ECE/CS 541
Computer System Analysis:
Introduction to Combinatorial Methods

Mohammad A. Noureddine
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

Fall 2018

Learning Objectives

- Or what is this course about?

- At the start of the semester, you should have
  - Basic programming skills (C++, Python, etc.)
  - Basic understanding of probability theory (ECE313 or equivalent)

- At the end of the semester, you should be able to
  - Understand different system modeling approaches
    - Combinatorial methods, state-space methods, etc.
  - Understand different model analysis methods
    - Analytic/numeric methods, simulation
  - Understand the basics of discrete event simulation
  - Design simulation experiments and analyze their results
  - Gain hands-on experience with different modeling and analysis tools

Announcements and Reminders

- HW1 is out
  - Covers the probability review
  - Prepare you for the probability quiz
  - Due on **September 18, 2018 at the start of class**

- Probability quiz on **September 20, 2018**
  - First 30 minutes of class

- **Project Proposals due near the first week of October**
  - Start forming groups and thinking about your projects
  - Come to office hours for discussions
  - List of possible projects and ideas on the website soon

- TA office hours: MW: 4:00 – 5:00 pm in CSL 231

Objectives for this Module

- Introduce combinatorial (non state-space) methods of modeling
- Develop and formulate models of system reliability
- Introduce different reliability formalisms
- Combinatorial models for improved testing research at Internet scale
  - Technique generated out of UC Santa Cruz and adopted by Netflix

Lecture Outline

- Assumptions for combinatorial modeling
- Review definition of reliability
- Failure rate
- System reliability
  - Maximum
  - Minimum
  - k of N
- Reliability formalisms
  - Reliability block diagrams
  - Fault trees

Introduction to Combinatorial Methods

- Combinatorial validation methods are the simplest kind of analytical/numerical techniques and can be used for reliability and availability modeling under certain assumptions.

- Assumption 1:
  - The system being studied is composed of several elementary units, called *components*.

- Assumption 2:
  - The components of the system fail in a statistically independent manner. For availability analysis, they can be repaired independently.

- When these assumptions hold, simple formulas for reliability and availability exist.

Choosing Validation Techniques cont.

Criterion	**Combinatorial**	**State-Space-Based**	**Simulation**	**Measurement**
Stage	Any	Any	Any	Post-prototype
Time	Small	Medium	Medium	Varies
Tools	Formulae, tools	Languages & tools	Languages & tools	instrumentation
Accuracy	Low	Moderate	Moderate	high
Comparisons	Easy	Moderate	Moderate	Difficult
Cost	Low	Low/medium	Medium	High
Scalability	High	Low/medium	Medium	low

Reliability

- One key to building highly available systems is the use of reliable components and systems.

- Reliability:

  – The *reliability* of a system at time t ($R(t)$) is the probability that the system operation is proper throughout the interval $[0,t]$.

- Probability theory and combinatorics can be directly applied to reliability models.

- Let X be a random variable representing the time to failure (TTF) of a component. The reliability of the component at time t is given by

$$R_X(t) = P(X > t) = 1 - P(X \leq t) = 1 - F_X(t)$$

- Similarly, we can define *unreliability* at time t by

$$U_X(t) = P(X \leq t) = F_X(t)$$

Failure Rate

What is the rate that a component fails at time *t*? This is the probability that a component that has not yet failed fails in the interval $(t, t + \Delta t)$, as $\Delta t \to 0$.

Note that we are not looking at $f_X(t)dt = P(X \in (t, t + \Delta t))$. Rather, we are seeking $P\left(X \in (t, t + \Delta t) \mid X > t\right)$

Failure Rate

What is the rate that a component fails at time t? This is the probability that a component that has not yet failed fails in the interval $(t, t + \Delta t)$, as $\Delta t \to 0$.

