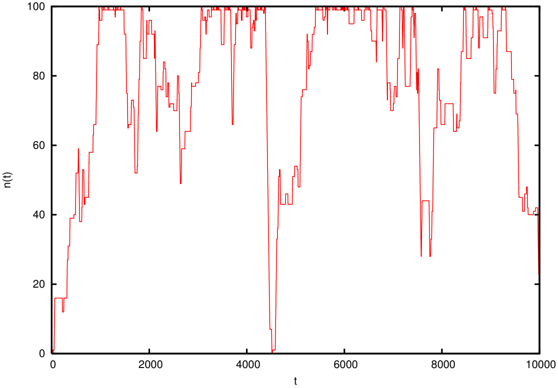
Example 2: $r_1 = .1, p_1 = .01, r_2 = .1, p_2 = .01, N = 100$

Two-Machine, Finite-Buffer Lines
Simulations



Example 3: $r_i = .1, i = 1, 2, p_1 = .02, p_2 = .01, N = 100$

Two-Machine, Finite-Buffer Lines
Simulations



Example 4: $r_i = .1, i = 1, 2, p_1 = .01, p_2 = .02, N = 100$

Two-Machine, Finite-Buffer Lines
Simulations

If we increase the buffer size of Example 4 from 100 to 200, what will happen to the production rate?

a. It will double it.

b. It will increase it substantially, but less than double.

c. It will increase it but only by a trivial amount.

d. It will have exactly zero effect.

e. None of the above.

Two-Machine, Finite-Buffer Lines
Simulations

If we increase the buffer size of Example 4 from 100 to 200, what will happen to the average inventory?

a. It will increase by something very close to 100.

b. It will double it.

c. It will increase it substantially, but less than double.

d. It will increase it but only by a trivial amount.

e. It will have exactly zero effect.

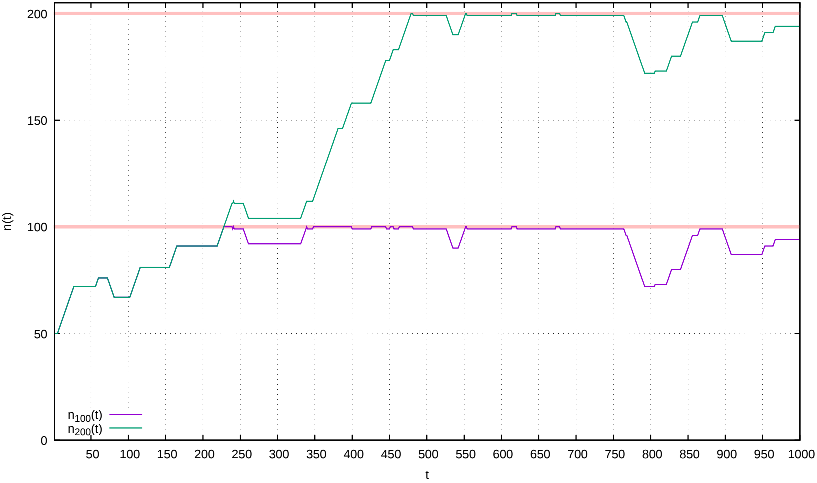
f. None of the above.

Two-Machine, Finite-Buffer Lines

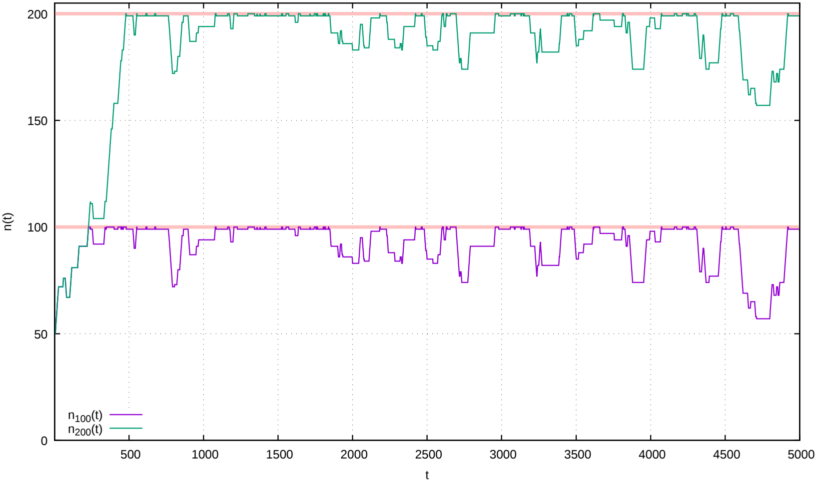
Simulations

Insert animated simulation here!!