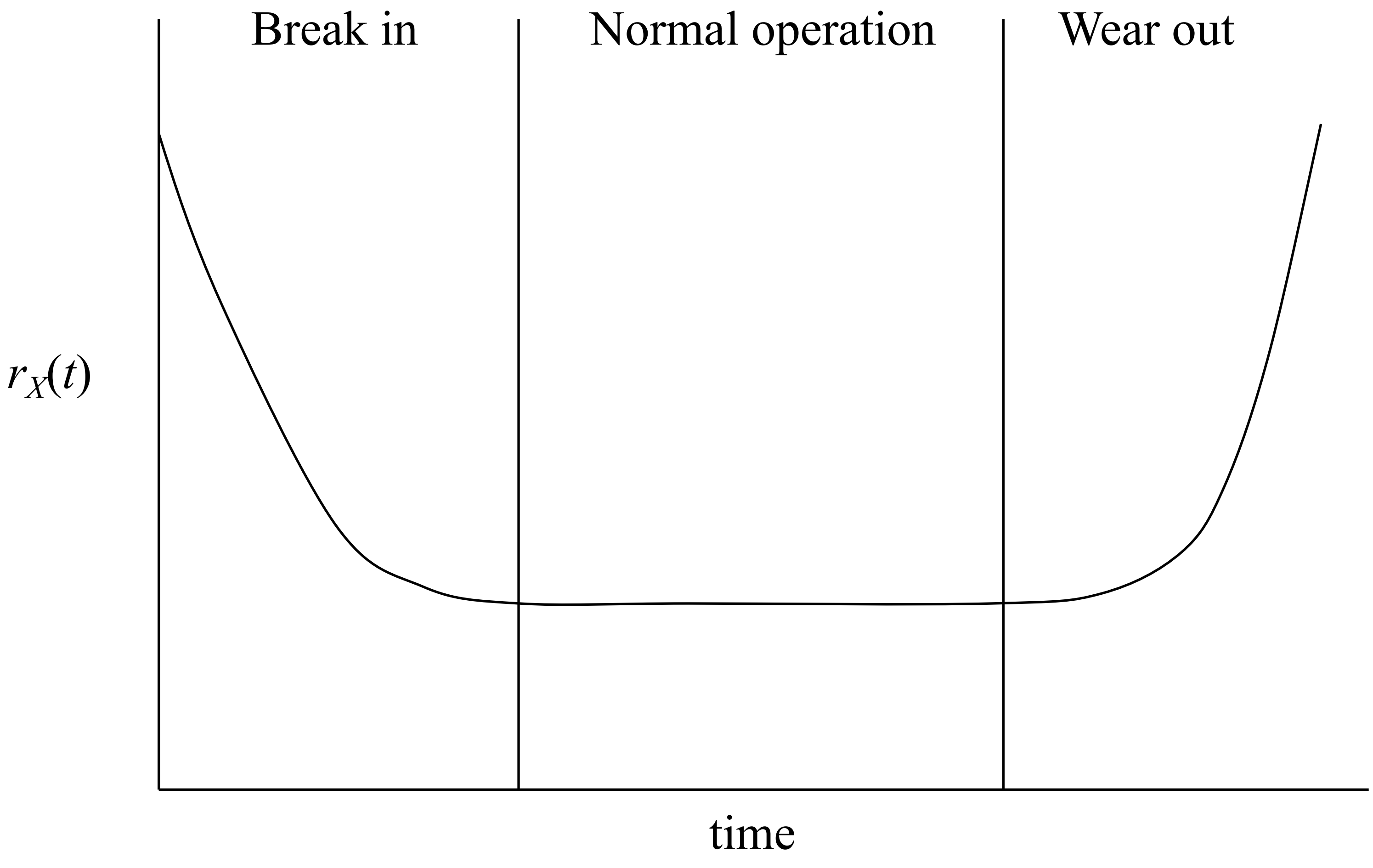
Note that we are not looking at $f_X(t)dt = P(X \in (t, t+\Delta t))$. Rather, we are seeking $P\left(X \in (t, t+\Delta t) \mid X > t\right)$