Two-Machine, Finite-Buffer Lines
Simulations

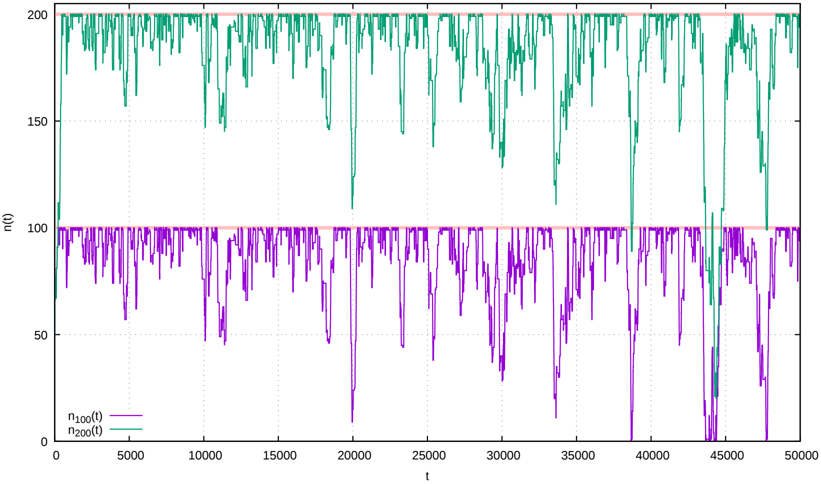


Two-Machine, Finite-Buffer Lines
Simulations



Two-Machine, Finite-Buffer Lines

Simulations



Two-Machine, Finite-Buffer Lines
Deterministic processing time model

- There are several models in the literature.
- We focus on the *Deterministic processing time,* or *Buzacott model:*
  - ⋆ deterministic processing time;
  - ⋆ geometric failure and repair times;
  - ⋆ operation-dependent failures;
  - ⋆ discrete state, discrete time.

Two-Machine, Finite-Buffer Lines
Markov Process Model

- *Discrete time-discrete state Markov process:*

$$\text{prob}\{X(t+1) = x(t+1) | X(t) = x(t), X(t-1) = x(t-1), X(t-2) = x(t-2), \ldots\} = \text{prob}\{X(t+1) = x(t+1) | X(t) = x(t)\}$$

- If we write $i = x(t+1), j = x(t)$, then

$$\text{prob}\{X(t+1) = i | X(t) = j\} = P_{ij}$$

is the transition matrix of the Markov process.

Two-Machine, Finite-Buffer Lines

State Space

- An analytical solution is available for a Markov process model of a two-machine line.

- Here, $X(t) = (n(t), \alpha_1(t), \alpha_2(t))$, where

  - ⋆ n is the number of parts in the buffer; $n = 0, 1, ..., N$.
  - ⋆ α_i is the repair state of M_i; $i = 1, 2$.
    - $\alpha_i = 1$ means the machine is *up* or *operational*;
    - $\alpha_i = 0$ means the machine is *down* or *under repair*.

Two-Machine, Finite-Buffer Lines

State Space

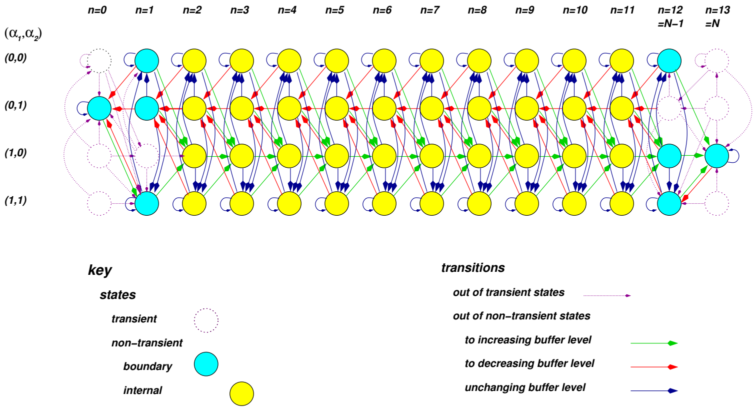
Examples of transition probabilities P_{ij}:

- If $j = x(t) = (3, 0, 1)$ and $i = x(t+1) = (7, 1, 1)$ then
  $P_{ij} = \text{prob}\{X(t+1) = x(t+1)|X(t) = x(t)\} = 0$
  - ⋆ because the buffer level cannot change by more than 1 part in 1 time unit.
- If $j = x(t) = (3, 0, 1)$ and $i = x(t+1) = (2, 0, 1)$ then
  $P_{ij} = \text{prob}\{X(t+1) = x(t+1)|X(t) = x(t)\} = (1 - r_1)(1 - p_2)$
  - ⋆ because the probability that the first machine does not get repaired is $1 - r_1$, the probability that the second machine does not fail is $1 - p_2$, and the buffer level goes down by 1.
- If $j = x(t) = (3, 0, 1)$ and $i = x(t+1) = ($**3**$, 0, 1)$ then
  $P_{ij} = \text{prob}\{X(t+1) = x(t+1)|X(t) = x(t)\} = 0$
  - ⋆ because the buffer level *must* go down by 1.