$$
\begin{aligned}
P\left(X \in (t, t+\Delta t) \mid X > t\right) &= \frac{P(X \in (t, t+\Delta t), X > t)}{P(X > t)} \\
&= \frac{P(X \in (t, t+\Delta t))}{1 - F_X(t)} \\
&= \frac{f_X(t)dt}{1 - F_X(t)} \triangleq r_x(t)dt
\end{aligned}
$$

$$
\boxed{r_x(t) = \frac{f_X(t)}{1 - F_X(t)} = \frac{f_X(t)}{R_X(t)}}
$$

$r_X(t)$ is called the *failure rate* or *hazard rate*.

Typical Failure Rate



System Reliability

While R_X can give the reliability of a component, how do you compute the reliability of a system?

System failure can occur when one, all, or some of the components fail. If one makes the *independent failure assumption*, system failure can be computed quite simply. The independent failure assumption states that all component failures of a system are independent, i.e., the failure of one component does not cause another component to be more or less likely to fail.

Given this assumption, one can determine:

1) Minimum failure time of a set of components
2) Maximum failure time of a set of components
3) Probability that k of N components have failed at a particular time t.

Maximum of *n* Independent Failure Times

Let $X_1, \ldots, X_n$ be independent component failure times. Suppose the system fails at time S if all the components fail.

Thus, $S = \max\{X_1, X_2, \ldots, X_n\}$

What is $F_s(t)$?

Maximum of *n* Independent Failure Times

Let $X_1, \ldots, X_n$ be independent component failure times. Suppose the system fails at time S if all the components fail.

Thus, $S = \max\{X_1, X_2, \ldots, X_n\}$

What is $F_s(t)$?

$$\begin{aligned} F_S(t) = P(S \leq t) &= P(X_1 \leq t \wedge X_2 \leq t \wedge \ldots \wedge X_n \leq t) \\ &= P(X_1 \leq t) \times P(X_2 \leq t) \times \ldots \times P(X_n \leq t) \\ &= F_{X_1}(t) F_{X_2}(t) \ldots F_{X_n}(t) \\ &= \Pi_{i=1}^{n} F_{X_i}(t) \end{aligned}$$

By definition!

By independence!

Minimum of *n* Independent Component Failure Times

Let $X_1, \ldots, X_n$ be independent component failure times. A system fails at time S if any of the components fail.

Thus, $S = \min\{X_1, \ldots, X_n\}$.

What is $F_S(t)$? Proof in Homework 1

$$F_S(t) = 1 - \Pi_{i=1}^{n}(1 - F_X(t))$$

k of N

Let $X_1, \ldots, X_n$ be component failure times that have identical distributions (i.e., $F_{X_1}(t) = F_{X_2}(t) = \ldots$).

- The system fails at time S if k of the N components fail

k of N

Let $X_1, \ldots, X_n$ be component failure times that have identical distributions (i.e., $F_{X_1}(t) = F_{X_2}(t) = \ldots$).

- The system fails at time S if *k* of the *N* components fail

$$F_S(t) = P\,(\text{at least } k \text{ components failed by time } t)$$
$$= P\,(\text{exactly } k \text{ failed } \vee \text{ exactly } k+1 \text{ failed } \vee \ldots \text{ exactly } N \text{ failed })$$
$$= P\,(\text{exactly } k \text{ failed}) + P\,(\text{exactly } k+1 \text{ failed}) + \ldots + P\,(\text{exactly } N \text{ failed})$$

k of N

Let $X_1, \ldots, X_n$ be component failure times that have identical distributions (i.e., $F_{X_1}(t) = F_{X_2}(t) = \ldots$).