Two-Machine, Finite-Buffer Lines

State Space

State Transition Graph for Deterministic Processing Time, Two−Machine Line



Two-Machine, Finite-Buffer Lines
Calculation of performance measures

- $\pi(n, \alpha_1, \alpha_2)$ is the steady-state probability that there are n parts in the buffer and the machine states are α_1 and α_2.

- To obtain $\pi(n, \alpha_1, \alpha_2)$ for all (n, α_1, α_2), let $i = (n, \alpha_1, \alpha_2)$ and $j = (n', \alpha_1', \alpha_2')$.

- Then

$$\pi_i = \sum_j P_{ij} \pi_j \quad \text{or} \quad \pi = P\pi$$

and

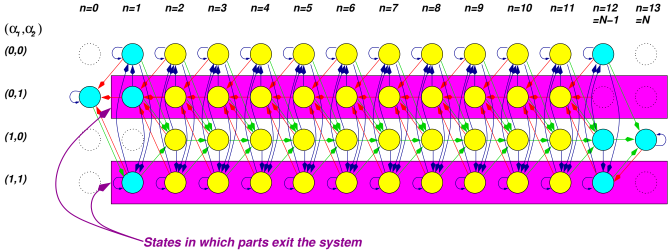
$$\sum_i \pi_i = 1 \quad \text{or} \quad \nu^T \pi = 1$$

- This is exactly what we discussed in the Markov Process videos, although the specific notation for this system is a little more complicated here.

Two-Machine, Finite-Buffer Lines

Calculation of performance measures

- To calculate the production rate in parts per operation time, add up the probabilities of the indicated states:



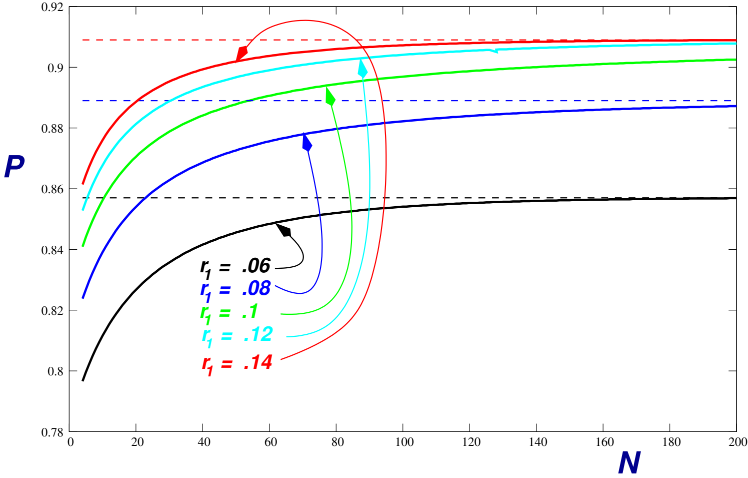
- To calculate the average inventory, evaluate

$$\bar{n} = \sum_{n=0}^{N} \sum_{\alpha_1=0}^{1} \sum_{\alpha_2=0}^{1} n\pi(n, \alpha_1, \alpha_2)$$

Two-Machine, Finite-Buffer Lines

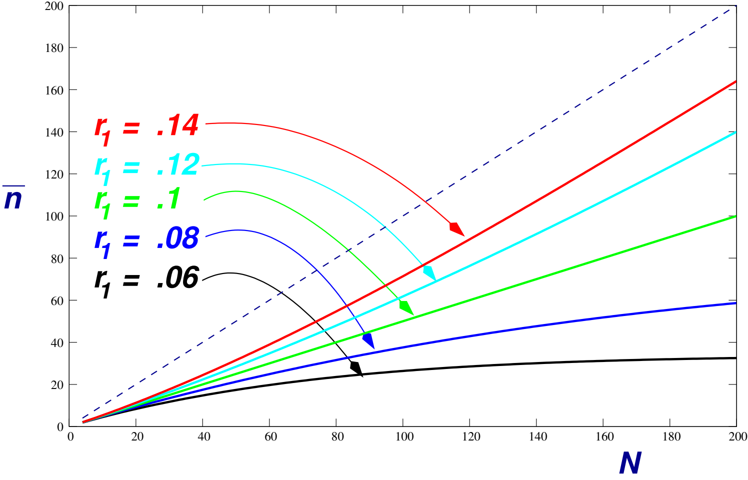
Production rate vs. Buffer Size

$p_1 = .01$
$r_2 = .1$
$p_2 = .01$

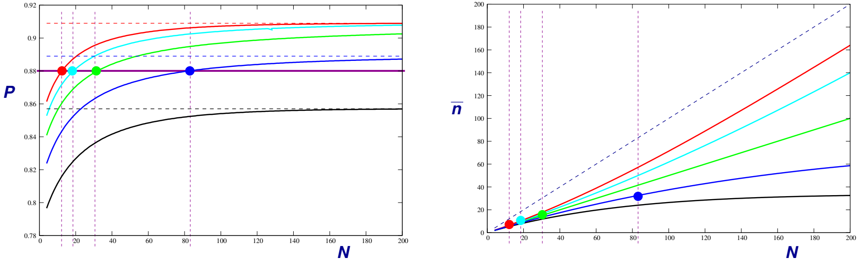


Two-Machine, Finite-Buffer Lines
Average Inventory vs. Buffer Size

$p_1 = .01$
$r_2 = .1$
$p_2 = .01$



Two-Machine, Finite-Buffer Lines
Line Design



Problem: Select M_1 and N so that $P = .88$.

Solutions:

r_1	MTTR	N	$\bar{n}$
.14	7.14	13	7.08
.12	8.33	19	10.12
.10	10.00	32	16.00
.08	12.50	82	32.21

Two-Machine, Finite-Buffer Lines

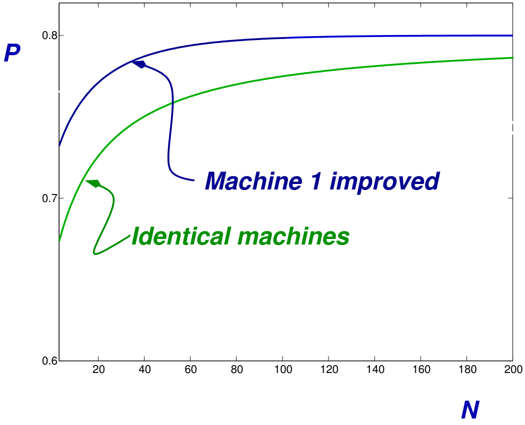
Improvements

Questions:

- If we want to increase production rate, which machine should we improve?
- What would happen to production rate if we improved any other machine?

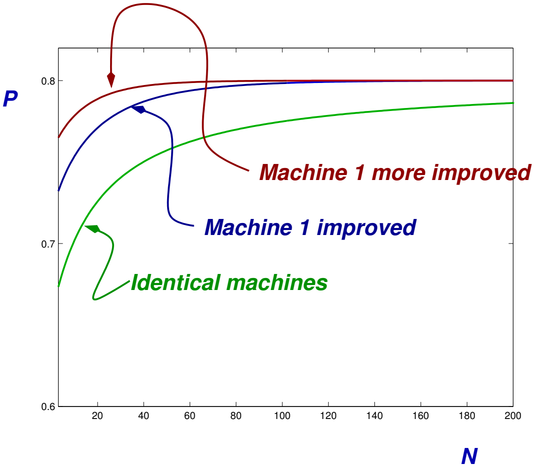
Two-Machine, Finite-Buffer Lines

Improvements to *non-bottleneck* machine.



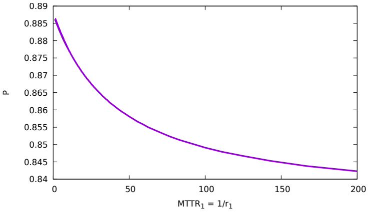
Two-Machine, Finite-Buffer Lines

Improvements to *non-bottleneck* machine.



Two-Machine, Finite-Buffer Lines
Failure frequency

Is it better to have short, frequent, disruptions or long, infrequent, disruptions?



- $r_2 = 0.1$, $p_2 = 0.009$, $N = 30$
- r_1 and p_1 vary together such that
  $$e_1 = \frac{r_1}{r_1+p_1} = \frac{\text{MTTF}}{\text{MTTF+MTTR}} = .9$$
- *Answer:* evidently, short, frequent failures.
- *Why?*

Introduction to Manufacturing Systems
Part 1

- I hope that you enjoyed this course and and that will find it useful.
- Part 2 continues the exploration of the issues that we have studied in Part 1: causes and consequences of variability in factories, and what to do about them.
- I hope you will come back for Part 2 and for the rest of the Principles of Manufacturing program.
- Please feel free to offer comments, criticisms, and suggestions for improvements.