- The system fails at time S if k of the N components fail

$$
\begin{aligned}
F_S(t) &= P(\text{at least } k \text{ components failed by time } t) \\
&= P(\text{exactly } k \text{ failed } \vee \text{ exactly } k+1 \text{ failed } \vee \ldots \text{ exactly } N \text{ failed }) \\
&= P(\text{exactly } k \text{ failed}) + P(\text{exactly } k+1 \text{ failed}) + \ldots + P(\text{exactly } N \text{ failed})
\end{aligned}
$$

$$
\begin{aligned}
P(\text{exactly } k \text{ failed}) &= P(k \text{ failed and } N-k \text{ have not}) \\
&= \binom{N}{k} F_X(t)^k (1 - F_X(t))^{N-k}
\end{aligned}
$$

Thus,

$$
\boxed{F_S(t) = \sum_{i=k}^{N} \binom{N}{i} F_X(t)^i (1 - F_X(t))^{N-i}}
$$

k of N in General

For non-identical failure distributions, we must sum over all combinations of at least k failures.

Let G_k be the set of all subsets of $\{X_1, \ldots, X_N\}$ such that each element in G_k is a set of size at least k, i.e.,

$$G_k = \{g_i \subseteq \{X_1, \ldots, X_N\} : |g_i| \geq k\}$$

k of N in General

For non-identical failure distributions, we must sum over all combinations of at least k failures.

Let G_k be the set of all subsets of $\{X_1, \ldots, X_N\}$ such that each element in G_k is a set of size at least k, i.e.,

$$G_k = \{g_i \subseteq \{X_1, \ldots, X_N\} : |g_i| \geq k\}$$

All possible failure scenarios

k of N in General

For non-identical failure distributions, we must sum over all combinations of at least k failures.

Let G_k be the set of all subsets of $\{X_1, \ldots, X_N\}$ such that each element in G_k is a set of size at least k, i.e.,

$$G_k = \{g_i \subseteq \{X_1, \ldots, X_N\} : |g_i| \geq k\}$$

All possible failure scenarios

Now F_S is given by

$$F_s(t) = \sum_{g \in G_k} \left(\prod_{X \in g} F_X(t) \right) \left(\prod_{X \notin g} (1 - F_X(t)) \right)$$

Component Building Blocks

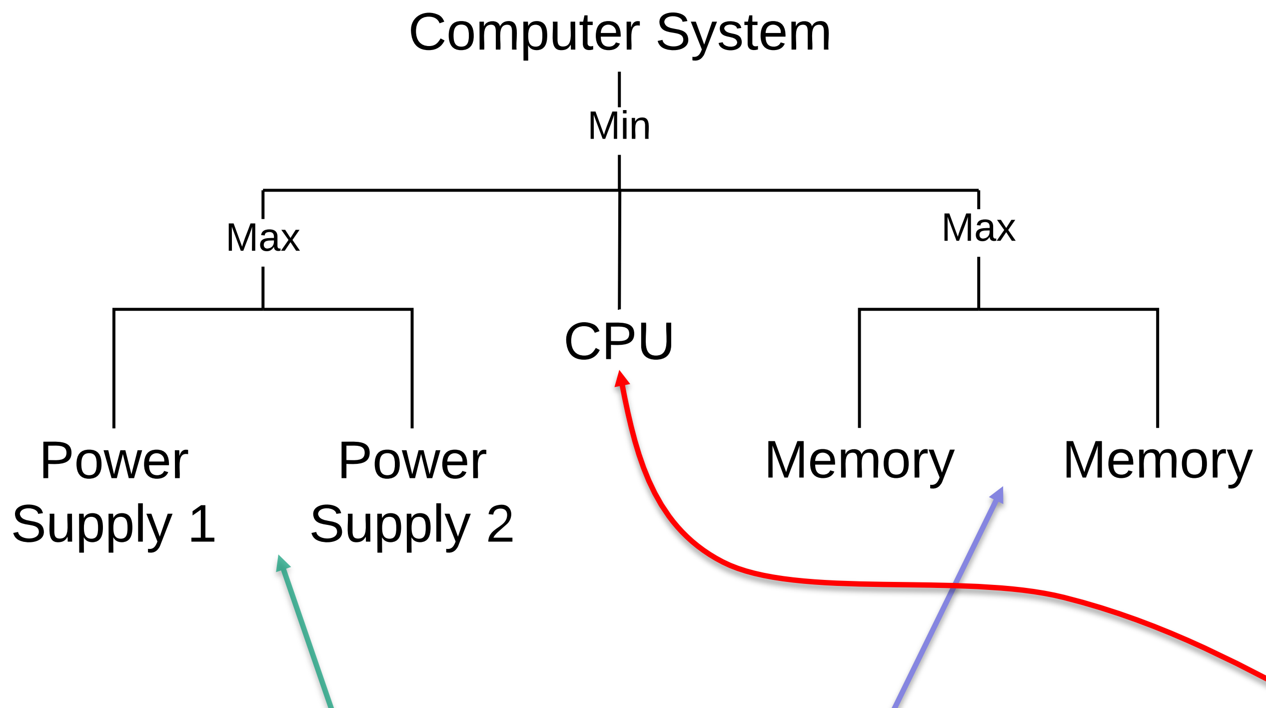
- Assumption 1 tells us that the systems we consider are composed of components.
  - So we can think about them hierarchically

- Consider a computer system that fails if:
  - Both power supplies fail, or
  - Both memories fail, or
  - The CPU fails

- Let's reason about the problem using our previously seen techniques.
  - Look at every component on its own
  - Build their composition

Component Building Blocks

- The computer system problem is one of a minimums
  - The system will fail when the first of its three subsystems fail



$$F_S(t) = 1 - \left(\left(1 - F_{P_1}(t) F_{P_2}(t)\right) \left(1 - F_{M_1}(t) F_{M_2}(t)\right) \left(1 - F_C(t)\right) \right)$$

Summary

A system comprises N components, where the component failure times are given by the random variables $X_1, \ldots, X_N$. The system fails at time S with distribution F_S if:

Condition	Distribution
All components fail	$F_S(t) = \prod_{i=1}^{N} F_{X_i}(t)$
One component fails	$F_S(t) = 1 - \prod_{i=1}^{N} (1 - F_{X_i}(t))$
k components fail, i.i.d	$F_S(t) = \sum_{i=k}^{N} \binom{N}{i} F_X(t)^i (1 - F_X(t))^{N-i}$
k components fail, general case	$F_S(t) = \sum_{g \in G_k} \left(\prod_{X \in g} F_X(t) \right) \left(\prod_{X \notin g} (1 - F_X(t)) \right)$

Reliability Formalisms

There are several popular graphical formalisms to express system reliability. The core of the solvers is the methods we have just examined.

In particular, we will examine

- Reliability Block Diagrams
- Fault Trees
- Reliability Graphs

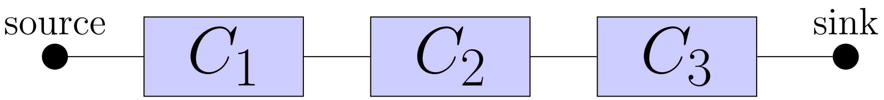
There is nothing particularly special about these formalisms except their popularity. It is easy to implement these formalisms, or design your own, in a spreadsheet, for example.

Reliability Block Diagrams

- Blocks represent components.
- A system failure occurs if there is no path from source to sink.

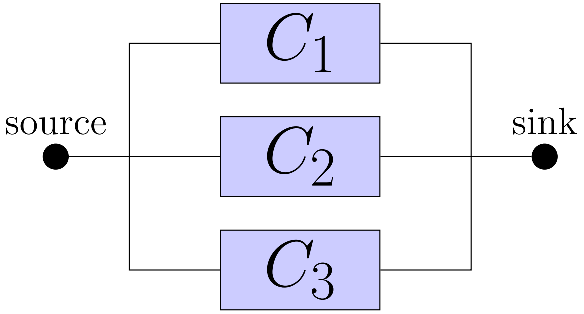
Series:

System fails if any component fails.



Parallel:

System fails if all components fail.



k of *N*:

System fails if at least *k* of *N* components fail.